

**Allelism and allele sequence divergence of *LOP*, the
locus of parthenogenesis in the model apomict
Hieracium praealtum (Asteraceae)**

A thesis submitted in partial fulfilment of the requirements for the degree

of

Master of Science in Plant Biology

at the

University of Canterbury



by **Robert McGee**

Supervised by:

Dr. Anthony M. Poole, Dr. Marie Hale

Dr. Ross Bicknell, Dr. Andrew Catanach

August 2013

School of Biological Sciences

TABLE OF CONTENTS

ACKNOWLEDGEMENTS	I
ABSTRACT	II
LIST OF FIGURES	III
LIST OF TABLES	V
ABBREVIATIONS	VI

CHAPTER 1 : GENERAL INTRODUCTION

1.1. Literature Review	1
1.1.1. Sexual Reproduction versus Apomixis	1
1.1.2. The Potential Value of Apomixis to Agriculture	1
1.1.3. The Genetic Basis of Apomixis.....	2
1.1.4. Understanding the <i>LOP</i> locus.....	4
1.1.5. Research Hypothesis and Aims	7

CHAPTER 2 : GENOTYPING

2.1. Introduction.....	9
2.1.1. Genetic Basis of Apomixis	9
2.1.2. The Alternative Alleles of Candidate Parthenogenesis Genes	10
2.1.3. Polyhaploids.....	11
2.1.4. Genotyping the Alternative Alleles	11
2.1.5. Inheritance of Alleles.....	12
2.1.6. Recombination of Alleles	12
2.1.7. Research Hypothesis.....	13

2.1.8. Aims of the Study	13
2.2. Materials and Methods	14
2.2.1. Plant Material	14
2.2.2. BAC Isolation and Sequencing	15
2.2.3. Genotyping - Fragment Length Polymorphism	18
2.2.4. Genotyping – High Resolution Melt	19
2.2.5. Allele Segregation and Recombination	20
2.3. Results	22
2.3.1. BAC Sequence Data	22
2.3.2. Genotyping - Fragment Length Polymorphism	26
2.3.3. Genotyping – High Resolution Melt	32
2.3.4. Segregation of Alleles.....	35
2.3.5. Recombination of Alleles	39
2.4. Discussion.....	41
2.4.1. Polyhaploid Genotyping	41
2.4.2. Segregation of Alleles.....	42
2.4.3. Recombination Between Genes of <i>LOP</i> and Alternative Alleles.....	44
2.4.4. Inheritance of Alleles	47
2.4.5. Future Work	47
2.4.6. Conclusion.....	49

CHAPTER 3 : ALIGNMENTS

3.1. Introduction.....	50
3.1.1. Allelic Sequence Divergence.....	50
3.1.2. Consequences of Allelic Sequence Divergence.....	50

3.1.3. ASD of Apomixis Genes and Loci.....	51
3.1.4. Allelic Sequence Divergence of the <i>LOP</i> Candidate Genes.....	52
3.1.5. Research Hypothesis	53
3.1.6. Aims of the Study	53
3.2. Materials and Methods	55
3.2.1. Sequence Data.....	55
3.2.2. Multiple Sequence Alignments using ClustalW, and Generation of Distance Trees ..	55
3.2.3. Genomic Sequence Alignments using Mauve	56
3.2.4. Transposable Elements	56
3.3. Results	57
3.3.1. Multiple Sequence Alignments	57
3.3.2. Distance Trees and Distance Matrices	58
3.3.3. Genomic Sequence Alignments using Mauve	60
3.3.4. Transposable Elements in the Genomic Sequence Alignments	61
3.4. Discussion.....	66
3.4.1. ASD of the <i>LOP</i> Allele.....	66
3.4.2. Causes of ASD of the <i>LOP</i> Allele.....	67
3.4.3. Origin of <i>LOPY</i>	69
3.4.4. Origin of <i>LOPX</i> and <i>LOPB</i>	69
3.4.5. Conclusion.....	70

CHAPTER 4 : GENERAL DISCUSSION

4.1. Overview	72
4.1.1. Hypothesis One: ‘Alternative Alleles are Allelic to the <i>LOP</i> Allele’	72

4.1.2. Hypothesis Two: ‘ <i>LOP</i> Alleles Display ASD as a Result of Repressed Recombination at <i>LOP</i> ’	73
4.1.3. Hypothesis Three: ‘All Alleles Originated From an Ancestral Allele’	74
4.2. Models of the Evolution of <i>LOP</i>	74
4.3. Future Work	76
4.4. Overall Conclusions	77
REFERENCES	78
APPENDICES	85

ACKNOWLEDGEMENTS

I would like to start by acknowledging the Plant and Food Research (PFR) summer studentship programme, which I undertook in 2011 and where I was first introduced to the study of apomixis. It was during this time that I decided to undertake my MSc thesis in that field. It was a collaborative effort between the PFR apomixis research team, based in Lincoln, New Zealand and the School of Biological Sciences at the University of Canterbury, Christchurch, New Zealand.

I would like to thank Dr. Ross Bicknell, the team leader of the PFR apomixis research team, for introducing me to the project and giving me access to the resources, plant material, genetic tools, laboratory space and expertise to undertake this MSc project. He is an excellent supervisor and mentor, and great at explaining how everything fits into the big picture. Thank you Ross. Within the apomixis research team, I would also like to acknowledge Dr. Andrew Catanach, my other supervisor. I cannot thank Andrew enough for the time he spent helping and advising me on my project. His day-to-day guidance was invaluable throughout the project from the planning phase to the final write-up. I could have not completed this project without his help. Thank you Andrew. I would also like to thank Dr. Anthony M. Poole and Dr. Marie Hale, my university supervisors, who have been a great help during the write-up and final stages of the project. Their comments and insights helped to improve the clarity of the thesis. You did a great job, thank you.

I am also grateful to Mrs Sylvia Erasmuson, a lab assistant within the PFR apomixis research team, who helped me complete all the DNA extractions, flow cytometry work, maintaining the polyhaploids in the greenhouse and providing valuable advice throughout the project. I would also like to thank Dr. Samantha Baldwin, PFR Lincoln, for advising me to use the HOT FIREPol[®] EvaGreen[®] HRM Mix and the associated PCR procedures for HRM genotyping. Dr. Robert Braun, my office and lab mate also provided valuable insights into the HRM analysis step as well as providing really helpful advice on undertaking laboratory procedures. Thanks to you all.

Finally, on a personal note I would like to thank my parents for their continued support and assistance during my master's thesis.

ABSTRACT

Apomixis, or asexual seed development, if introduced into crop species, has the potential to greatly improve global food production. Towards this goal, this study focused on uncovering the genetic mechanisms that control the parthenogenesis step within apomixis whereby fertilisation is avoided. In the model apomict, *Hieracium praealtum* (Asteraceae), parthenogenesis is controlled by the *LOSS OF PARTHENOGENESIS (LOP)* locus. Previous research showed that in addition to genomic copies of candidate genes at *LOP*, the genome has at least three other copies referred to as alternative alleles. The main goal of this study was to investigate four candidate genes, Genes B, X, H and Y, at *LOP* by generating segregation data of the alternative alleles. BAC clones containing alternative allele sequences were identified and Roche 454 pyrosequenced. These sequences were used to design alternative allele specific primers for genotyping two *Hieracium praealtum* polyhaploid populations (~ 300 plants).

Four major conclusions were drawn from this study. First, the alternative alleles were in fact acting like alleles to the *LOP* alleles of Genes B, X and Y. Second, allelic sequence divergence (ASD) of the *LOP* alleles of Genes B and X relative to the alternative alleles, indicated a recent and separate evolutionary history. Third and, unexpectedly, recombination was detected at the *LOP* locus, in contrast to other apomixis loci reported in the literature. Furthermore, Gene B was found to be very closely associated with parthenogenesis in the polyhaploid population indicating that it may be essential to parthenogenesis and therefore requires further investigation. On the other hand, the absence of Genes X, Y and H, due to recombination, had no impact upon parthenogenesis. Fourth, the sequence data suggested that the *LOP* and alternative alleles originated from a shared common allele ancestor. It is hoped that these findings have made a significant contribution towards the future goal of introducing apomixis into crop species.

LIST OF FIGURES

Figure 1.1.	Apomixis in <i>Hieracium praealtum</i>	4
Figure 1.2.	a. The four alleles of the tetraploid <i>Hieracium praealtum</i> parent plant. b. The expected allele arrangement of the three main polyhaploid classes	5
Figure 1.3.	Polyhaploid generation	6
Figure 1.4.	Simplified schematic tree of allelic sequence divergence of <i>LOP</i>	7
Figure 2.1.	Physical map of the <i>LOP</i> locus	10
Figure 2.2.	Identification of BAC plates containing the alternative allele sequences using gel electrophoresis of PCR products. a. Superpool plate testing. b. Pool plate testing	22
Figure 2.3.	Identification of BAC colonies containing the alternative allele sequences using gel electrophoresis of PCR products. a. Row testing in BAC plate. b. Colonies testing on identified row of BAC plates	23
Figure 2.4.	DNA sequence alignments of existing <i>LOP</i> alleles and alternative alleles sequences with new Sanger-derived BAC sequences of PCR products from Gene X (a) and Gene Z (b)	24
Figure 2.5.	DNA sequence alignments of existing <i>LOP</i> alleles and alternative alleles sequences with new Sanger-derived BAC sequences of PCR products from Gene B	24
Figure 2.6.	Fragment length polymorphism genotyping of Gene X. a. Fragment profiles. b. Association between fragment peaks and sequence data	29
Figure 2.7.	Fragment length polymorphism genotyping of Gene B. a. Fragment profiles. b. Association between fragment peaks and sequence data	30
Figure 2.8.	Fragment length polymorphism genotyping of Gene H. a. Fragment profiles. b. Association between fragment peaks and sequence data	31
Figure 2.9.	Distribution of the alternative alleles of Genes X, B and H in the two <i>Hieracium praealtum</i> polyhaploid populations generated through fragment length polymorphism genotyping	32
Figure 2.10.	High Resolution Melt (HRM) genotyping of Gene Y using PM92. a. Normalised melting curve profiles. b. Normalised melting peaks	33
Figure 2.11.	Association between the melting curves of Gene Y and the type of SNP	34

Figure 2.12.	Distribution of the alternative alleles of Gene Y in the EMS15C <i>Hieracium praealtum</i> polyhaploid population generated through HRM-based genotyping	34
Figure 2.13.	Bar chart of the frequency distribution of allele combinations of Genes X and B in the two <i>Hieracium praealtum</i> polyhaploid populations	35
Figure 2.14.	Bar chart of the frequency distribution of allele combinations of Genes B and H in the two <i>Hieracium praealtum</i> polyhaploid populations	36
Figure 2.15.	Bar chart of the frequency distribution of allele combinations of Genes B and Y in the two <i>Hieracium praealtum</i> polyhaploid populations	37
Figure 2.16.	Bar chart of the frequency distribution of allele combinations of Genes Y and H in the two <i>Hieracium praealtum</i> polyhaploid populations	38
Figure 2.17.	Confirmation of the ploidy of the recombinant plants from the 2011 polyhaploid population using flow cytometry	40
Figure 3.1.	Multiple sequence alignments of the four alleles of Gene B (a) and of Gene X (b)	57
Figure 3.2.	Multiple sequence alignment of the entire genomic region of the three sequenced Gene B alleles	58
Figure 3.3.	Neighbour-joining distance trees of the four alleles of Gene B (a) and of Gene X (b)	59
Figure 3.4.	Mauve alignment of 25 kb downstream and 25 kb upstream of Gene B between <i>LOPB</i> and the B1 and B2 alternative alleles	62
Figure 3.5.	Mauve alignment of 25 kb downstream and 25 kb upstream of Gene X between <i>LOPX</i> and the X1 alternative allele	63
Figure 3.6.	Mauve alignment of 25 kb downstream and 25 kb upstream of Gene Y between <i>LOPY</i> and the Y1 alternative allele	64
Figure 4.1.	Model proposed to explain the occurrence of ASD at the <i>LOP</i> allele	75

LIST OF TABLES

Table 2.1.	Read count and sequence statistics from BAC sequencing	25
Table 2.2.	The BAC assembly progress using gsAssembler v2.6 & v2.7 and associated statistics	26
Table 2.3.	Frequency distribution of allele combinations of Genes X and B in the two <i>Hieracium praealtum</i> polyhaploid populations	35
Table 2.4.	Frequency distribution of allele combinations of Genes B and H in the two <i>Hieracium praealtum</i> polyhaploid populations	36
Table 2.5.	Frequency distribution of allele combinations of Genes B and Y in the two <i>Hieracium praealtum</i> polyhaploid populations	37
Table 2.6.	Frequency distribution of allele combinations of Genes Y and H in the two <i>Hieracium praealtum</i> polyhaploid populations	38
Table 2.7.	The presence (+) or absence (-) of specific <i>LOP</i> alleles in eight recombinant plants identified in the two polyhaploid populations	39
Table 2.8.	Recombination rates shown as percentages between the alleles of Genes X and B	40
Table 3.1.	Distance matrices of the number of nucleotide substitutions per 100 nucleotides between each allele of Gene B (a) and of Gene X (b)	60
Table 3.2.	The number of transposable elements in the ~ 50 kb sequence alignments of Genes B, X and Y within the <i>LOP</i> allele and alternative alleles	65
Table A.1.	SSR-based primer mixes used to identify BACs containing alternative allele sequences	85
Table A.2.	Panel of existing or new SSR primers selected for genotyping	86
Table A.3.	Panel of existing or new HRM primers selected for genotyping	87
Table A.4.	Panel of additional <i>LOP</i> sequence-specific primers screened on the recombinant plants from the 2011 polyhaploid population	87
Table A.5.	Genotype data of Genes B, X and H from the 2008 polyhaploid population	88
Table A.6.	Genotype data of Genes B, X, H and Y from the 2011 polyhaploid population	94

ABBREVIATIONS

%	Percentage
°C	Degrees Celsius
µl	Microlitre
µM	Micromoles per litre
A	Adenine nucleotide base
ASD	Allelic sequence divergence
ATP	Adenosine-5'-triphosphate
BAC	Bacterial artificial chromosome
BLAST	Basic Local Alignment Search Tool
bp	Base pairs
C	Cytosine nucleotide base
cm	Centimetre
CpG	Cytosine-phosphate-guanine
DNA	Deoxyribonucleic acid
dNTP	Dinucleotide triphosphate
<i>E.coli</i>	<i>Escherichia coli</i>
EDTA	Ethylenediaminetetraacetic acid
eIF3e	Eukaryotic translation initiation factor 3 subunit E
eIF4	Eukaryotic translation initiation factor 4
EMS	Ethyl methanesulfonate
EMS15C	Mutant <i>Hieracium praealtum</i> parent plant derived from an ethyl methanesulfonate mutagenesis screen
EXO	Exonuclease I
g	Grams

<i>g</i>	Earth's gravitational acceleration
G	Guanine nucleotide base
HRM	High Resolution Melt
Kb	Kilobase
Kcal/mol	Kilocalorie per mole
LB	Lysogeny broth (formally known as Luria-Bertani medium)
<i>LOA</i>	<i>LOSS OF APOMEIOSIS</i>
<i>LOP</i>	<i>LOSS OF PARTHENOGENESIS</i>
Mb	Mega base pairs
Milli-Q	Purified and deionised water produced with Millipore Corporation equipment
ml	Millilitres
mM	Millimolar
mRNA	Messenger RNA
<i>n</i>	Number of chromosomes
ng	Nanogram
<i>p</i>	Probability value
PCR	Polymerase chain reaction
PFR	Plant & Food Research ®
pH	Hydrogen ion concentration
PM	Primer mix
qPCR	Quantitative real time polymerase chain reaction
R35	<i>Hieracium praealtum</i> parent plant
RNA	Ribonucleic acid
RNAi	RNA interference
RNA-seq	RNA sequencing

ROX	6-carboxy rhodamine
RT-PCR	Reverse transcription polymerase chain reaction
SAP	Shrimp Alkaline Phosphatase
SNP	Single nucleotide polymorphism
SSR	Simple sequence repeat (microsatellite marker)
T	Thymine nucleotide base
TAMRA-dUTP	Tetramethylrhodamine-5(6)-[5-(3-carboxyaminoallyl) 2'-deoxyuridine 5' triphosphate]
TEs	Transposable elements
U	Unit of enzyme
UV	Ultraviolet
ΔG	Gibbs free energy

CHAPTER 1

GENERAL INTRODUCTION

1.1. Literature Review

1.1.1. Sexual Reproduction versus Apomixis

Seed development in angiosperms typically arises through sexual reproduction. Seed development is initiated after pollination in which male pollen grains attach to the stigma and release male gamete cells which travel to the female embryo sac via the pollen tube. One male gamete cell fertilizes the egg forming a diploid zygote and the other gamete cell fertilizes two haploid polar nuclei forming a triploid cell, an event known as double fertilisation. The zygote develops into the embryo and the triploid cell becomes the nutritious endosperm of the seed.

Some angiosperms have circumvented the requirement for sexual reproduction and have developed the ability to produce seeds asexually through the process of apomixis (Richards, 2003). Apomixis proceeds through two critical steps, the avoidance of meiosis (apomeiosis) followed by the avoidance of fertilisation (parthenogenesis). The formation of the endosperm can either be autonomous or, as is commonly found in grasses, it can be pseudogamous, which requires pollination (Koltunow, 1993). Apomixis can further be classified into a sporophytic type and two gametophytic sub-types. These are known as diplospory and apospory sub-types, which are based on different mechanisms that lead to clonal seeds. During sporophytic apomixis, adventitious embryos arise spontaneously from nucellus or integument cells within the ovule, which then out-compete the existing zygotic embryos and mature into clonal seeds (Ozias-Akins, 2006). Gametophytic apomixis involves the production of unreduced embryo sacs either from megaspore mother cells in the diplospory sub-type (Whitton *et al.*, 2008) or from aposporous initials in the apospory sub-type (Bicknell & Koltunow, 2004).

1.1.2. The Potential Value of Apomixis to Agriculture

One of the primary motivations for apomixis research is the hope that it can be introduced into crop species. This is envisioned to give rise to two major advantages. First, it would allow for

large-scale production of genetically identical seed. Second, hybrid cereal cultivars with superior genotypes could be perfectly maintained through subsequent generations (Bicknell & Koltunow, 2004; Dwivedi *et al.*, 2010). Furthermore, it has been suggested that apomixis would provide superior cereal cultivars than those produced using current practices, which are typically maintained through inbreeding (Khush *et al.*, 1998). Other important advantages associated with inducing apomixis in crop species, include rapid generation time, reduction in the cost of plant breeding, prevention of viral transfer through propagated vegetation and the avoidance of pollination and cross-compatibility issues with other species (Jefferson & Bicknell, 1995; Savidan *et al.*, 2001). With these foreseeable advantages, apomixis is seen as one of the key future technologies that could improve global food production. This is of tremendous importance given a growing world population, shrinking environmental resources and the impacts of climate change (Germanà, 2011; Ronald, 2011). In one scenario based on the successful transfer of apomixis into rice with only modest rates of adoption, the economic improvement to welfare was estimated to be in excess of 4 billion US Dollars per annum (McMenimen & Lubulwa, 1997).

With the potential for apomixis to help alleviate pending world food shortages, the next critical step is to determine how it might be introduced into crop species. Progress has been hampered by the lack of an apomictic crop species to study. Almost all cultivated species show no apomixis. The few exceptions are either tropical fruit trees, which include mango, mangosteen and *Citrus*, or tropical forage grasses (Koltunow *et al.*, 1996). Efforts have been made to hybridise crop species with related apomicts but these have had only limited success. One such example is the apomictic maize-*Tripsacum* hybrids, which unfortunately have low seed sets and developmental defects, which are thought to be caused by unforeseen interspecific and ploidy barriers (Savidan, 1999; Leblanc *et al.*, 2009). In order to avoid these problems and to enable the introduction of apomixis into crop species, one of the proposed solutions has been to develop an understanding of the genes that control apomixis (Bicknell & Koltunow, 2004).

1.1.3. The Genetic Basis of Apomixis

The first apomixis research was unknowingly conducted by Gregor Johann Mendel in 1869 within the *Hieracium* genus (Asteraceae) (Mendel, 1869). *Hieracium* inheritance data collected by

Mendel was found to conflict with his groundbreaking *Pisum* data, and was therefore disregarded as a biological oddity. Rosenberg (1907) and Ostenfeld (1910) revisited the data and formally described apomixis in this genus.

With the development of molecular genetics tools, apomixis was found to be genetically controlled, and widely believed to be inherited as a single locus. This was reported in the aposporous grasses *Brachiaria* (Valle *et al.*, 1994), *Pennisetum* (Sherwood *et al.*, 1994), *Hieracium* (Bicknell *et al.*, 2000) and within *Taraxacum* (Van Dijk *et al.*, 1999). Subsequent research revealed that in some species there were more progeny types than previously described. As a result, the existing research was reassessed, which led to a shift in the prevailing viewpoint towards the existence of two unlinked loci in *Hieracium* (Bicknell & Koltunow, 2004) and in *Taraxacum officinale* (Van Dijk *et al.*, 2009).

Further investigations revealed that apomixis loci were typically situated within regions of repressed meiotic recombination, as described in *Pennisetum* (Goel *et al.*, 2003) and *Paspalum* (Labombarda *et al.*, 2002). This type of region prevents the use of conventional recombination-based linkage mapping to dissect the apomixis loci. With this in mind, an alternative approach was undertaken by Catanach *et al.* (2006) in which deletion mutagenesis was used to map apomixis loci in *Hieracium praealtum* (formally described as *Hieracium caespitosum*). Two dominant independent genetic loci were identified that control apomixis, termed *LOSS OF APOMEIOSIS* (*LOA*; Okada *et al.*, 2007) and *LOSS OF PARTHENOGENESIS* (*LOP*; Catanach *et al.*, 2006; see Figure 1.1). The *LOP* locus is currently being investigated by the PFR apomixis research team based in Lincoln, New Zealand, while the *LOA* loci is being investigated by collaborators. The genomic region of *LOP* has been sequenced, a BAC library was constructed and a number of candidate genes have been predicted (Dr. Andrew Catanach personal communication). Out of these genes, four have been selected for investigation in this study termed Gene B, Gene H, Gene X and Gene Y. The primary focus of this study was to gain a greater understanding of these candidate genes at the *LOP* locus within *Hieracium praealtum*. This will assist international research efforts to further understand the genes that control apomixis and is hoped to lead to its eventual introduction into crop species.

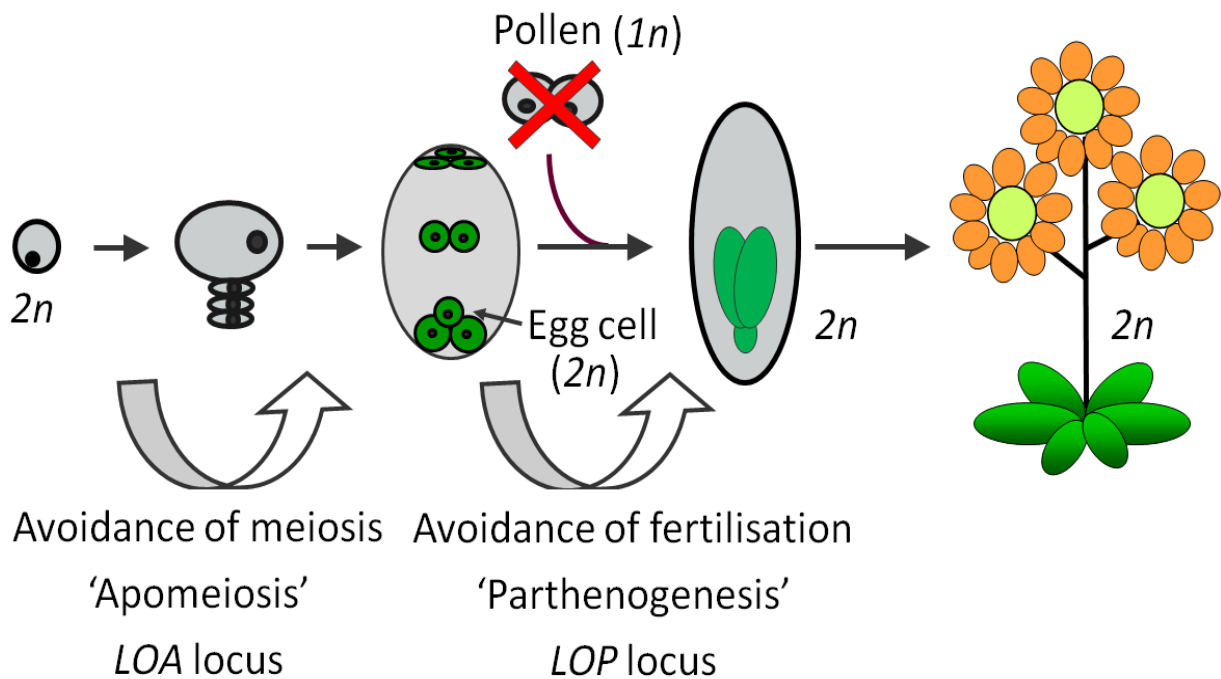


Figure 1.1. Apomixis in *Hieracium praealtum*. Somatic cells ($2n$) avoid meiosis through the process of apomeiosis controlled by the *LOA* locus resulting in an unreduced egg cell ($2n$). Fertilisation of the unreduced egg cell is avoided through the process of parthenogenesis controlled by the *LOP* locus. Finally, an endosperm develops via the process of autonomous endospermy. The resulting seed contains a $2n$ plant with the same genetic material as the parent plant. Modified from Catanach *et al.*, 2006.

1.1.4. Understanding the *LOP* locus

Alternative Alleles of the Candidate Genes

As part of my Plant & Food Research (PFR) summer studentship undertaken in 2010-2011 with the apomixis research team based in Lincoln, New Zealand; PCR based sequencing was performed using exon based primers across introns of Genes B, X and Y at the *LOP* locus within *Hieracium praealtum*. The PCR products from these genes were found to contain more than one variant, which significantly differed from the *LOP* sequence. This was not surprising since *Hieracium praealtum* is a tetraploid (Chapman & Bicknell, 2000), therefore each individual plant is expected to contain four alleles of each gene (Nielsen, 1980; Wolf *et al.*, 1990; Diwan *et al.*, 2000; Mahy *et al.*, 2000; Tiffin & Gaut, 2001). These alternative PCR sequences were hypothesised to represent recessive copies or recessive alleles (Figure 1.2a, McGee, PFR Summer Studentship 2010-2011,

unpublished data). During the current study, it was unknown if these recessive copies were acting as alleles to the *LOP* genes and therefore were referred to as ‘alternative alleles’.

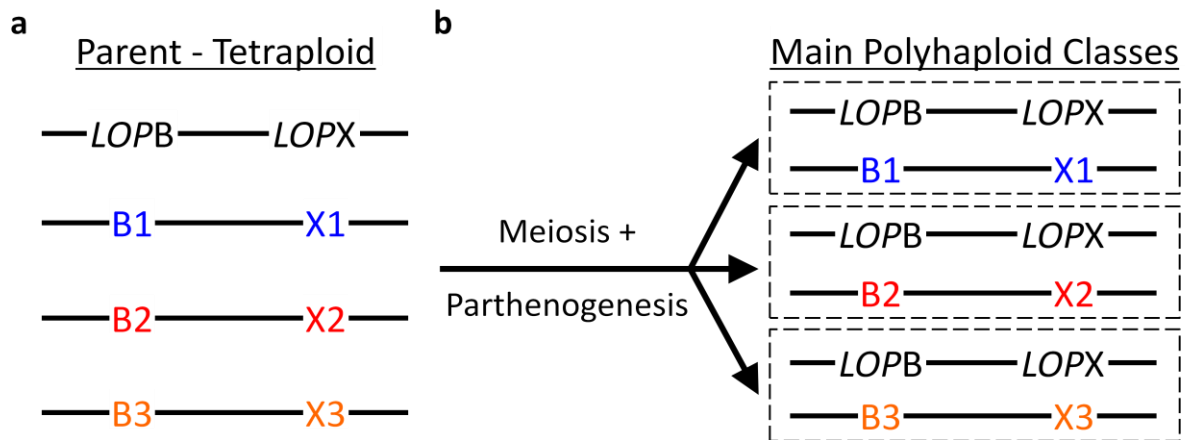


Figure 1.2. a. The diagram shows the four variants of each of Genes X and B discovered during my PFR summer studentship (McGee, PFR Summer Studentship 2010-2011, unpublished data). The *LOP* variants of Genes B and X are abbreviated to ‘*LOPB*’ and ‘*LOPX*’, and reside on the same chromatid based on previous deletion mapping of the *LOP* sequence (Catanach *et al.*, 2006). The ‘alternative alleles’ are also abbreviated, for example ‘X1’ refers to Gene X alternative allele 1. The arrangement of specific pairs of alternative alleles residing on the same chromatid was later confirmed in this study via BAC sequencing. **b.** Based on the expected allele arrangement, meiosis and parthenogenesis should produce polyhaploid plants that only contain two alleles; the *LOP* variant together with only one of the three previously identified alternative alleles. Three main polyhaploid classes are expected. It is also possible that some plants undergo recombination resulting in different combinations of alleles.

Polyhaploids

In order to gain a greater understanding of the alternative alleles, one of the focuses of this study was to generate segregation data of the alternative alleles using populations of a rare progeny type called a polyhaploid. Unlike true clonal seeds ($2n$), a polyhaploid is the product of a reduced meiotic egg that has then undergone parthenogenesis ($1n$) (see Figure 1.3; Bicknell *et al.*, 2003). Polyhaploids are a unique mapping resource for three main reasons. First, unlike true apomictic seeds which are clones, each polyhaploid plant is variable since it is a meiotically derived individual and has allele segregation which can be used for gene mapping. Second, all

polyhaploids contain only two copies of the haploid genome ($2x$) and are therefore expected to contain only two alleles. The *LOP* allele is always present (see Figure 1.2b) because it is required for the expression of parthenogenesis in the gamete (Koltunow *et al.*, 2011) and the subsequent production of polyhaploids. The second allele in the polyhaploid is expected to be one of the three previously identified alternative alleles. Third, since polyhaploids contain only maternal alleles, they carry only half the normal genome complement of the parent. This reduces the overall genetic complexity and simplifies the mapping process.

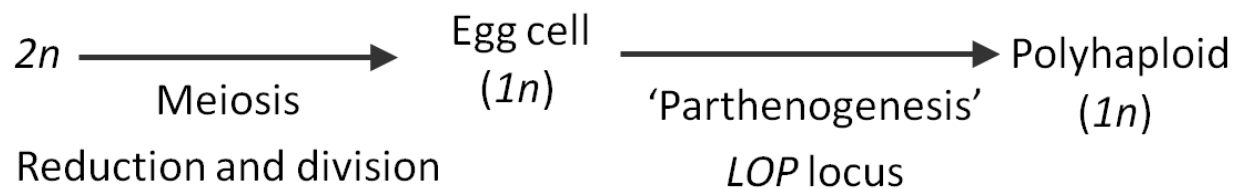


Figure 1.3. During normal gamete formation in *Hieracium praealtum*, a typical tetraploid cell ($2n = 3x + 8 = 35$) undergoes meiotic reduction and division to produce a reduced egg cell ($1n$). The reduced egg cell then undergoes parthenogenesis, which leads to the production of a polyhaploid ($1n$).

Allelic Sequence Divergence of the LOP Allele

Another significant finding of the PFR summer studentship I undertook (McGee, PFR Summer Studentship 2010-2011, unpublished data) was an indication of allelic sequence divergence (ASD) between the *LOP* allele and corresponding alternative alleles (see Figure 1.4). In the case of asexual plants, ASD is thought to arise as a result of the repression of meiotic recombination of certain alleles with the rest of the genome over evolutionary time (Corral *et al.*, 2009). Apomixis loci have been reported to exist within zones of repressed recombination in the apomicts, *Pennisetum* and *Paspalum* (Labombarda *et al.*, 2002; Goel *et al.*, 2003). To determine if the *LOP* allele has undergone ASD in *Hieracium praealtum*; the first step in this study was to generate sequence data of the alternative alleles. These sequences were generated via Roche 454 pyrosequencing of bacterial artificial chromosomes (BACs) containing alternative allele sequences. These BACs were originally derived from an existing *Hieracium praealtum* BAC library generated by Dr. Ross Bicknell as part of the PFR apomixis research team based in Lincoln, New Zealand. ASD could then be detected by comparing complementary genetic regions in the *LOP* allele to the alternative alleles. Large differences were detected in the sequence of the *LOP*

alleles, which is consistent with ASD. The alternative allele sequence data also assisted efforts to understand the evolution and origin of the *LOP* locus. One theory is that the *LOP* allele originated from the functional divergence of an ancestral locus shared by *LOP* and the alternative alleles (Tucker & Koltunow, 2009). Another theory is that the *LOP* allele originated in another species that was introduced into *Hieracium* through interspecific hybridization (Tucker & Koltunow, 2009).

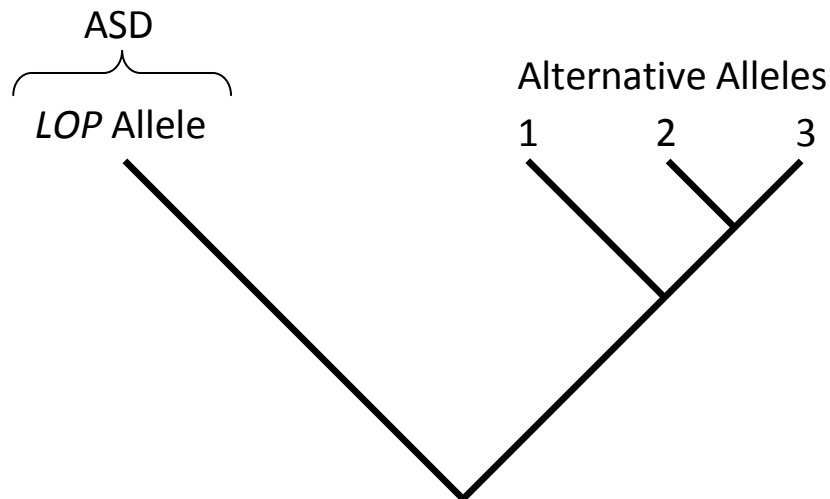


Figure 1.4. Simplified schematic tree showing the allelic sequence divergence (ASD) of the *LOP* allele in relation to the alternative alleles within the parthenogenesis candidate genes. Based on preliminary evidence (McGee, PFR Summer Studentship 2010-2011, unpublished data).

1.1.5. Research Hypothesis and Aims

Research Hypothesis

Introducing the apomixis trait into crop species is seen as a key future technology to improve global food production. Improving our knowledge of the genetic basis of this trait is an important step towards achieving this goal. While progress has been substantial, knowledge is lacking on the intricacies of the genetic control of the parthenogenesis step of apomixis. Parthenogenesis is thought to be controlled by the *LOP* in the *Hieracium* genus. In order to understand more about the *LOP* locus and the *LOP* alleles themselves, the following hypotheses were tested:

1. The alternative alleles of the candidate parthenogenesis genes act like alleles to the *LOP* alleles.
2. The *LOP* alleles display ASD from corresponding alternative alleles resulting from repressed recombination within *LOP*.
3. The *LOP* alleles and the alternative alleles originated from the same ancestral allele.

Aims and Research Summary

The first aim of this study was to determine whether the alternative copies of *LOP* genes were in fact alleles. This was done by generating allele segregation data of two *Hieracium praealtum* polyhaploid populations for the alleles of the four candidate genes. Genotyping was performed by detecting differences within microsatellites and SNPs between alleles present in individual polyhaploids. The genotyping data was also used to determine segregation of alleles.

The second aim of this study was to assess potential ASD of the *LOP* allele relative to the alternative alleles. This was done by aligning the complementary genetic regions of the *LOP* alleles with those of the alternative alleles and looking for large discernible differences. This also included identifying deletions, insertions, translocations and retrotransposons. In addition, the genotype data generated as part of the first aim was used to determine if recombination had occurred within *LOP*. The rate of recombination between each pair of alleles was then calculated based on the number of recombinant plants and their genetic composition.

The third aim of this study was to look for gene synteny between the *LOP* allele and alternative alleles, which would be an indication of a shared common ancestor. This was investigated by determining the level of sequence similarity between the *LOP* alleles and alternative alleles within the four candidate genes.

CHAPTER 2

GENOTYPING

2.1. Introduction

2.1.1. Genetic Basis of Apomixis

In order to introduce apomixis into crop species an understanding of the genetic basis of this trait is essential. Progress over the years has been substantial with the identification of a number of loci associated with apomixis within different apomictic species (Bicknell & Koltunow, 2004). Recent investigations also revealed that apomixis loci appear to be situated within zones of repressed meiotic recombination, as described in *Pennisetum* (Goel *et al.*, 2003) and *Paspalum* (Labombarda *et al.*, 2002). The lack of recombination in these zones hinders the use of conventional recombination-based linkage mapping to dissect the apomixis loci. With this in mind, a previous study used deletion mutagenesis to map apomixis loci in *Hieracium praealtum* (Catanach *et al.*, 2006). Two dominant independent genetic loci were identified, *LOSS OF APOMEIOSIS* (*LOA*; Okada *et al.*, 2007) and *LOSS OF PARTHENOGENESIS* (*LOP*; Catanach *et al.*, 2006) that control two of the key stages of apomixis, which are the avoidance of meiosis, or apomeiosis, and the avoidance of fertilisation known as parthenogenesis (see Figure 1.1 in Chapter 1; Bicknell & Catanach, 2004). Genetic sequences of the *LOP* locus of *Hieracium praealtum* were partially sequenced from a bacterial artificial chromosome (BAC) library of *Hieracium praealtum*. Two contigs were constructed, referred as *lop92* and *lop110*, totalling ~ 1.3 Mb of DNA sequence (see Figure 2.1) which is predicted to contain approximately 18 candidate genes (Dr. Andrew Catanach personal communication). In the current study, four candidate genes were selected for further investigation because they possess introns containing polymorphic microsatellites suitable for use as markers. These are Genes B, H, X and Y. Their location within the *LOP* locus is based on BAC sequencing as shown in Figure 2.1. BLASTn and BLASTx searches (www.ncbi.nlm.nih.gov) were performed on the sequences of these genes to assess their potential function. The Gene B coding sequence shows high similarity to that of the *Arabidopsis* annexin D3 protein, which is an essential component in calcium signalling (Ge *et al.*, 2007). The Gene H coding sequence shows high similarity to a helicase protein involved in chromosome stability (Pinter *et al.*, 2008) and the Gene

X coding sequence shows high similarity to the eukaryotic translation initiation factor 3 subunit E (eIF3e; Verlhac *et al.*, 1997). The Gene Y coding sequence shows high similarity to a transducin family protein which is expressed in megaspores and is essential for gametogenesis (Shi *et al.*, 2005). A greater understanding of these candidate genes is required to determine if they play an essential role in parthenogenesis.

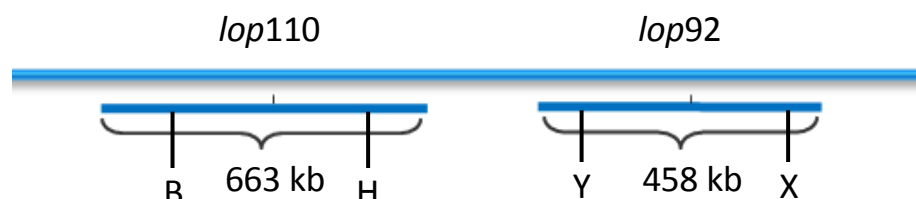


Figure 2.1. Physical map of the *LOP* locus showing the two *LOP* contigs (*lop110* and *lop92*), their size in kb and the relative positions of the four candidate genes within each contig (Dr. Andrew Catanach personal communication).

2.1.2. The Alternative Alleles of Candidate Parthenogenesis Genes

Hieracium praealtum was the model apomictic species used in the current study. It is a tetraploid (Chapman & Bicknell, 2000), therefore, each individual plant is expected to contain four alleles of each gene (Nielsen, 1980; Wolf *et al.*, 1990; Diwan *et al.*, 2000; Mahy *et al.*, 2000; Tiffin & Gaut, 2001). This was confirmed in *Hieracium praealtum* as part of my Plant & Food Research (PFR) summer studentship undertaken in 2010-2011 with the apomixis research team based in Lincoln, New Zealand. PCR products generated from sequencing Genes B, X and Y at the *LOP* locus were found to contain more than one variant that differed from the *LOP* sequence. These alternative homologous copies of *LOP* were hypothesised to represent recessive alleles or alternative alleles (see Figure 1.2a in Chapter 1; McGee, PFR Summer Studentship 2010-2011, unpublished data). Little was known about the characteristics of these alternative alleles, their genetic composition and if they are acting like alleles to the *LOP* copy of the gene. In this study this was tested by determining how the alternative alleles segregate with the *LOP* copy, in which segregation would infer that they are located on sister chromatids. The other possibility is that the alternative alleles undergo independent assortment relative to the *LOP* copy, which would indicate that *LOP* alleles and alternative alleles of these genes are paralogues and are not located on sister chromatids.

2.1.3. Polyhaploids

In order to understand more about the alternative alleles, this study set out to generate segregation data of the alleles of four candidate genes using two *Hieracium* polyhaploid populations. The production and expected genetic composition of a typical polyhaploid is outlined in section 1.1.4 in Chapter 1.

2.1.4. Genotyping the Alternative Alleles

In order to genotype the two existing *Hieracium praealtum* polyhaploid populations, new primers were designed using the new alternative allele sequence data generated in this study. The new sequence data was derived from BACs containing the alternative allele sequences of Genes B, X and Y which were identified from an existing BAC library of *Hieracium praealtum* maintained by the PFR apomixis research team based in Lincoln, New Zealand. These identified BACs underwent Roche 454 pyrosequencing. The new sequence data was then used to design primers to genotype each allele of the candidate genes. One class of polymorphism genotyped in this study were microsatellites, also known as simple sequence repeats (SSRs), which are repeat motifs typically of 1 - 6 base pairs (bp) in length. The number of repeats in a microsatellite can vary between individuals due to slippage errors caused during DNA replication (Li *et al.*, 2002). Variation in the number of repeats creates variation in the overall length of the amplified DNA fragment, which can be detected using high-resolution electrophoresis in a genetic analyser. In this study, microsatellites were used to distinguish the alleles of Genes B, H and X in a polyhaploid population.

As additional work, Gene Y was investigated to see if it behaves similar to Gene X, which it is known to be in linkage with (Dr. Andrew Catanach personal communication). However, SSR-based genotyping of Gene Y was not possible since the sequence data highlighted a lack of length polymorphisms. Therefore an alternative genotyping technique was selected, known as High Resolution Melt (HRM) which can detect single nucleotide polymorphisms (SNPs). HRM uses an optical detection system to accurately measure minute changes in the fluorescence-labelled genetic fragment as it was slowly heated (Distefano *et al.*, 2012). The shape and characteristics of the

generated melting curves were then associated with different SNP types, which had already been identified in the Gene Y sequence data. This genotyping technique was tested on the EMS mutant polyhaploid population.

2.1.5. Inheritance of Alleles

Genotype data was also used to reveal the nature of inheritance of the candidate genes and their alleles. Alleles of tetraploids typically undergo one of two patterns of segregation known as tetrasomic and disomic inheritance. With tetrasomic inheritance, all alleles are homologues (denoted as $A_1A_2A_3A_4$) and therefore each allele is equally likely to be inherited (Muller, 1914). This form of inheritance is commonly indicative of autotetraploids, the product of chromosome doubling in a diploid (Soltis *et al.*, 1993). Autopolyploidy is common among apomictic species, as reported in the apomicts *Paspalum rufus* (Quarín *et al.*, 1998) and *Tripsacum* (Grimanelli *et al.*, 1998). Similarly, autopolyploidy was detected in 10 different *Hieracium* species based on phylogenetic analysis of external transcribed spacers in the nuclear ribosomal DNA (Kraak *et al.*, 2013). In the case of genes at *LOP* in the tetraploid *Hieracium praealtum* (Chapman & Bicknell, 2000), tetrasomic inheritance would result in all polyhaploids inheriting the *LOP* allele, together with an equal chance of inheriting one of the three alternative alleles.

The other potential segregation pattern is disomic inheritance, which occurs when there are two homeologous sets of alleles in two homologous chromosomes (denoted as A_1A_2 or B_1B_2). In this case, homologues exclusively pair (e.g. B_1B_2 ; Winge, 1932). This type of inheritance is typically associated with allopolyploids, which arise from hybridisation between two different species (Soltis *et al.*, 1993; Ramsey & Schemske, 2002). Within *Hieracium*, disomic inheritance would result in all polyhaploids receiving the *LOP* allele and a skewed segregation of the alternative alleles, due to specific pairs of alleles behaving as homologues.

2.1.6. Recombination of Alleles

As found at apomixis loci in a number of apomictic species including *Pennisetum* (Goel *et al.*, 2003) and *Paspalum* (Labombarda *et al.*, 2002), the *LOP* locus was expected to exist within a zone of repressed recombination. The current study set out to investigate this characteristic by

determining if any polyhaploids in the populations underwent recombination within *LOP*. Recombination occurs during meiosis, and results in the exchange of genetic material via chromosomal crossover. Recombinant plants that arise can have different combinations of alleles than either parental plant. In the current study, recombinant plants were identified during the genotyping process and the number of these recombinant plants and their genetic composition were used to calculate the rate of recombination between the genes at *LOP*. In addition, some recombinant polyhaploids did not contain the *LOP* allele of certain genes. By assessing the genetic composition of these recombinant plants it was possible to determine which *LOP* genes were likely essential for parthenogenesis, based on which were absent or retained as a result of recombination.

2.1.7. Research Hypothesis

The first research hypothesis of this study was that the alternative alleles of the candidate parthenogenesis genes act like alleles to the *LOP* alleles. These alleles are also expected to undergo tetrasomic inheritance, which would be consistent with an autopolyploid origin of *Hieracium praealtum*. The second research hypothesis was that *LOP* exists within a zone of repressed recombination.

2.1.8. Aims of the Study

The first aim of this study was to generate allele segregation data of Genes B, H, X and Y within two *Hieracium praealtum* polyhaploid populations. Initially, alternative allele sequence data was generated by first screening an existing *Hieracium praealtum* BAC library for BACs containing alternative allele sequences of the four candidate genes. The identified BACs were then Roche 454 pyrosequenced. The new and existing sequence data was then used to design SSR-based or SNP-based primers that were used to genotype the two polyhaploid populations for each allele of the four candidate genes. The generated genotype data was used to determine the segregation of alleles. The second aim of the study was to determine the rate of recombination between these alleles. This was calculated based on the number of identified recombinant plants during genotyping as well as their genetic composition.

2.2. Materials and Methods

2.2.1. Plant Material

Genotyping Populations

Two polyhaploid populations were utilised in this study that were generated using different approaches by the PFR apomixis research team based in Lincoln, New Zealand, led by Dr. Ross Bicknell. Both consisted of the apomictic *Hieracium praealtum* Vill. ($2n = 3x + 8 = 35$), also referred to as R35, which was originally obtained from Dijon, France (Catanach *et al.*, 2006).

One of the polyhaploid populations was generated in 2008 from a single R35 parent plant. This plant carried a single copy of a construct containing the bacterial gene *codA* (cytosine deaminase A), which is unlinked to the *LOP* locus. *CodA* is lethal in the presence of 5-fluorocytosine, which is converted into toxic 5-fluorouracil (Dubeau *et al.*, 2009). Seeds were sown on medium containing 5-fluorocytosine; therefore normal apomicts that retain the construct were not recovered. In contrast, progeny that had undergone parthenogenesis and eliminated the construct were recovered. Using this method, a total of 178 polyhaploid seedlings were recovered and grown to maturity.

The other polyhaploid population was generated in 2011 from a single mutant R35 parent plant called EMS15C, derived from ethyl methanesulfonate (EMS) mutagenesis of *Hieracium praealtum* seeds (method described in Green & Krieg, 1961). This mutant is dysfunctional in apomeiosis but still capable of parthenogenesis, therefore it produces mostly polyhaploid progeny. It is similar to the *loaLOP* mutants generated by Koltunow *et al.*, (2009). A total of 119 polyhaploids from EMS15C were grown to maturity in a greenhouse with 16 hour supplementary lighting, as described by Bicknell *et al.* (2000).

DNA Extractions

Four young leaves were harvested from each plant and placed in a test tube containing glass beads and snap-frozen using liquid nitrogen. Each tube was vortexed for 12 seconds using a Silamat S6 shaker (Ivoclar Vivadent AG, Schaan, Liechtenstein) to grind the plant material to a fine powder. DNA was extracted using the reagents and manufacturer's protocol in the DNeasy[®] Plant Mini Kit

(Qiagen, Hilden, Germany). The resulting genomic DNA was eluted in 140 µl of Milli-Q water. Genomic DNA was quantified using a NanoVue Spectrophotometer (General Electric Healthcare, NJ, USA). DNA extractions were repeated on samples with low quality DNA, which is known to promote genotyping errors (Pompanon *et al.*, 2005), allelic dropout and false alleles (Taberlet *et al.*, 1996). DNA was diluted with Milli-Q water to volumes of 100 µl at 20 ng/µl and stored at -20 °C.

2.2.2. BAC Isolation and Sequencing

BAC Library Screening Using PCR

The BAC library screened in this study was generated from a single *Hieracium praealtum* (R35) apomict plant previously developed by Dr. Ross Bicknell (unpublished data). PCR screening was performed on this BAC library in order to identify the BAC colonies containing alternative allele sequences. This was done using primers designed to amplify Genes A, B, Z, and X (see Table A.1) located within the *LOP* contig. PCR amplification was performed on a Mastercycler Ep Gradient S Thermocycler (Eppendorf, Hamburg, Germany). The same PCR machine was used in all subsequent PCR reactions in this study. PCR reactions were performed in a total volume of 15 µl, composing of forward and reverse primers (5 µM), 2 µl of template DNA at 40 ng/µl, dNTPs (2 mM), 1 × Buffer, magnesium chloride (1.5 mM), and 1 U Taq DNA polymerase (Fisher Biotec, Wembley, Australia). The PCR protocol utilised started with 94 °C for 3 minutes, followed by 40 cycles of 94 °C for 30 seconds, 60 °C for 30 seconds, 72 °C for 1 minute, with a final 72 °C extension cycle for 7 minutes. Aliquots of 5 µl of PCR product were electrophoresed on a 1% agarose gel containing 8 µl (0.008%) of GelRed[®] Nucleic Acid Gel Stain (Biotium, Hayward, USA) alongside a 1 kb Plus DNA Ladder (600 ng; Invitrogen, Carlsbad, USA). Gels were visualised under UV light using a Molecular Imager[®] Gel Doc[™] XR System with the Quantity One v4.6.5 analysis software (Bio-Rad, Hercules, USA). Initially, PCR screening was performed on an 84 well superpool plate, which consists of pools of seven 96 well pool plates. A well that contains a positive PCR sample of the superpool plate corresponded to a specific column within one of the pool plates. The identified column of the specific pool plate was then PCR screened to identify the positive well of the pool plate. The position of the positive PCR sample within the pool

plate corresponded directly to a single 384 well BAC plate within the BAC library that contained the positive clone. This plate was then screened as described below.

BAC Culturing

To identify specific bacterial colonies containing the alternative allele sequence, each identified BAC plate was stamped in triplicate on LB plates containing 12.5 µg/ml chloramphenicol, using a 384-pin manual plate replicator (V&P Scientific, San Diego, USA). The plates were grown overnight at 37 °C in an Ecotron Incubator Shaker (Infors HT, Bottmingen, Switzerland). Each row of BAC colony on the LB plate was then individually sampled by drawing a pipette tip through each row. This small sample of cells from each row was transferred into PCR wells containing 50 µl of Milli-Q water. To lyse the cells, the PCR plate was heated to 96 °C for 3 minutes. The lysate then cleared by centrifugation at 4000 x g for two minutes on a Centrifuge 5810 (Eppendorf, Hamburg, Germany). Aliquots of 1 µl from the previous centrifugation step were used as template for new PCRs. These PCRs utilised the same PCR protocol and 60 °C annealing temperature as previously used to screen the BAC library. Individual colonies of the row that had a positive PCR product were then screened by transferred a small sample of each colony into individual PCR wells containing 50 µl of Milli-Q water using a pipette tip. The PCR and gel electrophoresis steps were repeated on the individual colonies, as outlined above, to identify positive BAC colonies. These were then re-streaked using a sterilised inoculating loop onto a new LB plate containing 12.5 µg/ml chloramphenicol for later colony purification and BAC DNA preparation.

BAC Confirmation

Identified BAC colonies were confirmed to contain alternative allele sequences by Sanger sequencing of the PCR products. Amplified products from single colonies were purified using the EXO SAP technique (Werle *et al.*, 1994) in 10 µl reactions containing 5 µl of the previous PCR product from the BAC culturing step, as well as 20 U/µl Exonuclease I (Fermentas, Thermo-Fisher Scientific, Waltham, USA), and 1 U/µl of Shrimp Alkaline Phosphatase (Roche Applied Science, Penzberg, Germany). Purification reactions were incubated at 37 °C for 30 minutes followed by deactivation for 15 minutes at 80 °C and cooling to 4 °C. Sequencing reactions contained 2 µl of the PCR cleanup product, forward or reverse primers (5 µM), approximately 0.5 µl of Big Dye®

v3.1 (Applied Biosystems, Foster City, USA), 1 × Sequencing Buffer (Applied Biosystems), in a total volume of 10 µl. The sequencing reactions underwent thermocycling as follows: 1 minute at 96 °C, followed by 50 cycles of 96 °C for 10 seconds, 50 °C for 5 seconds, and 60 °C for 4 minutes concluding with a 4 °C hold step. Upon completion, the sequencing reactions were precipitated by adding 13.75 µl of Milli-Q water, 1.25 µl of 500 mM EDTA at pH 8.0 and 60 µl of 100% ethanol into each well. The PCR plate was then inverted four times, followed by 15 minutes of room temperature incubation out of the light. The PCR plate was then centrifuged for 30 minutes at 3000 x g at 4 °C on a Centrifuge 5810R. Once completed, the PCR plate was inverted and centrifuged at 185 x g for approximately 30 seconds to remove the existing ethanol. 60 µl of 70% ethanol was then added to each well and the plate was centrifuged at 3000 x g for another 15 minutes at 4 °C. The ethanol was once again removed by inverting and then centrifuging the PCR plate at 185 x g for 1 minute. The PCR plate was then air dried at 37 °C in an Ecotron Incubator Shaker (Infors HT). On completion, 10 µl of Hi-Di™ formamide (Applied Biosystems) was added to each well and the PCR plate was briefly centrifuged. Extension products were denatured by heating at 96 °C for 3 minutes, followed by a brief centrifugation, and then electrophoresed on a 3130xl Genetic Analyser (Applied Biosystems). The generated sequence data was compared with existing sequences using pair-wise alignment in Geneious 5.5 (Biomatters Ltd).

BAC Preparation and 454 Pyrosequencing

In preparation for Roche 454 pyrosequencing, BAC DNA was isolated following the ‘User-Developed Protocol’ (QP01, Qiagen) which utilises the Plasmid Midi Kit (Qiagen). This and the following step was undertaken by staff of the PFR apomixis research team. Individual BACs were uniquely labelled using a multiplex identifier before pooling into a 16-BAC pool. Axseq Technologies (Seoul, South Korea) sequenced the BACs using the 454 GS FLX Plus platform (Roche Diagnostics, Basel, Switzerland). BAC sequence data was made available for download from Axseq Technologies.

BAC Sequence Assembly

BACs were assembled using gsAssembler v2.6 (454 Life Sciences, Branford, USA). Reads were trimmed for BAC vector (pAgiBAC1) using the UniVec database (<ftp://ftp.ncbi.nih.gov/pub/UniVec>) and screened for *E. coli* contamination using the sequence of

the DH10B strain. Default parameters were used for the initial assembly process. In order to construct a continuous draft sequence, assembled contigs were ordered based on the contigs that overlapped and those that spanned the same allele. The *LOP* candidate genes were then queried against the draft sequences using nucleotide BLAST (Altschul *et al.*, 1997) and the output was used for annotation in Geneious 5.6 (Biomatters Ltd).

2.2.3. Genotyping - Fragment Length Polymorphism

SSR Primer Design

Using Primer3web (v.4.0.0, Rozen & Skaletsky, 2000) and PrimerQuest[®] (Integrated DNA Technologies, Coralville, USA), primers were designed to anneal to exons which flanked introns containing SSRs within the candidate genes. Primer generated by both programs were selected based on specific characteristics outlined in Dieffenbach *et al.* (1993). The primers were also checked for self annealing using OligoAnalyzer 3.1[®] (Integrated DNA Technologies) which calculates the ΔG (Gibbs free energy). Ideally, ΔG values were above -9 kcal/mol which is known to reduce the formation of inhibitory self-dimers (Integrated DNA Technologies). Primers that fulfilled these criteria were then supplied by Integrated DNA Technologies.

Fragment Length Polymorphism Genotyping

Prior to genotyping the two populations, 11 SSR primers listed in Table A.2 were tested on a sub-population of samples consisting of randomly selected polyhaploids, the *Hieracium praealtum* parent (R35), the EMS15C parent, a water control and four *LOP*-minus mutants (R35 mutants 138, 143, 156 and 164). The *LOP*-minus mutants were part of a set generated by Catanach *et al.* (2006). The PCR reactions consisted of forward and reverse primers (5 μ M), approximately 40 ng/ μ l of template DNA, dNTPs (2 mM), 1 \times manufacturers supplied buffer, magnesium chloride (1.5 mM), 1 U Taq DNA polymerase, 20 μ M of fluorescence TAMRA-dUTP (Molecular Probes, Eugene, USA) in a total volume of 10 μ l. The PCR cycling parameters followed the same 60 °C annealing protocol previously utilised to screen the BAC library (section 2.2.2). The precipitation step was performed by adding 6.875 μ l of Milli-Q water, 0.625 μ l of 500 mM EDTA at pH 8.0 and 30 μ l of 100% ethanol to the amplified PCR mix. The subsequent PCR precipitation steps were as described in section 2.2.2. Prior to fragment analysis, 0.075 μ l of GeneScan[™] 600 LIZ[®] Size

Ladder and 10 µl of Hi-Di™ formamide (Applied Biosystems) were added to the precipitated PCR product. The PCR plate was spun at 1000 x g for 30 seconds and then denatured by heating at 94 °C for 3 minutes, followed by brief centrifugation, and then underwent electrophoresis on a 3130xl Genetic Analyser. The fragment data was visualised using GeneMarker® v2.2.0 (SoftGenetics, Pennsylvania, USA). The most suitable primers for genotyping the two populations required three features. First, one of the peaks in the electropherogram must be associated with the amplified microsatellite. Second, there must be variation in the size and pattern of the peaks associated with the microsatellites, which can therefore be used to genotype the alternative alleles. Third, a peak associated with the *LOP* allele needed to be present in all the polyhaploids and parents but absent in the *LOP*-minus mutants. The primers which met all three requirements were then used to genotype the two populations for the alleles of Gene X, B, and H. Ten plants were found to contain the parental genotype and were therefore removed from further analysis, leaving a total of 287 plants. To increase the precision of the genotyping process, each fragment result was rescored a minimum of three times. In addition, the genotype process was repeated on individuals with unclear genotyping results.

2.2.4. Genotyping – High Resolution Melt

HRM Primer Design

HRM primers were designed to flank SNPs downstream of Gene Y using the LightScanner Primer Design Software 1.0 (Idaho Technology, Salt Lake City, USA). The primers generated by this program were selected based on a number of ideal characteristics outlined in Dieffenbach *et al.* (1993). The HRM primers which fulfilled these criteria were then supplied by Integrated DNA Technologies.

HRM Genotyping

Initially six HRM primers listed in Table A.3 were PCR tested at different annealing temperatures on the same sub-population previously amplified with the SSR primers (section 2.2.3). Each PCR reaction contained forward and reverse primers (5 µM), approximately 40 ng/µl of template DNA and HOT FIREPol® EvaGreen® HRM Mix (No ROX; Solis BioDyne, Tartu, Estonia) in a total

volume of 8 μ l. The PCR mix was dispensed into a 96-well Blackwell PCR plate (Bio-Rad), followed by adding 20 μ l of mineral oil (Sigma-Aldrich, St Louis, USA). Prior to PCR, optically clear sealing tape was placed over the plate, which was then spun at 1600 x g for 30 seconds. The PCR protocol started with 95 °C denaturing for 15 minutes, followed by seven cycles of a touchdown PCR starting with 30 seconds at 94 °C, with a variable annealing temperature (62 - 74 °C) for 30 seconds, which decreased by 1 °C per cycle, followed by a 1 minute extension cycle at 72 °C. The touchdown step was followed by 45 cycles of 94 °C for 30 seconds, 53 °C for 30 seconds, finishing with a 72 °C extension cycle for 15 seconds. Once the PCR reaction was completed the plate was spun at 1000 x g for 30 seconds and scanned immediately on the LightScanner[®] System (Idaho Technologies). The HRM data was visualised as melting curves using the LightScanner[®] Software using Call-IT 2.0 (Idaho Technologies). The most suitable HRM primers, and the optimal annealing temperature for that primer, were selected based on which combination produced the highest level of fluorescence and the clearest and most repeatable melting curves, which could also be linked to different alternative alleles. The HRM primers that met these criteria were selected to genotype the EMS15C population. To increase the precision of the data each individual was genotyped in triplicate and rescored a minimum of three times. Any unclear genotyping results were repeated.

The EMS15C population was also genotyped for the *LOP* allele of Gene Y using PM83, a Gene Y *LOP* allele-specific primer combination listed in Table A.2, using the same PCR mix and 60 °C annealing temperature PCR protocol used to screen the BAC library (section 2.2.2). The PCR products also underwent gel electrophoresis using the same technique utilised in the primer screening of the BAC library (section 2.2.2). Samples containing positive gel band of a specific size signified the presence of the *LOP* allele of Gene Y.

2.2.5. Allele Segregation and Recombination

Segregation of Alleles

The genotyping data was compiled into genotype frequency tables, which were used to determine alleles in linkage disequilibrium. The statistical significance of these relationships was determined

by calculating the probability (p) value using the inbuilt Fisher's exact test in Minitab 16 (Minitab Inc.).

Identification and Confirmation of Recombinants

During the course of genotyping, recombinant plants were detected that did not contain the peak or the gel band associated with the *LOP* allele. Recombinant plants from the 2008 population were confirmed by repeating the fragment and HRM genotyping process using the existing DNA. For the recombinants from the 2011 population, DNA from these recombinants was re-extracted using the same procedure as previously described section 2.2.1. The genotyping process was then repeated on the re-extracted DNA. Recombinants were also confirmed by testing additional *LOP* sequence-specific primers listed in Table A.4 using the same PCR mix, PCR protocol and gel electrophoresis protocols utilised to screen the BAC library (section 2.2.1).

An additional confirmation step was conducted using flow cytometry to confirm the polyhaploid status of the recombinant plants. First, 1 cm² of leaf tissue from each of the two reference standards, EMS15C parent and White Clover (*Trifolium repens*; an internal standard) were chopped up and submersed in Otto I buffer (100 ml deionised water, 2.1 g citric acid, 0.5 g Tween-20; Otto, 1990). The mixture was filtered through a Cell Trix Filter (Partec GmbH, Münster, Germany) into a collection tube. The collection tube was assayed on the PAII Ploidy Analyser flow cytometer (Partec) to detect UV intensity using the inbuilt FloMax[®] program (Partec). This process was then repeated on the recombinant plants. The level of ploidy of the recombinant plants was determined by comparing the intensity of the UV peaks of the recombinant plants with the reference peaks. To increase the precision of the flow cytometry data, each plant was tested three times. All confirmed recombinant plants were inputted into a finalised genotype dataset.

Recombination Rates

The finalised genotype dataset was used to calculate the rate of recombination between the alleles of two genes, based on the number of recombinant plants with only one of these alleles divided by the total number of polyhaploids containing one or both alleles. This process was completed for all possible combination of alleles between the four candidate genes.

2.3. Results

2.3.1. BAC Sequence Data

Identifying the Alternative Allele BAC colonies

The superpool plate of the BAC library was screened using SSR-based primer mixes (PM) 34, 61, 79 and 80 listed in Table A.1, which amplified sequences within Genes A, B, X and Z. 22 BAC plates were identified that contained alternative allele sequences (Figure 2.2). Each row and then each colony of an identified row was PCR tested in order to identify the specific BAC colonies containing alternative allele sequences (Figure 2.3). In total 11 BAC colonies were identified, of which five contained Gene A and Gene B sequences, three contained Gene X sequences, and the remaining three contained Gene Z sequences.

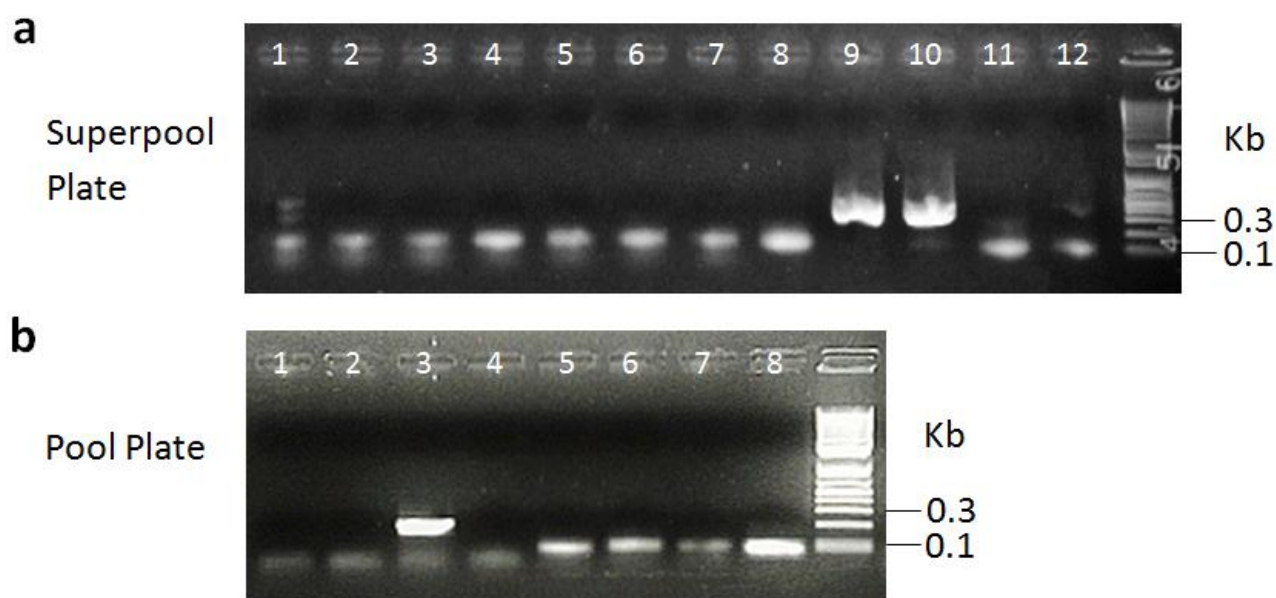


Figure 2.2. Identification of BAC plates containing the alternative allele sequences using gel electrophoresis of PCR products. **a.** Each row of the superpool plate was PCR tested. In this gel, lanes 9 and 10 contained the positive gel band used to identify a specific column within a pool plate. **b.** The identified column of a pool plate was then PCR tested. In this gel, the positive band in lane 3 was used to determine the BAC plate containing the alternative allele sequences. The size standards (far right lane of each gel) are 1 kb Plus DNA ladder.

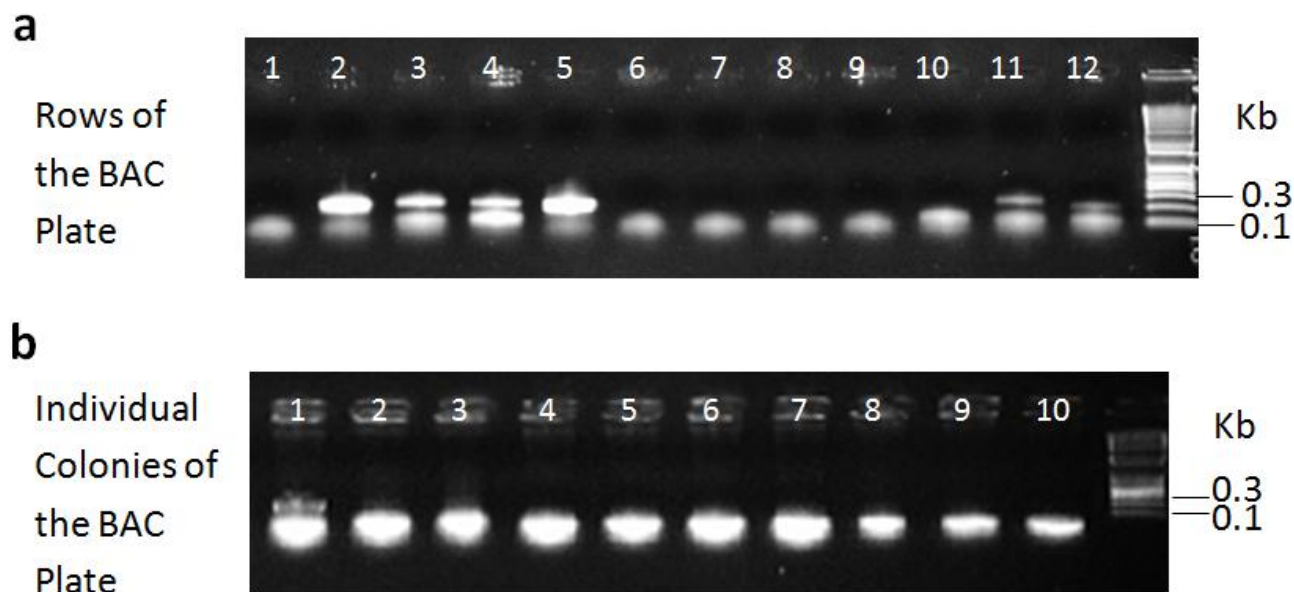


Figure 2.3. Identification of positive BAC colonies using gel electrophoresis of PCR products. **a.** Each row of the cultured BAC plate was PCR tested. In this gel, lanes 2 - 5 contained the positive gel bands used to identify the rows containing the BAC colony of interest. **b.** Each BAC colony within the identified rows was then PCR tested. In this gel, the positive band in lane 1 identified the BAC colony containing the alternative allele sequence. The size standards (far right lane of each gel) are 1 kb Plus DNA ladder.

Confirming the Alternative Allele BACs

Sanger sequencing of the BAC DNA confirmed that the identified BAC colonies contained alternative allele sequences and not *LOP* sequences (Figure 2.4 & 2.5). The alternative allele sequences were classified into different groups based on their similarity to the previously generated alternative allele sequences identified by the PFR apomixis research team (Dr Andrew Catanach, personal communication). As shown in Figure 2.4, three BAC colonies contained alternative allele sequences which matched the alternative allele 1 grouping of Gene X. Three other BAC colonies contained sequences that matched the alternative allele 1 grouping of Gene Z (Figure 2.4). Gene B alternative allele 1 was identified in a single BAC colony, while alternative allele 2 of Gene B was identified in four other BAC colonies (Figure 2.5).

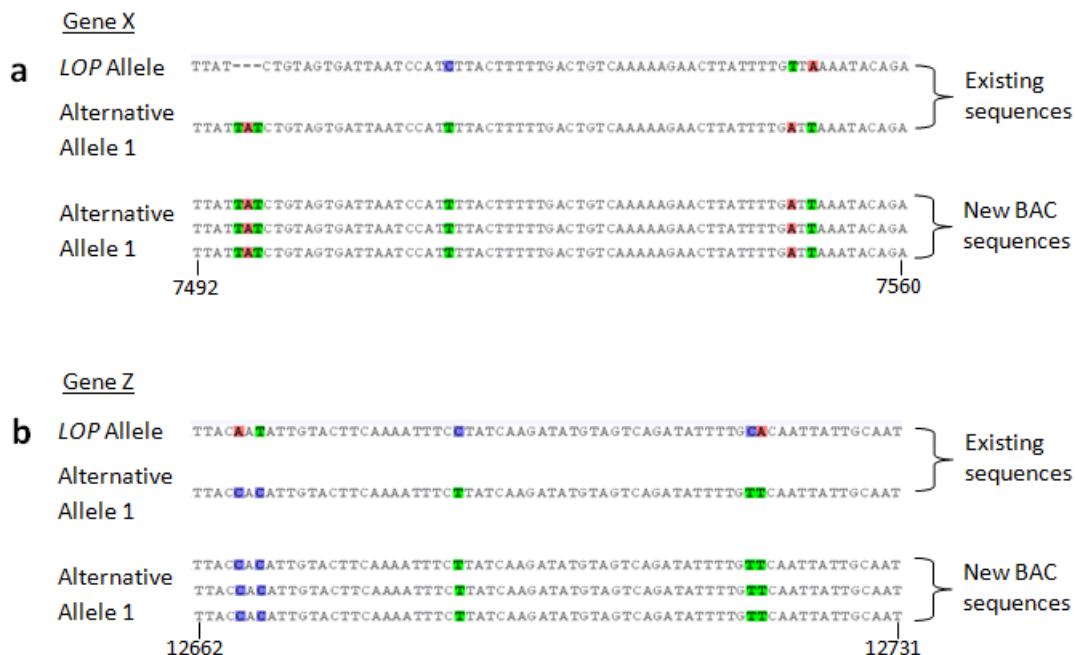


Figure 2.4. DNA sequence alignments of amplified regions of Genes X and Z, of previously known *LOP* and alternative allele sequences, with PCR sequences from identified BACs. **a.** The BAC sequence data derived from three BAC colonies aligns with the previously known sequence of alternative allele 1 of Gene X. **b.** Similarly, the BAC sequence data derived from the other three BAC colonies aligns with the previously known sequence of alternative allele 1 of Gene Z. The relative position in base pairs along the *lop92* contig is noted under each alignment.



Figure 2.5. DNA sequence alignments of amplified regions of Gene B of previously known *LOP* and alternative allele sequences, with sequences of PCR products from identified BACs. The BAC sequence data derived from five BAC colonies aligns with the previously known sequence of alternative allele 2 of Gene B, and one other colony aligns with alternative allele 1 of Gene B. The relative position in base pairs along the *lop110* contig is noted under each alignment.

454 Sequencing of the Alternative Alleles

Eight BACs were sent to Axseq Technologies for Roche 454 pyrosequencing. The 454 pyrosequencing resulted in 277 027 reads giving a total of 151 400 031 bp of DNA sequence from all eight BACs combined (Table 2.1).

Table 2.1. Read count and sequence statistics from BAC *BAC Sequence Assembly* sequencing.

BAC No.	Gene	Number of reads	Total DNA sequence (bp)
152	Z	14 451	7 732 994
160	A & B	47 675	25 450 375
308	Z	15 194	8 227 024
326	X	45 436	23 999 124
344	X	31 998	17 256 243
353	A & B	21 394	12 193 568
393	A & B	66 425	37 062 568
430	A, B, Z	34 454	19 478 135
	Total	277 027	151 400 031

Initially, gsAssembler v2.6 was used to trim the raw 454 reads for BAC vector and *E. coli* contamination. The remaining 454 reads were assembled with gsAssembler v2.6 using the default assembly parameters (minimum read length of 20 bp, 40 bp minimum overlap and 90% minimum overlap identify). The first assembly resulted in an average N50 value of 29.2 kb for all BACs (Table 2.2). On average 13 contigs were constructed from each BAC, with an average length of 12

363 bp (Table 2.2). To further improve the assembly statistics, the second assembly was conducted using gsAssembler v.2.7 with altered assembly parameters where the minimum contig and scaffold length were set to 200 bp. This resulted in an improvement in the average N50 value to 31.1 kb, with a marked increase in the average contig size to 14 699 bp, and a decrease in the average number of contigs constructed per BAC to 12 (Table 2.2). The third assembly step was performed by Dr. Andrew Catanach using gsAssembler 2.7. In this step, assembled contigs were ordered based on contigs of overlapping BACs, either of the same alternative allele or of the highly homologous allele. Subsequently, three continuous draft sequences were generated for three separate alternative alleles (Table 2.2) that were subsequently annotated with the *LOP* candidate genes.

Table 2.2. The BAC assembly progress using gsAssembler v2.6 & v2.7 and associated statistics.

Assembly Step	BAC No.	Number of Contigs	N50	Largest Contig (bp)	Average Contig Size (bp)
1	152	11	39 976	51 454	14 581
	160	22	27 858	31 887	6 813
	308	14	15 937	35 517	10 595
	326	13	38 336	71 564	14 979
	344	11	30 531	48 064	11 850
	353	5	38 677	48 489	10 995
	393	9	26 588	80 192	20 400
	430	19	15 924	35 573	8 693
	Average	13	29 228	50 343	12 363
2	152	11	40 013	51 466	14 621
	160	14	28 596	32 029	10 568
	308	13	15 938	35 552	11 426
	326	11	43 018	71 572	17 848
	344	5	40 052	48 176	25 996
	353	14	38 670	48 745	11 879
	393	11	26 847	80 192	14 958
	430	16	15 967	35 587	10 295
	Average	12	31 138	50 415	14 699
3	152	39	40 033	70 518	7 548
	160				
	308				
	430				
	326	18	37 251	54 286	15 858
	344				
353	15	48 280	80 191	14 940	
393					

2.3.2. Genotyping - Fragment Length Polymorphism

Primer Design

Seven new primers were designed, PM81, 82, 84-88 listed in Table A.2, to amplify microsatellites identified when comparing the newly assembled alternative allele sequences with the known *LOP* sequences.

Fragment Length Polymorphism Genotyping

Sub-Population Testing

Initially, 11 primer combinations consisting of both the existing primers used by the PFR apomixis research team and of newly designed primers were selected for sub-population testing. PM1 for Gene X, PM34 for Gene B and PM81 for Gene H (Table A.2) were found to be most suitable for genotyping the two populations. Their suitability was based on the criteria detailed in the methods, such as the clear association between different fragment peaks and different microsatellites, and an identifiable peak associated with the *LOP* allele.

PM1 was used to generate the fragment length profiles for the three alternative alleles of Gene X (Figure 2.6a). Allele 1 of Gene X was distinguished by a 391 bp peak, allele 2 by a 379 bp peak, and allele 3 by a 383 bp peak (Figure 2.6a). The 346 bp peak was found to be associated with the *LOP* allele peak since it was present in all polyhaploids and both parents (R35 & EMS15C) but absent from the *LOP*-minus mutants 138, 143, 156 and 164 (Figure 2.6a). It was also possible to confirm these allele sizes by linking the size of the fragment peaks directly to the sequence data. The size of the associated fragment peak of each allele could be correlated directly to the length predicted by previous Sanger sequence data (Figure 2.6b).

PM34 was used to generate the fragment length profiles of the three alternative alleles of Gene B. Allele 1 was characterised by a 248 bp peak, allele 2 by a 243 bp peak, and allele 3 also had a 243 bp peak but possessed a different fragment length profile from that of allele 2 (Figure 2.7a). The *LOP* allele of Gene B was found to be associated with the 259 bp peak which was present in all polyhaploids and parents, but absent from the *LOP*-minus mutants 138, 143, 156 and 164 (Figure 2.7a). It was also possible to correlate the size of the fragment peaks of each allele to the length predicted by previous Sanger sequence data (Figure 2.7b).

The fragment length profiles for Gene H, generated using PM81, indicated the presence of four alleles (Figure 2.8a). Allele 1 was characterised by a 323 bp peak and allele 2 was characterized as containing both a 323 bp and a 328 bp peak (Figure 2.8a). Allele 3 contained a single 328 bp peak and allele 4 was likely a null allele since no distinguishing peak was PCR amplified by PM81 (Figure 2.8a). The *LOP* allele was associated with the 333 bp peak (Figure 2.8a). The size of the *LOP* peak could be correlated directly to the length predicted by the known *LOPH* sequence

(Figure 2.8b). Sequence data of the alternative alleles of Gene H was unavailable in this study; therefore, it was not possible to validate the peaks of the alternative alleles. However, based on the known *LOPH* sequence, it was possible to infer the alternative allele sequence data. In the absence of one or both microsatellites, (A)₇ and (A)₄, the length of the inferred sequences accurately matched the fragment peak size observed with alternative alleles 1 – 3.

Population Genotyping

Gene X

Genotyping the two populations for the alleles of Gene X revealed that around 80% of the polyhaploids had either alternative allele 1 (117 plants), which was referred to as X1, or alternative allele 2 (124 plants), which was referred to as X2 (Figure 2.9). The least common allele was alternative allele 3 (or X3), which was only detected in 35 plants out of a total of 287. In addition, contrary to the repression of recombination at *LOP* predicted in the hypothesis, eleven recombinant plants were detected, which did not contain the *LOP* allele of Gene X, which is referred to as *LOPX* (Figure 2.9).

Gene B

The three alternative alleles of Gene B were distributed in a similar pattern to those observed in Gene X, where approximately 80% of the population was found to contain either the B1 (120 plants) or the B2 (128 plants) alleles (Figure 2.9). Similarly, the B3 allele of Gene B was the least common in the polyhaploid populations occurring in only 38 plants, which was analogous to the distribution of the X3 allele (Figure 2.9). It is interesting to note that no plants were found that did not contain the *LOP* allele of Gene B, which is referred to as *LOPB*. This was in contrast to the 11 recombinants detected which lacked *LOPX* (Figure 2.9).

Gene H

The segregation of the Gene H alleles differed from that of Genes X and B. One major difference was that the H1 allele was represented in over 60% (179 plants) of the population (Figure 2.9), which contrasted with the more even distribution of the X1, X2 and B1, B2 alleles. Another major difference was the presence of four detectable alternative alleles, as opposed to three in Genes X and B. Of these four Gene H alternative alleles, three appeared infrequently. This included the H2

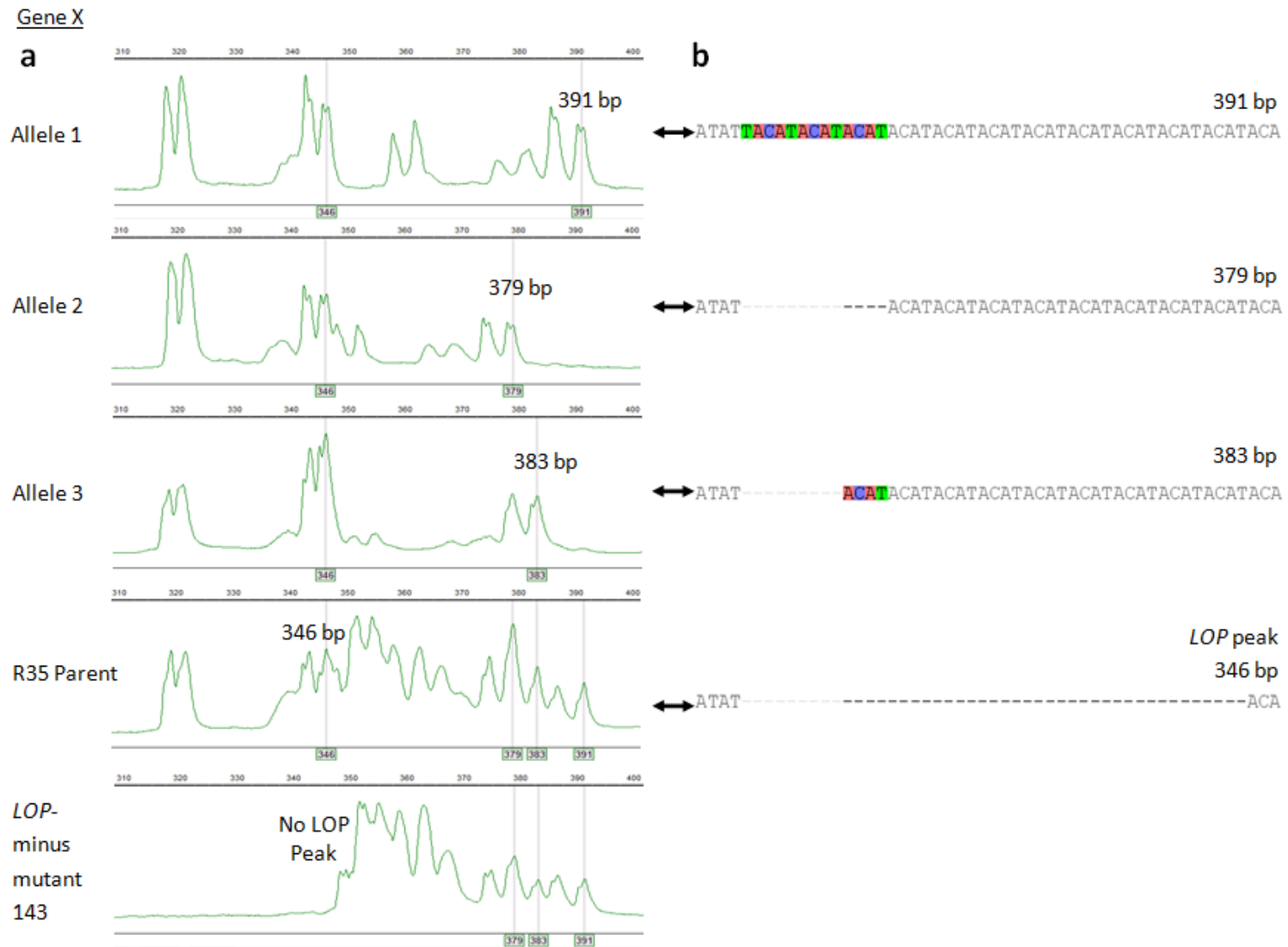


Figure 2.6. Fragment length polymorphism genotyping of Gene X. **a.** Polyhaploid-derived fragment profiles showing sizes in base pairs (bp) of distinguishing peaks of the three alternative alleles of Gene X identified using PM1. The 346 bp peak was associated with the *LOP* allele, as it was absent in the *LOP*-minus mutant 143. **b.** Previously-generated sequence data corresponds with the size of the distinguishing peak for each allele.

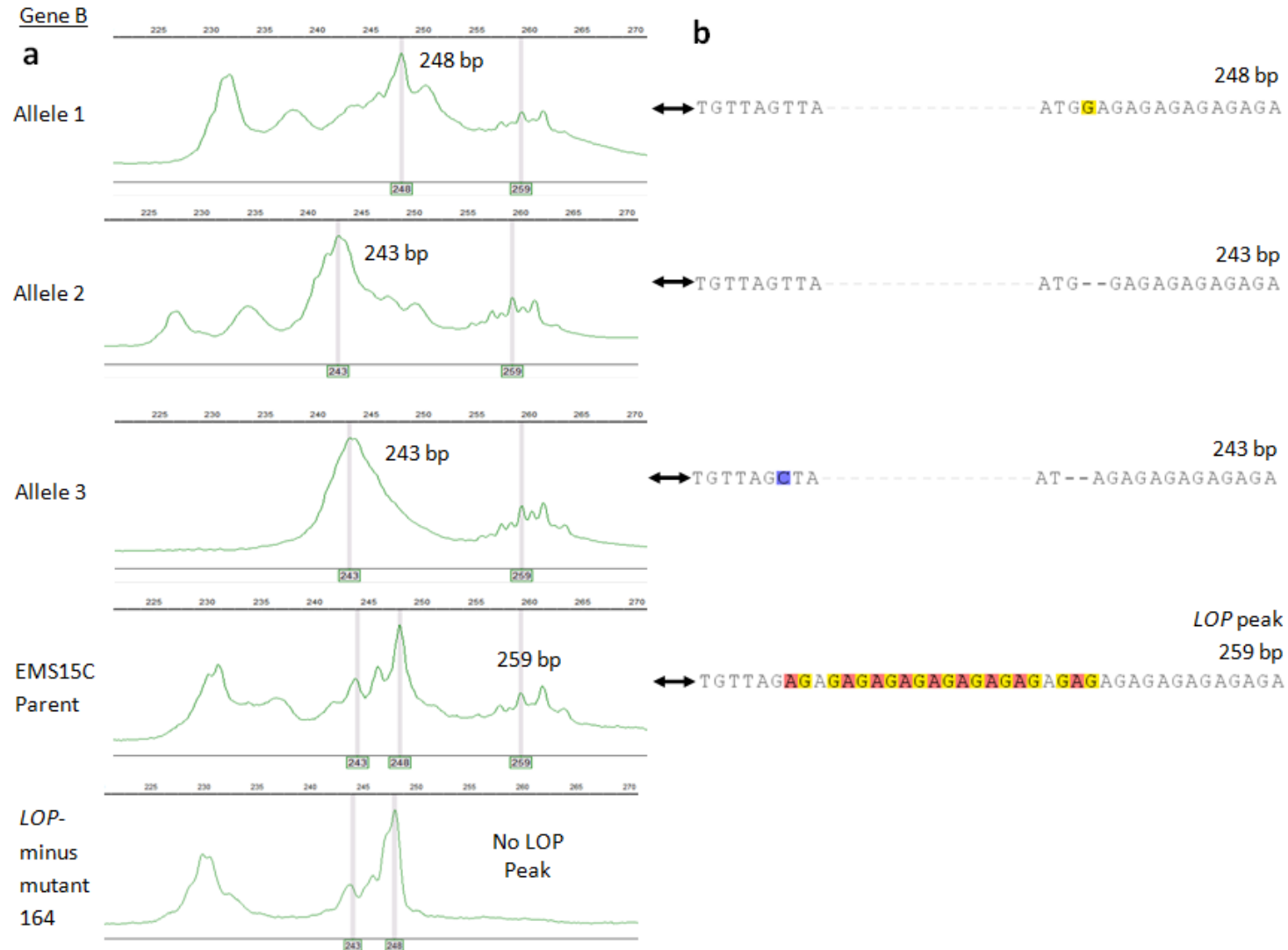


Figure 2.7. Fragment length polymorphism genotyping of Gene B. **a.** Polyhaploid-derived fragment profiles showing sizes in base pairs (bp) of distinguishing peaks of the three alternative alleles of Gene B identified using PM34. The 259 bp peak was associated with the *LOP* allele, as it was absent in the *LOP*-minus mutant 164. **b.** Previously-generated sequence data corresponds with the size of the distinguishing peak for each allele.

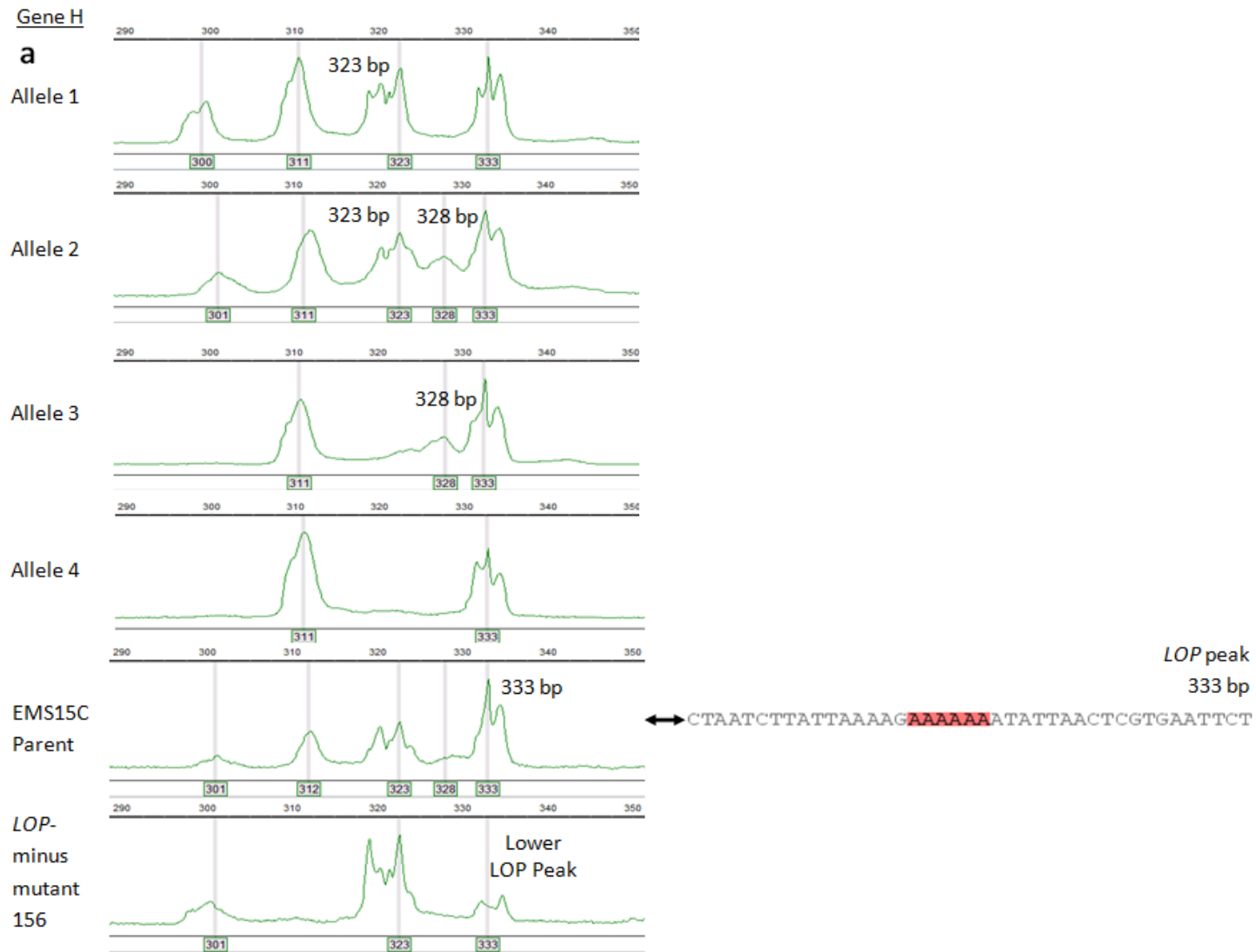


Figure 2.8. Fragment length polymorphism genotyping of Gene H. **a.** Polyhaploid-derived fragment profiles showing sizes in base pairs of distinguishing peaks (bp) of the four alternative alleles of Gene H identified using PM81. Allele 4 is consistent with a null allele, which was not amplified by PM81. The 333 bp peak was associated with the *LOP* allele since it was consistently lower in the *LOP*-minus mutant 156. **b.** The previously generated *LOP* sequence data corresponds to the size of the distinguishing *LOP* peak.

allele, found in 47 plants, the H3 allele found in 18 plants and the H4 allele found in 24 plants (Figure 2.9). Based on the fragment profile, the H4 allele was likely a null allele. In addition, only one recombinant plant was detected that did not contain the *LOP* allele of Gene H, which is referred to as *LOPH*.

Total Polyhaploid Population - #287 plants

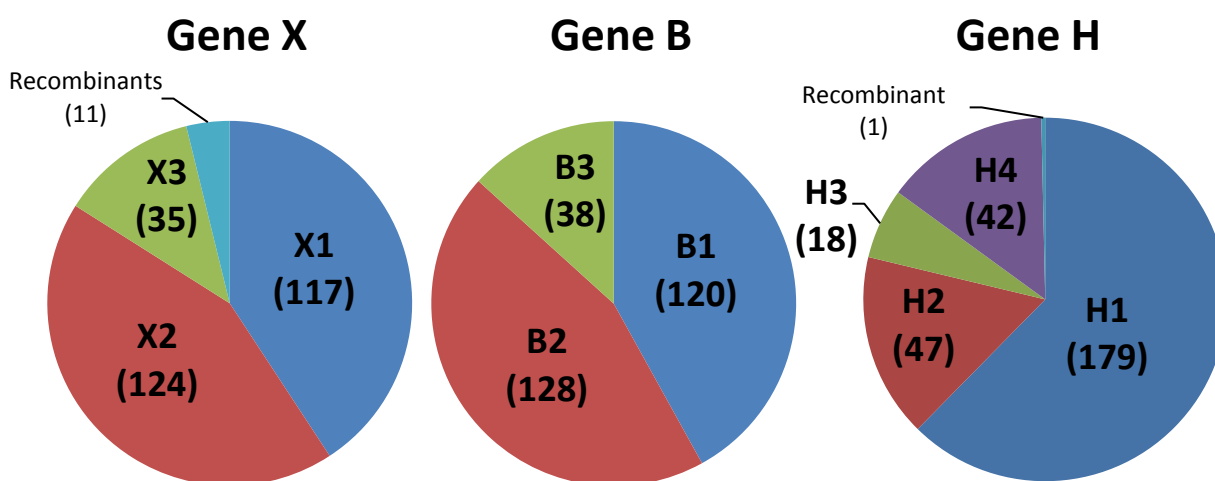


Figure 2.9. Distribution of the alternative alleles of Genes X, B and H in the two *Hieracium praealtum* polyhaploid populations generated through fragment length polymorphism genotyping. Alleles are abbreviated. For example, ‘X1’ refers to the alternative allele 1 of Gene X, etc. The number of plants with a given allele is shown in brackets. Plants that did not contain an *LOP* allele were classified as recombinant plants.

2.3.3. Genotyping – High Resolution Melt

For HRM genotyping, two new alternative allele-specific HRM primers, PM93 & PM94 listed in Table A.3, were designed around a SNP located 300 bp downstream of Gene Y.

HRM Genotyping

Sub-Population Testing

In total, six HRM primer combinations were selected for sub-population testing at a range of annealing temperatures. The primers were from two existing HRM primers used by the PFR apomixis research team and four newly designed primers (Table A.3). PM92 was found to be the most suitable primer combination for whole population genotyping (Table A.3) because of the ease

with which each plant could be classified into different alleles based on the shape of the melting curve profile (Figure 2.10a). The profile for allele 1 was distinguished by two peaks of fluorescence at a melting temperature of 75.5 °C and 77.5 °C (Figure 2.10b). Allele 2 was characterised by a different melting curve profile, with the highest level of fluorescence at 77.5 °C. Allele 3 was characterised by the highest level of fluorescence at 78 °C (Figure 2.10b).

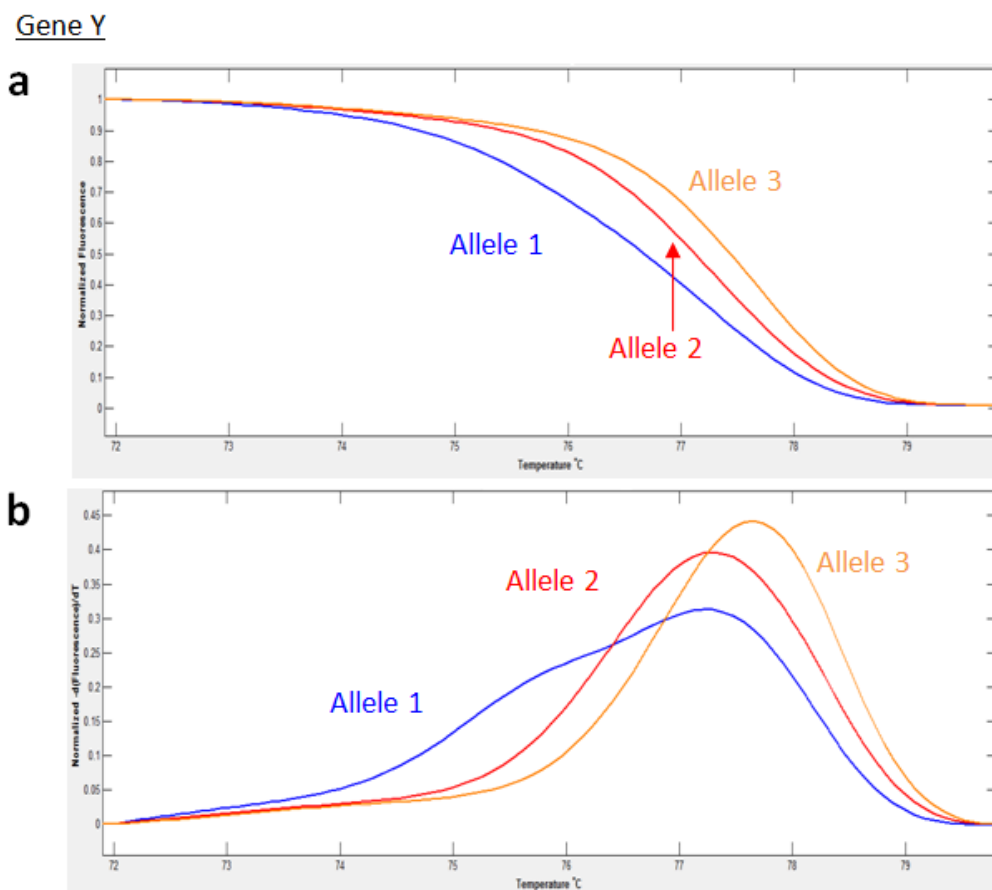


Figure 2.10. High Resolution Melt (HRM) genotyping of Gene Y using PM92. **a.** Normalised melting curve profile of the three alternative alleles showing melting temperature on the x-axis and normalised fluorescence on the y-axis, generated by LightScanner[®] Software (Idaho Technologies). **b.** Normalised melting peaks of the three alleles showing temperature on the x-axis and maximum change in fluorescence on the y-axis, generated by LightScanner[®] Software.

Furthermore, the shape and temperature of the melting curve were analysed so they could be associated with different SNPs identified within the BAC sequence data. This analysis was based on melt-curve characteristics described by MacKay *et al.* (2008) and Distefano *et al.* (2012), where

pairs of G and C nucleotides melt at a higher temperature than pairs of A and T nucleotides. Therefore based on the high melting temperature of allele 3, it likely contained a G & C nucleotide pairing within the double SNP region shown in Figure 2.11. Allele 1 likely contained a T & A pairing and allele 2 was an intermediate between allele 1 and 3, with a T & C pairing (Figure 2.11). However, sequence data of individual polyhaploids would be necessary to confirm the composition of these SNPs.



Figure 2.11. The shape and melting temperature of the Gene Y melting curves could be associated with different types of SNPs identified within the BAC sequence data. The double SNP region is also highlighted.

Genotyping the EMS15C Population

All 118 plants of the EMS15C population were genotyped for Gene Y. Even though three alternative alleles were detected, they were not distributed in a similar way to Genes X and B, in which the Y1 allele was found to be more common than the Y2 and Y3 alleles. In addition, gel-based genotyping was used to identify the *LOP* alleles of Gene Y (*LOPY*), which was not detected in three recombinant plants (Figure 2.12). These recombinant plants were the same as those lacking *LOPX*.

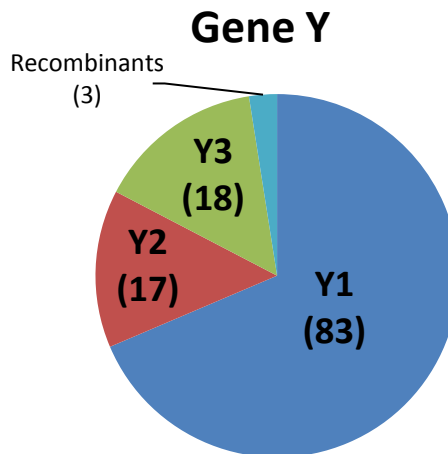


Figure 2.12. Distribution of the alternative alleles of Gene Y in the EMS15C *Hieracium praealtum* polyhaploid population, detected by HRM-based genotyping. Alleles are abbreviated to ‘Y1’ referring to alternative allele 1 of Gene Y, etc. The number of plants with a given allele is shown in brackets. Recombinant plants, which were did not contain *LOPY*, were detected using gel-based genotyping.

2.3.4. Segregation of Alleles

As expected from previous deletion-based mapping of *LOP* (Catanach *et al.*, 2006), the *LOP* alleles of Genes X and B co-segregated. Co-segregation was also seen between alternative alleles of these genes ($p = < 0.001$) reflected in three highly represented genotypes of the population (box highlighted in Table 2.3). All other genotypes were represented at low frequencies (see Figure 2.13) and were therefore under repulsion, i.e. they do not appear together ($p = < 0.001$; Table 2.3).

Table 2.3. Frequency distribution of allele combinations of Genes X and B in the two *Hieracium praealtum* polyhaploid populations. Low p values denote significant linkage in the boxed highlighted allele combinations, and significant repulsion in the other alleles. $N = 287$.

Allele				Frequency	P value
<i>LOPB</i>	<i>LOPX</i>	B1	X1	117	< 0.001
<i>LOPB</i>	<i>LOPX</i>	B1	X2	0	< 0.001
<i>LOPB</i>	<i>LOPX</i>	B1	X3	1	< 0.001
<i>LOPB</i>	<i>LOPX</i>	B2	X1	1	< 0.001
<i>LOPB</i>	<i>LOPX</i>	B2	X2	124	< 0.001
<i>LOPB</i>	<i>LOPX</i>	B2	X3	0	< 0.001
<i>LOPB</i>	<i>LOPX</i>	B3	X1	0	< 0.001
<i>LOPB</i>	<i>LOPX</i>	B3	X2	1	< 0.001
<i>LOPB</i>	<i>LOPX</i>	B3	X3	35	< 0.001

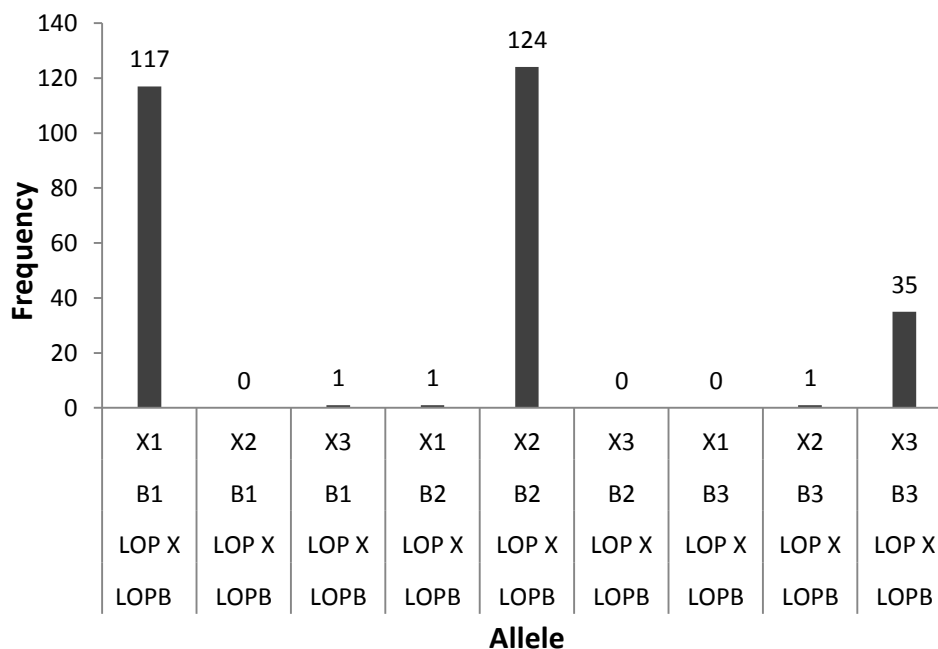


Figure 2.13. Bar chart of the frequency distribution of allele combinations of Genes X and B in the two *Hieracium praealtum* polyhaploid populations.

On the other hand, the frequency data of the alternative alleles of Genes B and H showed no significant linkage disequilibrium or co-segregation ($p = 0.10 - 1$; Table 2.4). It should be noted that even though the B1 & H2 and B2 & H1 allele combinations appear more frequently (see Figure 2.14) they do not represent statistically significant linkage disequilibrium, as shown by the high p values. They have only arisen because the H1 allele was the most common allele of Gene H.

Table 2.4. Frequency distribution of allele combinations of Genes B and H in two *Hieracium praealtum* polyhaploid populations. The high p values indicate no significant linkage. $N = 287$.

Allele				Frequency	P value
LOPB	LOPH	B1	H1	68	0.10
LOPB	LOPH	B1	H2	23	0.33
LOPB	LOPH	B1	H3	6	0.62
LOPB	LOPH	B1	H4	23	0.13
LOPB	LOPH	B2	H1	86	0.14
LOPB	LOPH	B2	H2	18	0.34
LOPB	LOPH	B2	H3	9	0.61
LOPB	LOPH	B2	H4	15	0.18
LOPB	LOPH	B3	H1	25	0.77
LOPB	LOPH	B3	H2	6	1.00
LOPB	LOPH	B3	H3	2	0.68
LOPB	LOPH	B3	H4	5	1.00

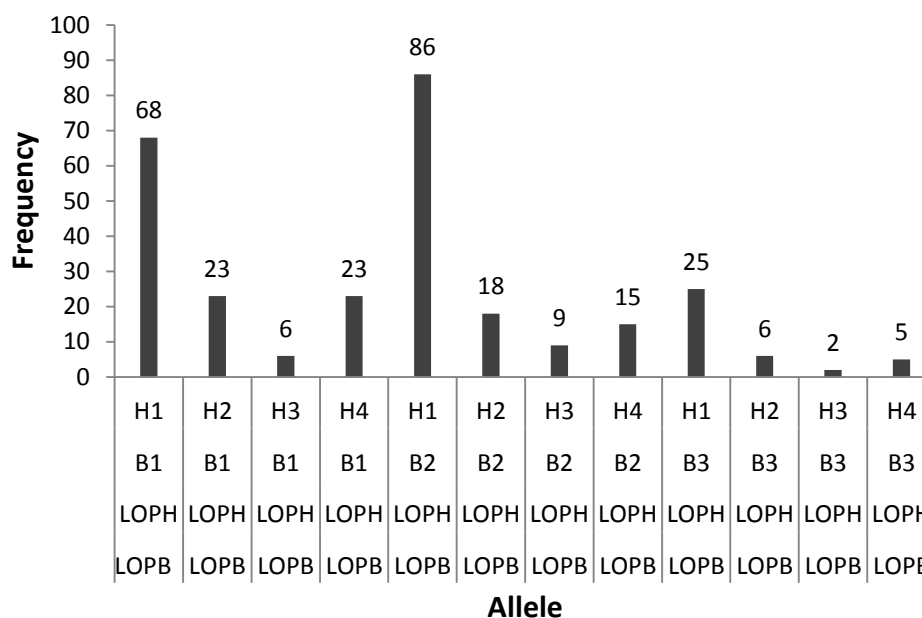


Figure 2.14. Frequency distribution of allele combinations of Genes B and H in two *Hieracium praealtum* polyhaploid populations.

A similar lack of co-segregation was also observed between all allele combinations of Genes B and Y and between Genes Y and H ($p = 0.13 - 1$; Table 2.5, Table 2.6). It should be noted that even though the B1 & Y1, B2 & Y1 and H1 & Y1 allele combinations appear more frequently (see Figure 2.15 & Figure 2.16) they do not represent statistical significant linkage disequilibrium, as shown by the high p values. They have only arisen because the H1 and Y1 alleles are the most common alleles of Genes Y and H.

Table 2.5. Frequency distribution of allele combinations of Genes B and Y in the two *Hieracium praealtum* polyhaploid populations. P values were generated from a Fisher's exact test. $N = 118$.

Allele				Frequency	P value
LOPB	LOPY	B1	Y1	37	0.15
LOPB	LOPY	B1	Y2	6	0.60
LOPB	LOPY	B1	Y3	5	0.29
LOPB	LOPY	B2	Y1	42	0.15
LOPB	LOPY	B2	Y2	10	0.79
LOPB	LOPY	B2	Y3	13	0.13
LOPB	LOPY	B3	Y1	1	1.00
LOPB	LOPY	B3	Y2	1	0.46
LOPB	LOPY	B3	Y3	0	1.00

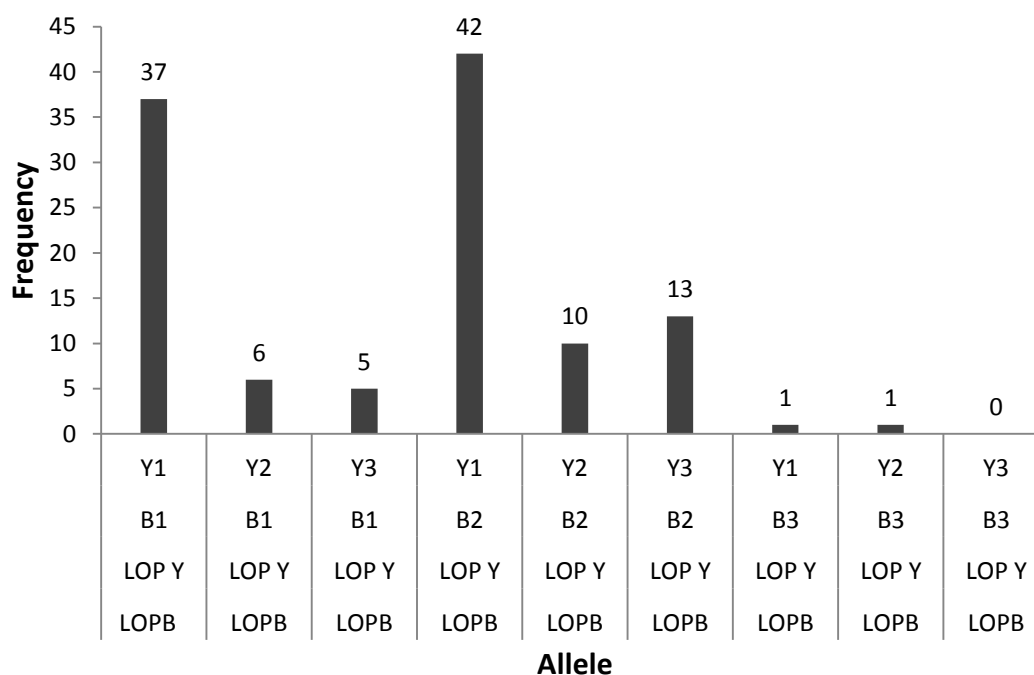


Figure 2.15. Frequency distribution of allele combinations of Genes B and Y in the two *Hieracium praealtum* polyhaploid populations.

Table 2.6. Frequency distribution of allele combinations of Genes Y and H in the two *Hieracium praealtum* polyhaploid populations. *P* values were generated from a Fisher's exact test. *N* = 118.

Allele				Frequency	<i>P</i> value
<i>LOPY</i>	<i>LOPH</i>	Y1	H1	50	0.53
<i>LOPY</i>	<i>LOPH</i>	Y1	H2	15	1.00
<i>LOPY</i>	<i>LOPH</i>	Y1	H3	6	0.67
<i>LOPY</i>	<i>LOPH</i>	Y1	H4	9	0.14
<i>LOPY</i>	<i>LOPH</i>	Y2	H1	10	0.79
<i>LOPY</i>	<i>LOPH</i>	Y2	H2	3	1.00
<i>LOPY</i>	<i>LOPH</i>	Y2	H3	1	1.00
<i>LOPY</i>	<i>LOPH</i>	Y2	H4	3	0.71
<i>LOPY</i>	<i>LOPH</i>	Y3	H1	10	0.60
<i>LOPY</i>	<i>LOPH</i>	Y3	H2	3	1.00
<i>LOPY</i>	<i>LOPH</i>	Y3	H3	0	0.59
<i>LOPY</i>	<i>LOPH</i>	Y3	H4	5	0.13

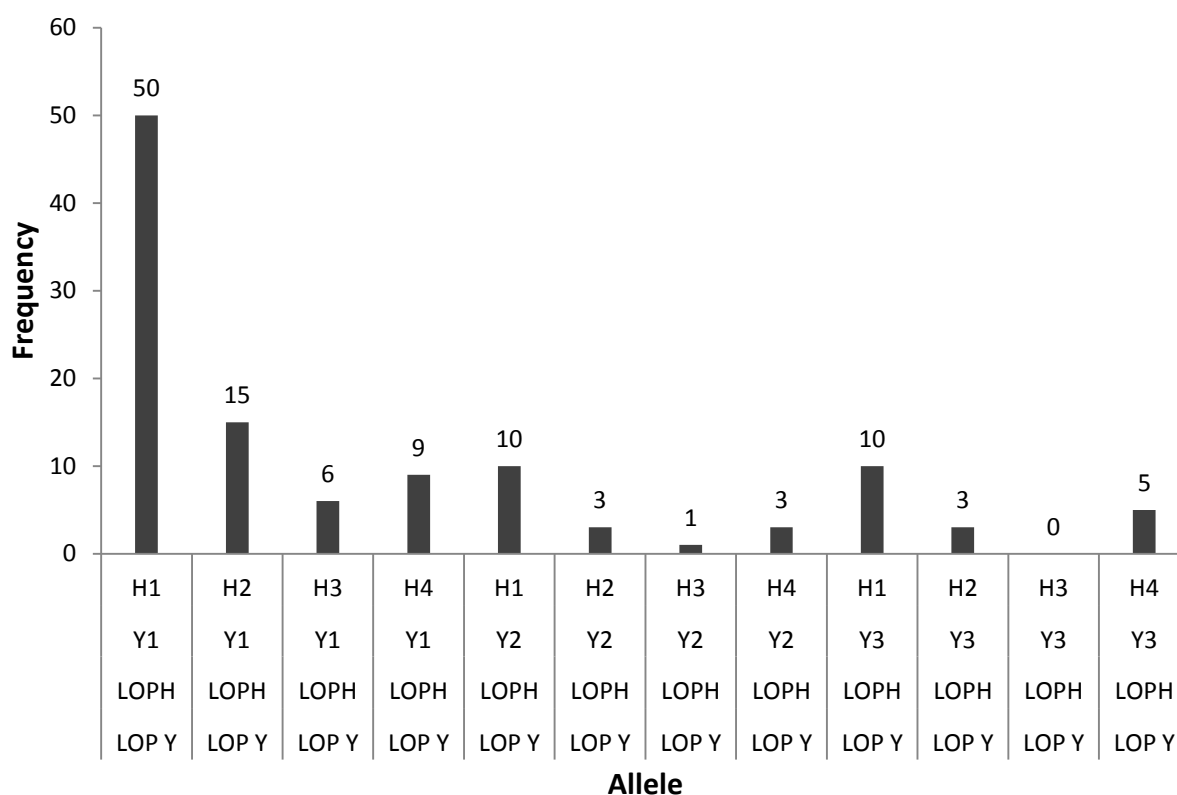


Figure 2.16. Frequency distribution of allele combinations of Genes Y and H in the two *Hieracium praealtum* polyhaploid populations.

2.3.5. Recombination of Alleles

Identification and Confirmation of Recombinant Plants

In total eight recombinant plants were identified that did not contain both *LOPX* and *LOPY*, but retained *LOPB* and *LOPH* (Table 2.7). Of these eight recombinant plants, only recombinant plant #392 did not contain *LOPH* (Table 2.7). Half of the recombinant plants contained both the X1 and X2 alleles (#172, #218, #386, #392, see Tables A.5, A.6), and the other half contained the X2 and X3 alleles (#124, #244, #268, #380, see Tables A.5, A.6).

Table 2.7. The presence (+) or absence (-) of specific *LOP* alleles in eight recombinant plants identified in the two polyhaploid populations.

Polyhaploid Population	Recombinant Plant ID	<i>LOPB</i>	<i>LOPH</i>	<i>LOPX</i>	<i>LOPY</i>
2008 - NE5	#124, #172, #218, #244, #268	+	+	-	-
2011- EMS15C	#380, #386,	+	+	-	-
2011 - EMS15C	#392	+	-	-	-

The genetic composition of the recombinant plants were confirmed by repeating the genotyping steps using the existing DNA from the 2008 polyhaploid population and using re-extracted DNA from the 2011 polyhaploid population. To eliminate the possibility of any recombinant plant being chimeric, individual rosettes of each plant were extracted separately. Any possibility of sampling error was eliminated by extracting DNA and genotyping neighbouring plants to the recombinant plants in the greenhouse. In addition, the absence of the *LOP* allele in the 2011 recombinant plants was further confirmed by PCR testing three additional *LOP* sequence-specific primers listed in Table A.4. Recombinant plant #392 did not contain ~ 100 kb of *LOP* sequence downstream of Gene B as well as at least 200 kb of *LOP* sequence between the coding regions of Gene X and Gene Y. The ploidy of the 2011 polyhaploid recombinant plants was also confirmed using flow cytometry. This was because the average fluorescence value of the recombinant plants was 350,

which as expected was half the fluorescence value of the polyploid ($4n$) EMS15C parent (Figure 2.17).

Recombination Rates

Importantly, recombination was found to have occurred between the *LOP* alleles of Genes B, X, Y and H, contrary to the commonly cited hypothesis which maintains that parthenogenesis loci are non-recombining (Labombarda *et al.*, 2002; Goel *et al.*, 2003). *LOPB* and *LOPX* underwent 2.8% recombination (Table 2.8) and similar rates were evident between the *LOPX*, *LOPY* and *LOPH* (data not shown). The rate of

recombination was also low between the previously identified alternative allele clusters of Genes X and B. The X1 & B1 alleles recombined at a rate of 3.2%, X2 & B2 alleles recombined at a rate of 5.2% and the X3 & B3 alleles recombined at a rate of 16.7% (Table 2.8). On the other hand, all other alternative allele combinations between Genes B,

X, Y and H were found to have undergone over ~ 50% recombination (Data not shown), indicating both a lack of linkage disequilibrium between these alleles and the action of random assortment.

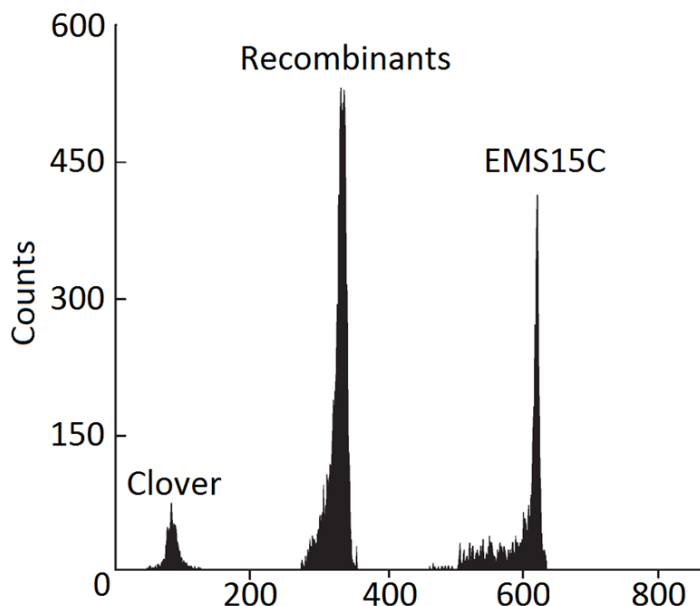


Figure 2.17. Confirmation of the ploidy of the recombinant plants from the 2011 polyhaploid population using flow cytometry. The two reference samples were white clover (*Trifolium repens*) and the EMS15C parent.

Table 2.8. Recombination rates shown as percentages between the alleles of Genes X and B.

Allele	<i>LOP</i> X & B	X1 & B1	X2 & B2	X3 & B3
Total polyhaploids with allele of 1 or both genes	287	124	134	42
# recombinants (allele of one gene absent)	8	4	7	7
Recombination Rate	2.8%	3.2%	5.2%	16.7%

2.4. Discussion

2.4.1. Polyhaploid Genotyping

The genotyping data of Gene X and B both show that all polyhaploids (excluding the recombinants) contained the *LOP* allele and only one other alternative allele for each respective gene. The data suggests that the alternative alleles of Gene X and B do segregate as alleles of the *LOP* alleles at the locus of Gene X and Gene B, as proposed in the hypothesis. In contrast, the genotyping data of Gene H suggests that the alternative genotypes of Gene H are not alleles of the *LOPH* genotype. This was demonstrated by the ‘H2’ polyhaploids that did not contain a single alternative genotype, which would be expected if the alternative genotypes and *LOPH* were segregating. Instead the ‘H2’ polyhaploids contained the *LOPH* and two alternative genotypes (the H1 and H3 peak), which points towards independent assortment of these genotypes.

Distribution of Genes X and B Alleles

The genotyping data of Genes X and B showed that the alleles are similarly distributed throughout the polyhaploid population. Two alternative alleles are common of each gene: X1 and X2 of Gene X and B1 and B2 of Gene B, followed by the rarer X3 and B3 alternative alleles. The X3/B3 alleles may be rarer because they may negatively impact on the seed viability. It is possible that the X3 allele may cause a reduction in Gene X activity. Gene X shows high similarity to eIF3e (Verlhac *et al.*, 1997) and previous studies have reported a general inhibition of translation associated with reduced eIF3e activity (Garcia-Barrio *et al.*, 1995). On the other hand, the X3 allele may cause overexpression of Gene X, which in the case of eIF4e has been shown to lead to tumour formation (Lazaris-Karatzas *et al.*, 1992). Similarly, the B3 allele may alter the activity of Gene B, which shows high similarity to *Arabidopsis* annexin proteins found in the embryos (Cantero *et al.*, 2006; Ge *et al.*, 2007; Yu *et al.*, 2005). The annexin proteins are involved in a number of key processes including cell division and growth, development, gravitropism and the response to pathogens (Mortimer *et al.*, 2008). Therefore, it is possible that any alterations in the functions of Genes B and Gene X, as a result of the inheritance of the X3/B3 alleles, have the potential to cause semi-lethal side effects at the gamete level. However, this possibility cannot be completely validated as the fitness of each plant was not assayed in this study.

Distribution of Genes H and Y Alleles

A key finding from the Gene H genotype data was that the alternative genotypes may not represent alleles. This is because the polyhaploids classified as ‘H2’ contained two genotypes or peaks (H1 and H3) as well as the *LOPH* peak. However, only a single genotype segregating with *LOPH* would be expected if the alternative genotypes were alleles of *LOPH*. Therefore, the data is consistent with the alternative copies of the Gene H being copies of the gene located at another locus or loci of the genome and not at *LOP*.

The genotype data also that the H4 genotype represents a null allele since no discernible fragment peak was associated with this genotype. This indicates that the sequence associated with the H4 genotype was unsuccessfully amplified by PCR using the primer combination PM81. If this is the case, then alternative methods are required to determine if the H4 genotype represents a homologue of Gene H. A null allele may be detected by redesigning the primer binding sites (Callen *et al.*, 1993), using different DNA polymerase (Shinde *et al.*, 2003) using higher quality DNA (Gagneux *et al.*, 1997), or by genotyping neighbouring genetic regions within Gene H. Additional techniques include identifying the number of alleles in Gene H by performing a Southern blot analysis (Southern, 1975), or using RNA-seq (Wang *et al.*, 2009) to identify potential RNAs produced by the hypothetical null allele.

The Gene Y genotype data showed that the Y1 allele was the most common, followed by the rarer Y2 and Y3 alleles. The rarity of the Y2/Y3 alleles may be as a result of these alleles negatively influencing the function of Gene Y. The coding sequence of Gene Y shows high similarity to the SWA1 gene found in functional megaspores and female gametophytic cells, which is essential for the formation of reduced egg cells, i.e. gametogenesis, as reported in *Arabidopsis thaliana* (Shi *et al.*, 2005). Therefore it is possible that any alternations in Gene Y expression, potentially caused by the inheritance of the Y2/Y3 alleles, has the potential to negatively influence the seed viability, as similarly postulated for the X3/B3 alleles.

2.4.2. Segregation of Alleles

Co-segregation of Genes X and B

The frequency data identified three alternative allele clusters, or pairs of alleles, of Genes X and B, that co-segregated frequently. These same genes were also found to co-segregate in a previous study of *LOP* using deletion-based mapping (Catanach *et al.*, 2006). The presence of co-segregation demonstrates that these allele combinations are under genetic linkage. Genetic linkage commonly arises because certain allele combinations may be necessary for a specific function. In the context of *LOP*, parthenogenesis may only function as a result of interactions between these allele combinations, known as epistatic interactions. Epistatic interactions between the *LOP* alleles and the other alleles may impart an advantage to the gamete. This provides an explanation for the higher frequency of the X1, X2, B1 and B2 alleles and the lower frequency of the X3/B3 alleles. One way in which epistatic interactions can cause such a positive effect is by influencing the level of expression and transcription at the gene level (Hirsch *et al.*, 2003). For example, the production of amylase increased with the presence of multiple interacting alleles of the *AMY1* gene (Perry *et al.*, 2007). Similarly, the percentage of milk fat increased in cows that possess a specific combination of alleles from the *DGAT1* gene and the *DGAT1* VNTR promoter (Kühn *et al.*, 2004). This may also be evident at *LOP* in which the expression or transcription level of Genes X and B is increased to a level necessary for parthenogenesis. Epistatic interactions can also induce specific gene functions, known as functional epistasis (Phillips, 2008). Functional epistasis was documented in the innate immune response in humans in which the progression of the AIDS virus was delayed only with the inheritance of a specific combination of receptor and antigen alleles in natural killer cells (Martin *et al.*, 2002). The process of parthenogenesis may be similar, whereby epistatic interactions between the alternative allele combinations are necessary for the trait to function.

The co-segregation of the alleles of Genes X and B may also represent an example of polygenic inheritance, whereby multiple genes are required for a specific phenotypic function (Lander & Schork, 1994). Previous studies have shown that multiple genes can be required to overcome a specific threshold of expression in order for a complex phenotypic trait to function (Lander & Schork, 1994). This was documented in canopy wilting in soybean (Charlson *et al.*, 2009), height differences in humans (Allen *et al.*, 2010) and complex physiological disorders like schizophrenia and bipolar disorder (Purcell *et al.*, 2009). In the context of this study, it is possible that the inheritance of the alternative allele combinations identified between Genes X and B are required to overcome a specific threshold in order for parthenogenesis to function. Further research is required

to investigate the association of parthenogenesis with the co-segregation of alleles of Genes X and B. Specifically, future research could focus on determining if co-segregation still occurs between the same allele combinations when genotyping other regions within these gene. Furthermore, future investigations may benefit from genotyping unlinked microsatellites within non-coding regions to act as control, as similarly undertaken by Corral *et al.* (2009).

Independent Assortment in Genes H and Y

No combination of alternative alleles between Genes X, Y and H, or between Genes B, Y and H, were found to co-segregate at a high frequency. This indicates that the alternative genotypes of Genes Y and Gene H are segregating independently from the *LOP* genotypes. It should be noted that, while the frequency data shows specific pairs of alleles occurring more frequently than others, this may reflect some genotypes (e.g. H1 & Y1) being more common than others, that is, representing more than one sequence in the genome.

2.4.3. Recombination Between Genes of *LOP* and Alternative Alleles

One of the key results of the current study was that recombination was documented between *LOPB*, *LOPX*, *LOPY* and *LOPH*. This was unexpected and contrary to the complete repression of recombination predicted at *LOP* in the hypothesis. The *LOP* locus was not expected to undergo any recombination based on the repressed recombination evident in the apomixis loci in a number of other apomictic species including *Pennisetum* (Goel *et al.*, 2003) and *Paspalum* (Labombarda *et al.*, 2002). Instead, in the current study, a number of polyhaploids were identified that have undergone recombination, referred to as recombinants.

Genetic Composition of Recombinant Plants

Eight recombinant plants were identified in the current study that did not contain the *LOP* regions of specific candidate genes. Genotyping these recombinants revealed that the eight recombinant plants contained very similar genetic compositions. Of interest was that all recombinants, and indeed all polyhaploids, contained *LOPB*. Similarly, *LOPH* was retained in seven recombinants, but was absent in recombinant plant #392. Furthermore, *LOPX* and *LOPY* were also absent in all

eight recombinants. These findings are very informative about which candidate genes are likely to be essential for parthenogenesis.

Gene B

The fact that *LOPB* was present in all 287 polyhaploids including the recombinants provides strong evidence of the essential role of this gene within parthenogenesis. This is also supported by the reported functions of Gene B, which appear to be closely associated with parthenogenesis. Gene B shows high similarity to the calcium signalling protein, annexin D3, which has a critical role in plant reproduction and germination (Cantero *et al.*, 2006; Ge *et al.*, 2007). Calcium signalling is also required for the provision of essential minerals to the ovule (Bednarska, 1989) and is necessary for fertilization, i.e. the fusion between the male and female gametophytes (Ge *et al.*, 2007). More importantly in the context of parthenogenesis, calcium signalling is known to be specifically involved in the avoidance of fertilization that utilises the plant's incompatibility mechanisms (Bednarska, 1989). Previous studies have shown that an incompatibility response produced by the plant will direct annexin proteins to induce a high calcium influx into the pollen tubes, stalling their elongation and avoiding fertilization (Bednarska, 1989; Franklin-Tong *et al.*, 2002; Franklin-Tong & Franklin, 2003). The reported functions of Gene B and the presence of *LOPB* in every member of the population indicate an essential role in parthenogenesis and warrants further investigation.

Gene H

The identification of a recombinant plant lacking *LOPH* indicates that this gene is most likely not essential for parthenogenesis. This argument is further strengthened by the lack of association between the reported functions of Gene H and parthenogenesis. Gene H shows high similarity to an ATP-dependent PIF1-like DNA helicase, which has a number of essential roles in chromosome stability and maintenance (Pinter *et al.*, 2008) and within DNA replication (Budd *et al.*, 2006). While these reported functions do not create a strong link to parthenogenesis, it is still possible that Gene H is involved in the process of parthenogenesis.

Genes X and Y

Even though *LOPX* and *LOPY* were lacking in all eight recombinant plants, these genes may still play a role in parthenogenesis based on their reported functions. Gene X shows high similarity to the eukaryotic initiation factor 3 subunit E (eIF3e), which is involved in the initiation of translation, mating and cell proliferation (Verlhac *et al.*, 1997) which may be important in parthenogenesis. Furthermore, the related eukaryotic initiation factor 4 (eIF4) has been reported to be involved in the transition of the egg cell to an embryo (Evsikov *et al.*, 2006), further reinforcing the prospects of an association between Gene X and parthenogenesis. Likewise, Gene Y shows high similarity to the *SWA1* gene, which is expressed in functional megaspores and female gametophytic cells and is reported to be essential in the formation of reduced egg cells (gametogenesis) in *Arabidopsis thaliana* (Shi *et al.*, 2005). Therefore, based on these reported functions, it is possible that Genes X and Y play a role in parthenogenesis. However, the recombination data does show the absence of *LOPX* and *LOPY* in recombinant polyhaploids, implying that they were not essential for the trait.

Recombination Rates

One of the major findings of this study was the detection of recombination within the *LOP* locus. This was contrary to the hypothesis, which predicted a complete repression of recombination at *LOP*. This hypothesis was based on the repression of recombination documented in the apomixis loci of other apomictic species including *Pennisetum* (Goel *et al.*, 2003) and *Paspalum* (Labombarda *et al.*, 2002).

A recombination rate of 2.8% was calculated between the *LOPX* and *LOPB*. A recombination rate of 6% may be estimated between these same genes in lettuce, in which they are 10.2 Mb apart and 1 cM (centiMorgan) is 1.7 Mb (Truco *et al.*, 2013). One potential reason for the lower rate of recombination in the current study is that recombination events associated with double reduction may not have been detected, which may be a feature of the polyhaploid population and for tetraploid populations in general. For example, if the B1 and X1 underwent recombination with *LOP* followed by migration to the same gamete, the recombination event would not be detected in this study. Therefore, it is important to note that the rate of recombination may be underestimated in this study.

The alternative alleles of Genes X and B underwent a higher rate of recombination between 3.2% - 16.7%. Of these alternative alleles, the B3 and X3 alleles may be seen to recombine more often (16.7%) as recombination may eliminate effects of reduced seed viability that may be associated with these alleles. Amongst the other candidate genes, the recombination rate was also lower between the *LOP* alleles, for example 0.3% recombination was calculated between *LOPX* and *LOPY*, and substantially higher between all other alternative alleles combinations (~ 50%).

2.4.4. Inheritance of Alleles

The genotyping and recombination data showed no evidence of disomic inheritance at the *LOP* locus and therefore allopolyploidy is unlikely in *Hieracium praealtum*. If disomic inheritance was present, specific sets of alleles would be expected to act like homologues, whereby each allele shows a strong propensity to segregate only with its homologue. This was not observed in the current study in which almost all possible combination of alleles were present in the polyhaploid populations. Similarly, under conditions of disomic inheritance, each allele would be expected to exclusively recombine only with one other which is effectively its homologue. As shown in the genetic composition of the recombinant plants, this was not observed in the current study in which the *LOP* allele did not preferentially recombine with a specific allele but instead underwent equal levels of recombination with all three alternative alleles.

The genotype data from the *LOP* locus suggests tetrasomic inheritance, which is an indication of autopolyploidy, both of which were predicted in the hypothesis. Autopolyploidy is common among apomict species, as reported in *Paspalum rufus* (Quarín *et al.*, 1998), *Tripsacum* (Grimanelli *et al.*, 1998) and within 10 different *Hieracium* species (Chrtek *et al.*, 2009; Krak *et al.*, 2013). In tetrasomic inheritance, each alternative allele has an equal chance of being coinherited with *LOP*. This was observed in the current study, in which the X1, X2, B1 and B2 alleles of Genes X and B were equally distributed in the population. While the alleles X3 and B3 were less commonly represented, but they maybe because they are under negative selection.

2.4.5. Future Work

An important focus of future investigations is the potentially pivotal role of Gene B in parthenogenesis. This recommendation is based on the consistent representation of Gene B in each

member of the polyhaploid population, which indicates that it may be required for polyhaploid production. Future investigations could focus on identifying regulatory elements, promoter regions, and CpG islands within the sequence of *LOPB*. These investigations may produce evidence of active transcription and/or regulation, and provide further support for the important functional role of this gene. Another potential indicator is up-regulation of Gene B expression during critical stages of parthenogenesis. This could be detected using qPCR (Livak & Schmittgen, 2001). In a similar way, elevated levels of Gene B mRNA during parthenogenesis would be a good indicator of an association between Gene B and parthenogenesis. This could be detected using RT-PCR (Gibson *et al.*, 1996). Furthermore, the annexin D3 protein that shows high similarity to Gene B could be the focus of future investigations. A range of methods are available to detect and purify the protein, uncover the structure and detect protein-DNA and protein-protein interactions, as outlined by Bollag *et al.* (1996). Studying the annexin protein may shed light on the potentially pivotal role of Gene B in parthenogenesis.

Another key area for future research could be to confirm the pattern of recombination discovered in this study. This would aim to prove that *LOPB* is always associated with parthenogenesis while *LOPX*, *LOPY* and *LOPH* are not necessary. Future research could seek to generate new populations and different species of plants which are both polyhaploids and tetraploids. Newly identified recombinants could be genotyped using a variety of *LOP* sequence-specific sequences to confirm that *LOPX*, *LOPY* and *LOPH* are always absent due to recombination.

The scope of future *LOP* research could also be expanded to investigate other candidate genes identified at the *LOP* locus. Further down the track, once important candidate genes have been more accurately determined, they could be transferred into a well known model system such as *Arabidopsis thaliana* using the floral dip method (Clough & Bent, 1998). This method operates by dipping the flowers of a plant into a solution of *Agrobacterium* carrying the desired genes within their plasmids, which are then transferred into some of the seeds (Clough & Bent, 1998). By using such a technique, it would be possible to selectively add candidate genes and their alleles into a model plant to determine which ones are truly essential for introducing the parthenogenesis trait. This will be an important step towards the ultimate goal of introducing the apomixis trait into crop species.

2.4.6. Conclusion

The genotyping data from this study demonstrated that the alternative alleles of Genes X and B are allelic to the *LOP* alleles, which was not apparent in the genotypes of Gene H. The genotyping data also showed that specific alternative allele clusters or pairs of alleles of Genes X and B were found to frequently co-segregate. Co-segregation indicates genetic linkage, which has likely arisen, or at least been maintained, as these allele combinations may be necessary for parthenogenesis. Epistatic interactions between these allele clusters may increase the expression or transcription of Genes X and B or be required to induce a specific function at *LOP*. Similarly, the influence of both Genes X and B may be necessary to overcome a threshold of expression necessary for parthenogenesis. Another possibility is the combined influences of Genes X and B may be required to overcome an expression threshold necessary for parthenogenesis, which is typical of polygenic inheritance.

Another significant finding of this study was that recombination was documented within *LOP*, contrary to the complete repression of recombination predicted. Eight recombinant plants were identified in this study, and their genetic composition provided important information regarding which candidate genes were likely essential for parthenogenesis. Of interest was *LOPB*, which appears to be essential for parthenogenesis since it did not undergo recombination and was present in all 287 polyhaploids including the recombinants. In contrast, *LOPX*, *LOPY* and *LOPH* were absent in a number of recombinant plants which had no impact upon their ability to undergo parthenogenesis. The reported functions of Gene B also reinforce the potential importance of this gene in parthenogenesis. Specifically, the annexin D3 calcium signalling protein, which shows high similarity to Gene B, ensures that fertilization is avoided by creating a high calcium influx into the pollen tube halting its elongation. However, further investigation is required to uncover the potential importance of Gene B within parthenogenesis. Future techniques could focus on investigating the regulatory elements within Gene B and looking at Gene B expression levels and mRNA activity during parthenogenesis.

The genotype and recombination data also indicated tetrasomic inheritance within *LOP*, which would be consistent with an autopolyploidy origin of *Hieracium praealtum*. This was based on the equal representation of the X1, X2, B1 and B2 alleles in the population. The X3 allele was less frequent as it suspected to negatively impact the seed viability.

CHAPTER 3

ALIGNMENTS

3.1. Introduction

3.1.1. Allelic Sequence Divergence

Some genetic loci are known to contain high levels of interallelic variation where certain alleles display greater sequence divergence relative to the other alleles. This phenomenon is known as allelic sequence divergence (ASD). Repressed recombination has been proposed as one of the likely causes of ASD within asexual lineages (Corral *et al.*, 2009). Alleles experiencing repressed recombination will rarely if at all recombine with the rest of the genome, which causes their sequence to diverge over evolutionary time. The mechanisms by which recombination is repressed are possibly similar to those identified within the zone of repressed recombination around the centromere, where recombination and transcription are suppressed to ensure that meiotic chromosomes undergo proper segregation (Ellermeier *et al.*, 2010). In this system, recombination is actively suppressed by the Clr4-Rik1 complex; a component of the heterochromatin, which is directed to the appropriate histone molecules by a RNA-induced transcriptional silencing complex (Ellermeier *et al.*, 2010; Castel & Martienssen, 2013).

Alleles can also display signs of ASD if they have been introgressed into a genome via interspecific hybridization. This has been documented in a previous study in which interspecific hybridization was attributed as the likely cause of high levels of divergence between alleles of surface antigens in *Plasmodium falciparum* (Roy *et al.*, 2006). Interspecific hybridization is well documented among the *Hieracium* genus (Morgan-Richards *et al.*, 2004; Krak *et al.*, 2013); it is therefore possible that *Hieracium praealtum*, the model species in the current study, might have undergone introgressive hybridization in the past.

3.1.2. Consequences of Allelic Sequence Divergence

ASD of certain alleles or genes can have beneficial or detrimental effects upon an organism. Beneficial consequences may include the acquisition of a novel function. In the case of an asexual

microinvertebrate (bdelloid rotifers), a diverged allele was found to impart a higher level of desiccation tolerance (Pouchkina-Stantcheva *et al.*, 2007). On the other hand, the repression of recombination, and the resulting ASD, can negatively impact upon the homologous repair mechanisms that removes mutations (Van Dijk *et al.*, 2009). The sequence of an isolated allele may, therefore, steadily accumulate mutations, some of which maybe deleterious (Bicknell & Koltunow, 2004; Docking *et al.*, 2006). This is known as the Meselson effect (Mark Welch *et al.*, 2008). Isolated alleles have also been shown to undergo a greater number of insertions and deletions (Conner *et al.*, 2008) and may acquire a growing number of transposable elements (Van Dijk *et al.*, 2009), all of which can have a detrimental effect on the organism. A previous study on *Taraxacum* found that the overall fitness of apomictic clones were negatively affected by the accumulation of transposable elements (Docking *et al.*, 2006). In some cases, the extinction of a cloned organism may be accelerated as a result of the steady accumulation of slightly deleterious elements (Van Dijk *et al.*, 2009).

3.1.3. ASD of Apomixis Genes and Loci

ASD has been documented within the apomixis genes of asexual *Boechera holbellii* populations, which is believed to have arisen due to repressed recombination at the apomixis locus (Corral *et al.*, 2009). Repressed recombination around apomixis-associated loci has also been reported in a number of other apomict species, including *Paspalum* (Labombarda *et al.*, 2002) and *Taraxacum* (Van Dijk *et al.*, 2009). Similarly, evidence of repressed recombination was found within the sequence that controls apomixis, known as the Apospory Specific Genomic Region (ASGR) in *Pennisetum* (Goel *et al.*, 2003). This large hemizygous region was also found to have undergone a number of gene duplications and to contain multiple retrotransposons (Conner *et al.*, 2008). Due to the repression of recombination, the ASGR locus could not be mapped using conventional linkage mapping. With this in mind, deletion mutagenesis was used to map the apomixis loci in *Hieracium praealtum* which identified the *LOSS OF PARTHENOGENESIS (LOP)* locus, which is associated with the avoidance of the fertilization step of apomixis (Catanach *et al.*, 2006). Sequences of *LOP* were generated and up to eighteen candidate genes predicted (Dr. Andrew Catanach personal communication). Three of these candidate genes were the focus of this chapter: Gene B, Gene X and Gene Y. The coding sequence of Gene B shows high similarity to an *Arabidopsis* annexin

protein involved in calcium signalling (Ge *et al.*, 2007) and the Gene X coding sequence shows high similarity to a eukaryotic translation initiation factor (eIF3e; Verlhac *et al.*, 1997). The Gene Y coding sequence shows high similarity to a transducin family protein, which is active in the megaspore and is potentially important for gametogenesis (Shi *et al.*, 2005).

Recent research on *Hieracium praealtum* revealed that there were multiple homologous copies of each of these candidate parthenogenesis genes across the *LOP* locus which are believed to represent recessive alleles. Each gene was found to encompass an ‘*LOP*’ copy as well as at least three other copies referred to as alternative alleles (Figure 1.2a in Chapter 1; McGee, PFR Summer Studentship 2010-2011, unpublished data). This was expected since *Hieracium praealtum* is a tetraploid (Nielsen, 1980; Wolf *et al.*, 1990; Diwan *et al.*, 2000; Mahy *et al.*, 2000; Tiffin & Gaut, 2001), therefore each individual will contain four alleles of each gene.

3.1.4. Allelic Sequence Divergence of the *LOP* Candidate Genes

Recently, short Sanger sequences were generated of three alternative alleles of Genes A, X and Y at *LOP* as part of a PFR summer studentship. These sequences were aligned with their *LOP* allele, revealing evidence of ASD of the *LOP* allele, relative to the alternative alleles (Figure 1.4 in Chapter 1; McGee, PFR Summer Studentship 2010-2011, unpublished data). The main focus of this study was to assess ASD in the selected candidate genes at *LOP* and the genomic regions surrounding these *LOP* alleles. One of the ways this was determined was by comparing allele sequences of each candidate gene to identify those alleles that were divergent or isolated. Multiple sequence alignments were created using the existing *LOP* sequence data from the PFR apomixis research team and the alternative allele sequence data generated from 454 pyrosequencing of bacterial artificial chromosomes (BACs) in Chapter 2.

Another approach to detect ASD was to construct distance trees of the multiple sequence alignments. In order to construct an accurate distance tree, the most appropriate DNA sequence evolution model was determined for each separate sequence alignment. The Jukes-Cantor model is one such evolution model, which assumes each nucleotide base is equally common and therefore equally probable to undergo nucleotide substitutions between one another (Jukes & Cantor, 1969). An alternative model, HKY, is more complex, which allows for unequal base frequencies and

contains a separate rate for transition and transversion (Hasegawa *et al.*, 1985). The models will be tested and the one with the highest likelihood value will most accurately represent a specific sequence alignment (Guindon & Gascuel, 2003). The identified model was then used to construct a distance tree, in which diverged alleles could be identified as those located on the longest branches of the distance tree. Divergent alleles were also identified by generating large-scale alignments by comparing ~ 50 kb of the sequences that flank the coding region within the alleles. Evidence of ASD was then detected within these flanking sequences, based on the allele which possessed the greatest number of retrotransposons (Matzk *et al.*, 2003), insertions, deletions and substitutions (Conner *et al.*, 2008). The large-scale alignments were generated using the global alignment package Mauve (<http://gel.ahabs.wisc.edu/mauve/>, Darling *et al.*, 2004).

Investigating the potential presence of ASD will also assist in efforts to understand the origin and evolutionary history of the *LOP* locus. Working towards this goal, the new allele sequence data generated in Chapter 2 was used to determine the level of sequence similarity between the alleles. A high level of similarity would indicate that all the alleles originated from an ancestral allele. This would add weight to the theory that the *LOP* allele originated from the functional divergence of an ancestral allele (Tucker & Koltunow, 2009). In contrast, a low level of sequence similarity may indicate that the *LOP* allele originated in another species and was introduced into the *Hieracium* genome through ancestral interspecific hybridization (Carman, 2007; Tucker & Koltunow, 2009).

3.1.5. Research Hypothesis

The central research hypothesis of this study was that *LOP* alleles display ASD from the corresponding alternative alleles as a result of repressed recombination within *LOP*. Furthermore, the *LOP* alleles and the alternative alleles originated from the same ancestral allele.

3.1.6. Aims of the Study

The main aim was to assess ASD of the *LOP* alleles of Genes B and X within the *LOP* region. To achieve this, two approaches were undertaken using the existing sequence of the *LOP* allele and new alternative allele sequence data generated from these genes in Chapter 2. The first approach

was to compare the candidate gene regions between the alleles of each gene and to construct associated distance trees. An allele which has undergone ASD will have a considerably different sequence to the alternative alleles. The second approach was to detect larger-scale evidence of ASD by comparing the entire gene and flanking sequences of the *LOP* allele with the alternative alleles, using the program Mauve. A greater number of retrotransposons and insertions and deletions would also be consistent with ASD. Ascertaining ASD of the *LOP* allele also provided important data relating to the evolutionary history and origin of the *LOP* locus.

3.2. Materials and Methods

3.2.1. Sequence Data

The sequence data of the *LOP* alleles of Genes B, X and Y, designated *LOPB*, *LOPX*, *LOPY* in the study, were derived from the concatenated *LOP* allele sequence assembled by the PFR apomixis research team (Catanach *et al.*, 2006). The sequence data of the X1, X2, B1, B2 and Y1 alternative alleles of Genes B, X and Y were generated via 454 pyrosequencing of BACs, which were subsequently assembled and concatenated into different alternative alleles (as described in the methods section in Chapter 2). The short Sanger sequences of the X3 and B3 were previously generated as part of my PFR summer studentship (McGee, PFR Summer Studentship 2010-2011, unpublished data).

3.2.2. Multiple Sequence Alignments using ClustalW, and Generation of Distance Trees

The genetic regions that were amplified by PM 1 and PM 34 (originally for genotyping purposes in Chapter 2) were first extracted from the concatenated sequences of *LOPB* and *LOPX* and from the B1, B2, X1 and X2 alternative allele sequences. The same genetic region in the X3 and B3 alleles was also extracted from previous Sanger data (McGee, PFR Summer Studentship 2010-2011, unpublished data). These four sequences, each representing an allele, were then aligned using the ClustalW alignment plug-in in Geneious 6.1 (Biomatters Ltd, <http://www.geneious.com>) with the default parameters of a 65% cost matrix, gap open penalty of 12 and a gap extension penalty of 3. Using these multiple sequence alignments, the number of unique SNPs in the *LOP* allele were counted.

In order to construct an accurate distance tree, the most appropriate DNA sequence evolution model for each sequence alignment was first determined using jModelTest 2.1.3. (Guindon & Gascuel, 2003; Durrbin *et al.*, 2012) at <https://code.google.com/p/jmodeltest2>. The most suitable model for a specific multiple sequence alignment was chosen when constructing a distance tree using the Geneious Tree Builder in Geneious 6.1, employing the neighbour-joining tree method. Distance matrices showing the number of nucleotide substitutions between each allele were also outputted as part of the distance tree construction process within Geneious 6.1.

The entire genomic regions of *LOPB*, B1 and B2 alternative alleles were also aligned using the ClustalW alignment plug-in within Geneious 6.1. These sequences were derived from the previously generated concatenated *LOP* contigs and the concatenated contigs of the B1 and B2 alternative alleles generated in Chapter 2. To simplify the comparison process, a 6.7 kb transposable element (TE) was removed from the genomic region of the B2 allele. In addition, the number unique SNPs were counted within the *LOP* allele.

3.2.3. Genomic Sequence Alignments using Mauve

In total, 25 kb downstream and 25 kb upstream of *LOPB*, *LOPX* and *LOPY* were extracted from the existing concatenated *LOP* allele sequence. The same sized region was extracted from the concatenated alternative allele sequences generated by 454 pyrosequencing in Chapter 2. These ~ 50 kb sequences each representing an allele, were then compared using the progressiveMauve alignment method (Darling *et al.*, 2010) employed by the Mauve program (Darling *et al.*, 2004). The visualised output from Mauve was used to identify deletions, insertions and genetic rearrangements along the entire ~ 50 kb region.

3.2.4. Transposable Elements

Long terminal repeat-retrotransposons were identified within the genomic sequence alignments using BLAST to scan against the RetroOryza library (Chaparro *et al.*, 2007) at <http://retroryza.fr>. Additional transposons were identified using the *RepeatMasker Web Server* (Smit *et al.*, 2013) at www.repeatmasker.org. Contingency tables were constructed showing the differences in the number of identified TEs between the *LOP* allele and the alternative alleles. Probabilities (*p*-value) were calculated for this relationship using the inbuilt 2x3 Fisher's exact test, in Minitab 16 (Minitab Inc., State College USA). In addition, the identified retrotransposons were annotated onto the ~ 50 kb regions.

3.3. Results

3.3.1. Multiple Sequence Alignments

ClustalW alignments showed that the primer-amplified sequence of *LOPB* varied considerably from the same region within the alternative alleles (Figure 3.1). *LOPB* was found to contain a greater number of AG repeats (labelled in Figure 3.1) than the alternative alleles and differed in overall length by 15 - 17 bp. This was in contrast to the alternative alleles which shared a high level of sequence similarity with one another.

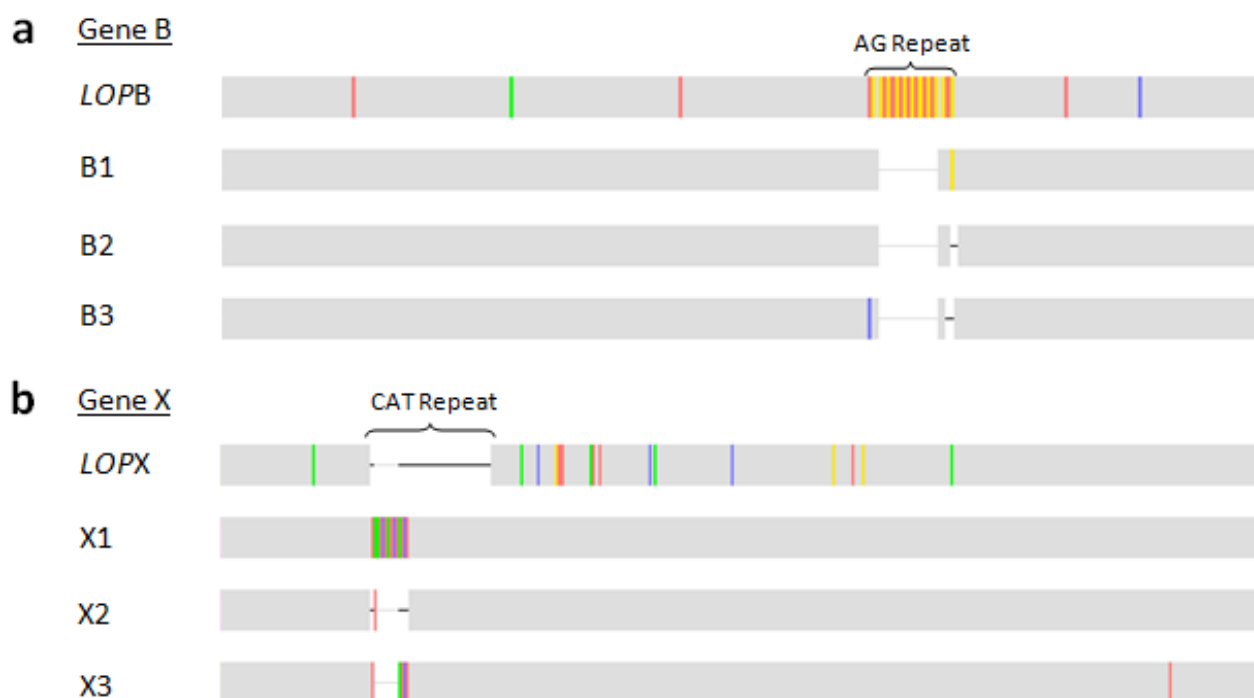


Figure 3.1. Multiple sequence alignments of the four alleles of Gene B (**a**) and of Gene X (**b**). The label for each allele was shortened. For example, *LOPB* refers to the *LOP* allele of Gene B and X1 refers to alternative allele 1 of Gene X, etc. The different colours denote different nucleotide changes between the sequences, in which sequence conservation is denoted in grey. The *LOP* sequences were derived from the PFR apomixis research team, the X1, X2, B1 and B2 alternative allele sequences were amplified using PM 1 and PM 34 in Chapter 2, and the X3 and B3 sequences were generated in my previous PFR summer studentship.

The sequence of *LOPB* was also found to contain 5 unique single nucleotide polymorphisms (SNPs), which were not observed in the alternative alleles (Figure 3.1).

The ClustalW alignments of Gene X also indicated that *LOPX* differed considerably from the alternative alleles (Figure 3.1). *LOPX* was found to differ in length by as many as 36 - 45 bp as a result of a lower number of CAT repeats when compared to the alternative alleles (labelled in Figure 3.1). The sequence of *LOPX* was also found to contain 16 unique SNPs, which were not shared by the alternative alleles (Figure 3.1).

The multiple sequence alignment of the entire genomic region of Gene B further highlighted that *LOPB* differed considerably from the alternative alleles. Specifically *LOPB* was found to possess 36 unique SNPs which were not shared by the B1 and B2 alternative alleles (Figure 3.2).



Figure 3.2. Multiple sequence alignment of the entire genomic region of the three sequenced Gene B alleles. The allele names are abbreviated. For example, *LOPB* refers to the *LOP* allele of Gene B and X1 refers to alternative allele 1 of Gene X, etc. The black lines indicate nucleotide differences between the sequences and grey regions signify sequence conservation. The asterisk “*” denotes a 6.7 kb TE removed from the B2 allele sequence to ease the alignment process.

3.3.2. Distance Trees and Distance Matrices

Prior to constructing the distances trees, jModelTest 2.1.3 was used to determine the most suitable model of DNA sequence evolution for the two previously generated multiple sequence alignments of Gene B and X. The Jukes-Cantor model (Jukes & Cantor, 1969) was found to be the most suitable for the Gene B alignment since it possessed the highest likelihood value of 424. The HKY model (Hasegawa *et al.*, 1985) was found to be most suitable for the Gene X alignment since it possessed the highest likelihood value of 727. The specific model for the specific sequence alignment was then selected when constructing a neighbour-joining distance tree.

In both the distance trees, the *LOPB* and *LOPX* were situated on a long branch, separate from their alternative alleles (Figure 3.3). The results were similarly within the distance matrices used to construct the distance trees, in which the *LOP* allele sequence had undergone a higher number of nucleotide substitutions when compared to the alternative alleles. *LOPB* was found to have undergone 3.3 – 3.7 nucleotide substitutions per 100 nucleotides when compared to the alternative alleles (Table 3.1a). In contrast to comparing the alternative alleles to one another, which were very similar with only 0.2 – 0.8 nucleotide substitutions per 100 nucleotides (Table 3.1a). A higher number of nucleotide substitutions in the *LOP* alleles compared to the alternative alleles was also apparent in Gene X (Table 3.1b). *LOPX* differed by as much as 4.9 – 7.9 nucleotide substitutions

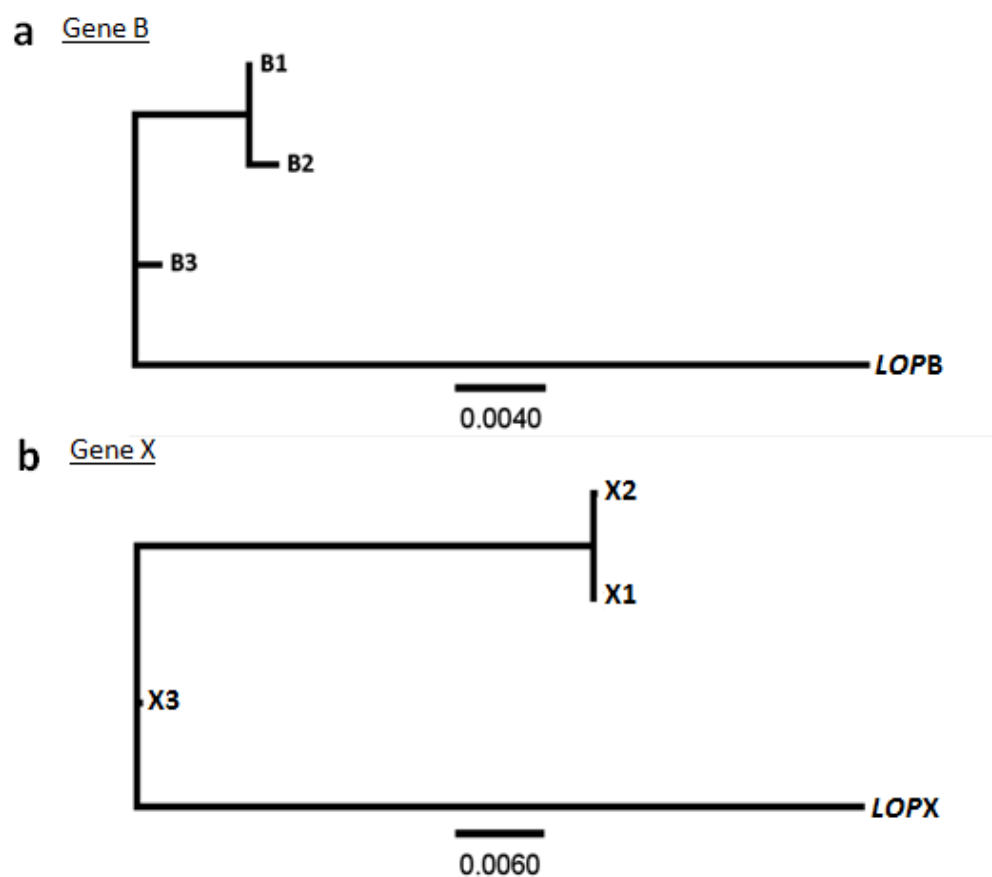


Figure 3.3. Neighbour-joining distance trees of the four alleles of Gene B (a) and of Gene X (b). The label for each allele was shortened. For example, *LOPB* refers to the *LOP* allele of Gene B and X1 refers to alternative allele 1 of Gene X, etc. The scale bar represents the expected number of nucleotide substitutions per site along the whole sequence. The distance trees were created using the previously generated multiple sequence alignments of Gene B and X.

per 100 nucleotides when compared to the alternative alleles (Table 3.1b). This was in contrast to the alternative alleles, which were more similar with only 2.9 – 5.9 nucleotide substitutions per 100 nucleotides difference between one another (Table 3.1b).

Table 3.1 Distance matrices of the number of nucleotide substitutions per 100 nucleotides between each allele in Gene B (**a**) and Gene X (**b**). The label ‘*LOPB*’ refers to the *LOP* allele of Gene B and ‘B1’ refers to alternative allele 1 of Gene B, etc. The distance matrices were created using the previously generated multiple sequence alignments of Genes B and C.

a	<i>LOPB</i>	B1	B2	B3	b	<i>LOPX</i>	X1	X2	X3
<i>LOPB</i>		0.037	0.037	0.033	<i>LOPX</i>		0.079	0.079	0.049
B1	0.037		0.002	0.004	X1	0.079		0.029	0.059
B2	0.037	0.002		0.008	X2	0.079	0.029		0.047
B3	0.033	0.004	0.008		X3	0.049	0.059	0.047	

3.3.3. Genomic Sequence Alignments using Mauve

Transposable elements (TEs) were identified by BLAST and annotated onto the ~ 50 kb sequences of *LOPB* and the B1, B2, X1 and Y1 alternative alleles. The Mauve alignment of the three Gene B alleles indicated that the sequence of *LOPB* allele differed considerably from that of the two alternative alleles. On the other hand, the alternative alleles shared a high level of similarity. One large-scale difference apparent in *LOPB* was an insertion of ~ 2 kb (labelled ‘TE Insert’ in Figure 3.4), which showed homology to a known transposon (data not shown). Another difference was *LOPB* appeared to have undergone a deletion of ~ 5 kb of sequence common to the two alternative alleles (labelled ‘TE Del’ in Figure 3.4). Based on the proximity of a neighbouring TE, it is possible that this sequence was deleted as a result of retrotransposition. In addition, BLAST output indicated a 6.7 kb retrotransposon insertion within the intron of the alternative allele B2 of Gene B (labelled ‘TE Insert’ in Figure 3.4). There are also small insertions / deletions apparent in *LOPB* relative to the alternative alleles (labelled ‘In/del’ in Figure 3.4). Overall, there appeared to be a greater number of insertion / deletions in the *LOPB* sequence compared to the alternative alleles.

The Mauve alignment of Gene X highlighted a number of key differences in *LOPX* when compared to the X1 alternative allele. One of the key differences was that *LOPX* contained a ~ 7 kb insertion (labelled 'TE Insert' in Figure 3.5), which may have been the result of a TE identified nearby. *LOPX* also contained evidence of multiple deletions of sequences common to the X1 alternative allele (labelled 'Del' in Figure 3.5). However, it cannot be stated with complete certainty that these deletions took place at the *LOP* alleles. It was possible that this sequence variation may reflect an insertion at the X1 allele. A similar uncertainty exists with respect to the translocation event identified between *LOPX* and X1 allele (labelled 'Translocation' in Figure 3.5), as the available data does not distinguish the allele that had undergone the translocation. This uncertainty stems from the lack of sequence data as only the X1 allele was available for comparison with *LOPX*.

Greater certainty can be placed on interpretations made from the Mauve alignment of *LOPY* and Y1 as their flanking regions shared almost no sequence similarity to one another (Figure 3.6). Out of the ~ 50 kb of sequence, only the ~ 2 kb of sequence that encompassed the coding sequence of Gene Y was found in both alleles (Figure 3.6). Based on the sequence data generated from this study, this suggested that the Gene Y had been inserted into the *LOP* locus.

3.3.4. Transposable Elements in the Genomic Sequence Alignments

The number of TEs identified in the ~ 50 kb sequences were analysed in a contingency table (Table 3.2). *LOPX* and *LOPY* contained a slightly higher number of retrotransposons than the alternative alleles (Table 3.2). However, *LOPB* was found to contain only two retrotransposons compared to 8 retrotransposons identified in the B1 allele and 7 in the B2 alleles (Table 3.2).

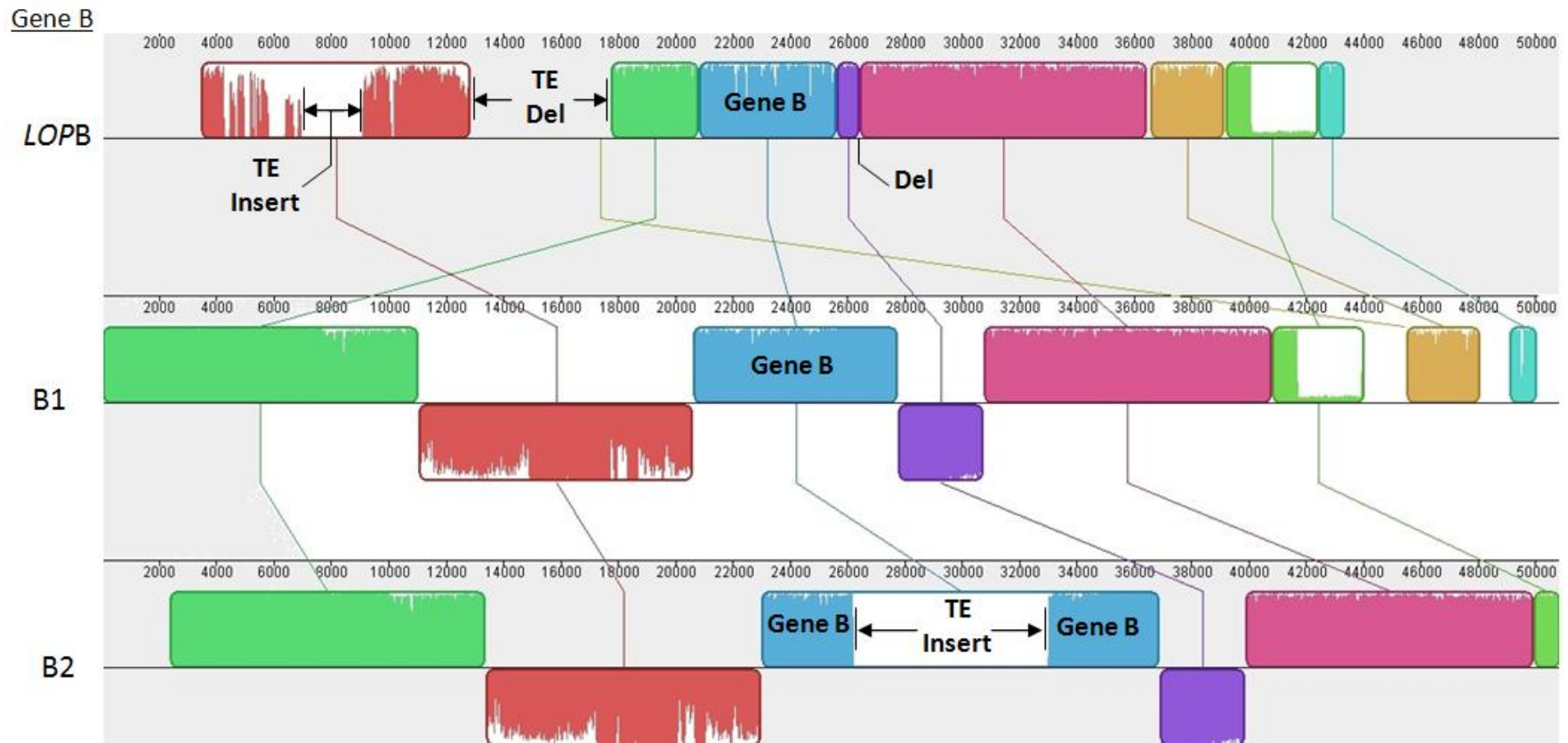


Figure 3.4. Mauve alignment of 25 kb downstream and 25 kb upstream of Gene B between *LOPB* and the B1 and B2 alternative alleles. Regions of the same colour indicate high sequence similarity and are connected by lines of the same colour, while white areas indicate a lack of sequence similarity. The location of Gene B is labelled. Deletions between two alleles are labelled as ‘Del’ while insertions or deletions thought to be caused by retrotransposition are labelled as ‘TE Insert’ or ‘TE Del’. The distance in bp along the entire alignment is noted on the top axis.

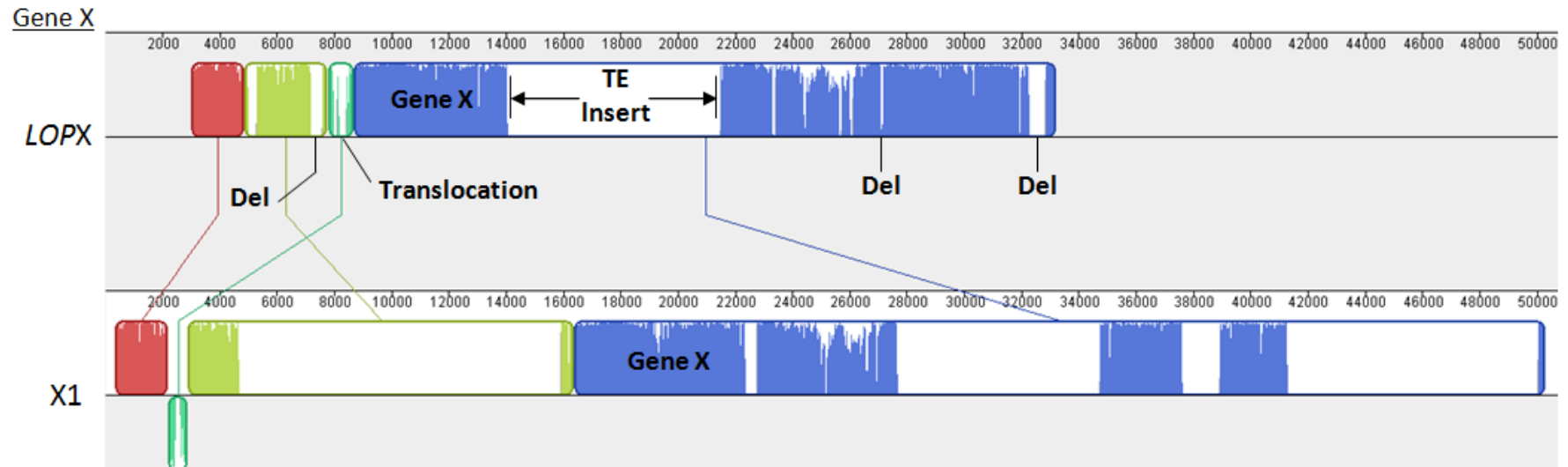


Figure 3.5. Mauve alignment of 25 kb downstream and 25 kb upstream of Gene X between *LOPX* and the X1 alternative allele. Regions of the same colour indicate high sequence similarity and are connected by lines of the same colour, while white areas indicate a lack of sequence similarity. The location of Gene X is labelled. Deletions between two alleles are labelled as ‘Del’ while insertions thought to be caused by retrotransposition are labelled as ‘TE Insert’. An apparent translocation event is also labelled. The distance in bp along the entire alignment is noted on the top axis.

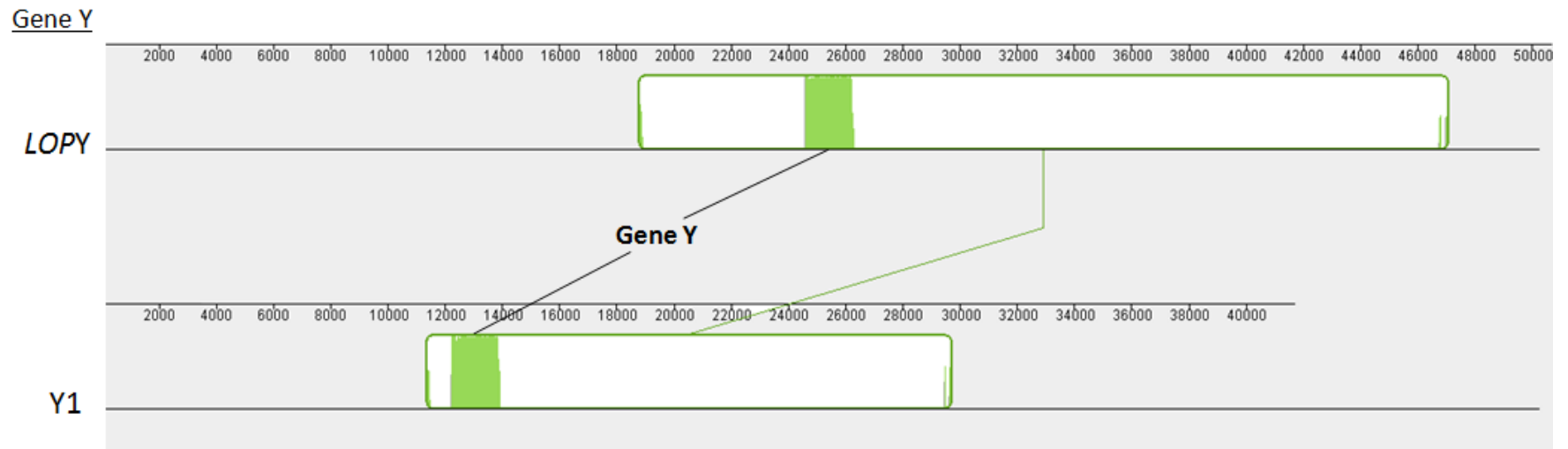


Figure 3.6. Mauve alignment of 25 kb downstream and 25 kb upstream of Gene Y between *LOPY* and the *Y1* alternative allele. Regions of the same colour indicate high sequence similarity and are connected by lines of the same colour, while white areas indicate a lack of sequence similarity. The location of Gene Y is labelled. The distance in bp along the entire alignment is noted on the top axis.

Table 3.2. The number of transposable elements in the ~ 50 kb sequence alignments of Genes B, X and Y within the *LOP* allele and the alternative alleles.

	Gene B	Gene X	Gene Y
<i>LOP</i> Allele	2	9	7
Alt Allele 1	8	6	6
Alt Allele 2	7	-	-

3.4. Discussion

3.4.1. ASD of the *LOP* Allele

Multiple Sequence Alignments

In agreement with the hypothesis, *LOP* allele sequences display ASD as detected by comparing complementary genetic regions at the *LOP* allele to those of the alternative alleles. The sequences of *LOPX* and *LOPB* were considerably longer and possessed a number of unique SNPs. Similarly, when comparing the entire genomic region of *LOPB* to the B1 and B2 alternative alleles, *LOPB* contained 36 unique SNPs not shared by the alternative alleles. These polymorphisms pointed to the isolation and divergence of *LOPB* from the alternative allele sequences over evolutionary time. A similar conclusion was also drawn in another study in which specific alleles were deemed to be undergoing ASD in relation to the other alleles, because of a high level of divergence found within the SFO-12 microsatellite region (Angers & Bernatchez, 1997).

Distance Trees and Distance Matrices

The distance trees also support ASD of *LOPB* and *LOPX*. The *LOP* alleles were located on a long branch, separate from the alternative alleles, further supporting the likelihood of the divergence and isolation of the *LOP* allele over evolutionary time. A study by Corral *et al.*, (2009), on the apomictic species *Boechera holboellii*, drew similar conclusions in which alleles located on isolated branches of a split decomposition network were deemed to have undergone ASD (Corral *et al.*, 2009). The apparent ASD of the *LOP* allele was also shown in the distance matrices used to construct the distance trees, in which a high number of nucleotide substitutions were evident between the *LOP* and alternative allele sequences. This further supports the suggestion that the *LOP* alleles have been isolated and undergone divergence. On the other hand, the low number of nucleotide substitutions between the different alternative alleles indicates a higher degree of sequence similarity and therefore a likely shared recent evolutionary history.

Genomic Sequence Alignments and Transposable Elements

The genomic sequence alignments also exhibited features of ASD in the *LOP* alleles within the flanking regions surrounding the candidate genes. The sequences of the *LOP* alleles were visually diverged from the alternative alleles and appeared to have undergone a greater number of

insertions, deletions and translocations when compared to the alternative alleles. These may have been due to the activity of neighbouring transposable elements (TEs). The greater number of nucleotide alternations in the *LOP* allele is also consistent with ASD (Matzk *et al.*, 2003; Bicknell & Koltunow, 2004; Docking *et al.*, 2006; Van Dijk *et al.*, 2009).

The alignments also showed that *LOPX* and *LOPY* contained a greater number of neighbouring TEs than the alternative alleles. Alleles undergoing ASD are theorised to accumulate TEs as a result of impaired homologous repair mechanisms (Van Dijk *et al.*, 2009). This was also observed in the apomict *Taraxacum*, in which TEs were found to steadily accumulate over time, which negatively impacted upon the fitness of a number of apomict *Taraxacum* clones (Van Dijk *et al.*, 2009).

Conversely, the numbers of TEs were lower nearer *LOPB* than in the alternative alleles of Gene B. One potential reason for the difference in the number of TEs between different genes could be based on the way in which TEs can act within an asexual plant. As discussed previously, while TEs can accumulate in asexual plants because of a lack of purifying selection, it is also possible that they degenerate and gradually become extinct (Hickey, 1982; Wright & Finnegan, 2001).

3.4.2. Causes of ASD of the *LOP* Allele

Repressed Recombination

One of the leading theories to explain the occurrence of ASD is the repression of recombination (Mark Welch *et al.*, 2008). A lack of meiotic recombination, and a possibly lack of involvement of a sister chromatid for DNA repair, will isolate an allele so it will diverge over evolutionary time (Corral *et al.*, 2009). Apomixis loci of species such as *Paspalum* (Labombarda *et al.*, 2002), *Taraxacum* (Van Dijk *et al.*, 2009), *Pennisetum* (Goel *et al.*, 2003) were found to be recombinationally isolated and the *LOP* locus was expected to be similar. Indeed, the findings from Chapter 2 demonstrate that the *LOP* alleles of the candidate genes underwent a reduced rate of recombination in comparison to the alternative alleles. Reduced or repressed recombination is thought to occur because of the process of apomixis. Apomixis represses sexual reproduction and the associated processes of meiotic recombination. The lack of meiosis, and therefore purifying

selection, is postulated to cause the mutation load to steadily accumulate at apomixis loci (Van Dijk *et al.*, 2009). Over time, this will cause the sequence of the alleles within the locus to diverge, known as the Meselson effect (Welch & Meselson, 2000) making homologous recombination with the other alleles outside the apomixis locus impossible (Opperman *et al.*, 2004; Li *et al.*, 2006). At that point, the process is self-reinforcing and the lack of recombination causes further and continuing divergence (Van Dijk *et al.*, 2009). The mechanisms in which recombination is repressed are not fully understood but they may be similar to those identified within the centromere, where recombination is actively repressed to allow for the proper segregation of meiotic chromosomes (Ellermeier *et al.*, 2010). This is undertaken by the Clr4-Rik1 complex, a component of the heterochromatin, which is recruited by an RNA-induced transcriptional silencing complex (Ellermeier *et al.*, 2010; Castel & Martienssen, 2013). Similar mechanisms may also exist at the *LOP* locus, resulting in ASD of the *LOP* alleles. However, further investigation is required.

Hybridization

An alternative explanation for ASD seen at *LOP* is that the locus introgressed into *Hieracium praealtum* as part of an ancestral hybridisation event. In this case, ASD would have arisen because of the different sequence composition of an introgressed allele since it originated in a different species. Previous studies have similarly attributed interspecific introgression as the likely cause of high levels of divergence between alleles of surface antigens in *Plasmodium falciparum* (Roy *et al.*, 2006). Similarly, hybridization with a diverged sexual species was proposed as the cause of ASD of microsatellite loci within the apomict *Boechera holboellii* (Corral *et al.*, 2009). Furthermore, high levels of heterozygosity in asexual lineages of *Rhopalosiphum padi* were attributed to a hybridization event (Delmotte *et al.*, 2002). Hybridization of the *LOP* allele is also consistent with an existing theory of the evolution of apomixis in which this trait may have originated from an ancient hybridization event between species with developmental plasticity within reproduction (Carman, 2007). These unique hybrids may then have undergone polyploidization (increase in chromosome set number), epigenetic modifications and genetic stabilization that eventually gave rise to the apomixis locus (Carman, 2007).

A number of approaches could be undertaken to determine whether ancient hybridisation events are behind the sequence divergence seen at *LOP*. One approach could be the sequence analysis of the alleles of candidate genes in *Hieracium praealtum* and those of close relatives that share

ancestral lineages. Constructing a phylogenetic network using these sequences may reveal an ancient hybridisation event. A similar approach was undertaken in the *Salmonidae* family, where the origin of ancestral microsatellite repeats was determined by comparing allele sequences from a number of related species (Angers & Bernatchez, 1997). Another approach could be to investigate the likelihood of interspecific hybridization within the species under investigation (Delmotte *et al.*, 2002). Within the *Hieracium* genus, interspecific hybridization is well documented (Morgan-Richards *et al.*, 2004; Krak *et al.*, 2013). Examples include the recovery of apomictic hybrids from crosses of *Hieracium caespitosum* and *Hieracium praealtum* (Chapman & Bicknell, 2000) and the identification of multiple hybrids types of *Hieracium pilosella* (Morgan-Richards *et al.*, 2004). Therefore, in the current study, an ancient hybridization event may explain ASD of the *LOP* allele. At the same time there has likely been interplay between potential hybridization events and reduced recombination of *LOP* creating a complex pattern. However, this should be the focus of further investigation before it is accepted.

3.4.3. Origin of *LOPY*

The Mauve alignment of the alleles of Gene Y showed that only the coding region of Gene Y was shared by the Y1 alternative allele and *LOPY*. No other sequence similarity existed between the two alleles in any of the neighbouring DNA. One possible explanation for this is that a copy of Gene Y introgressed into *LOP* and the Y1 allele by the action of retrotransposons, potentially those same TEs identified upstream of *LOPY* and downstream of the Y1 allele. Previous studies in humans have shown that retrotransposons can insert sequences of 2 kb - 46 kb, and these insertions have resulted in a number of genetic diseases (Hancks & Kazazian, 2012). TEs within plants have also demonstrated similar abilities (Feschotte *et al.*, 2002) and may therefore account for the origin of *LOPY* via a retrotransposition event.

3.4.4. Origin of *LOPX* and *LOPB*

In agreement with the research hypothesis, gene synteny was observed between the alleles of Genes B and X, as shown within the multiple sequence alignments and distance trees, which support the likelihood of a shared common ancestor among these alleles. These results also showed

that since the time of a shared common ancestor, *LOPB* and *LOPX* have diverged and undergone independent and isolated evolutionary histories from the alternative alleles. This was based on the *LOP* sequences containing numerous unique SNPs not shared by the alternative alleles and the fact that *LOP* alleles shared an overall low sequence similarity to the alternative alleles.

Further research is needed to strengthen these conclusions. Specifically, future research should seek to generate sequences from all the alternative alleles as well as the complete candidate gene sequence in the relatives and ancestors of *Hieracium*. Comparing these sequences will help to further illuminate the origin of the *LOP* allele. As references earlier, a similar approach was used in a previous study in which the SFO-12 microsatellite sequences were generated from different alleles within the same species, within closely related species and within species in the wider *Salmonids* genus (Angers & Bernatchez, 1997). These sequences were then compared and the ancestral patterns of microsatellite repeats was determined and used to decipher the origin of different alleles (Angers & Bernatchez, 1997). A similar approach could be taken in reference to *LOP* in the future.

3.4.5. Conclusion

In agreement with the research hypothesis, the results from the current study were consistent with the *LOP* alleles displaying ASD. The multiple sequence alignments showed that the *LOP* alleles varied considerably in overall length relative to the alternative alleles and possessed a number of unique SNPs. Similarly, the distance trees support the observation that the *LOP* alleles display ASD as they were located on long branches, separate from the alternative alleles. This most likely indicates the isolation and divergence of the *LOP* alleles over evolutionary time. In addition and consistent with ASD, the genomic sequence alignments of the flanking regions showed that the *LOPB* sequence has undergone a greater number of insertions and deletions than the alternative alleles. The numbers of TEs were also higher in *LOPX* and *LOPY*, which is consistent with ASD of these alleles. On the other hand, the numbers of TEs were lower near *LOPB*. This inconsistency in the number of TEs between different genes may reflect differences in the way TEs can act within an asexual plant in which some can proliferate while others become extinct.

In this study, ASD was thought to be caused by a complex interplay of reduced recombination within the *LOP* locus and the potential occurrence of ancient hybridization events. Both possibilities are supported by the reduced recombination documented in Chapter 2 and the likely occurrence of introgressive hybridization of Gene Y based on the sequence data.

The sequence data also showed gene synteny between the *LOP* alleles and the alternative alleles, which indicated that these alleles likely shared a common ancestral allele, supporting the hypothesis. The sequence data also supports that more recently the *LOP* alleles appeared to have undergone different evolutionary paths. This was demonstrated by the low level of sequence similarity of the *LOP* alleles when compared to the alternative alleles and the fact that the *LOP* sequences contained numerous unique SNPs not found in the alternative alleles. Further research could validate these findings by comparing the sequences of the candidate gene regions in the relatives of *Hieracium praealtum*, which will also help shed light on the events that marked the evolution of *LOP*.

CHAPTER 4

GENERAL DISCUSSION

4.1. Overview

One of the ultimate goals of apomixis research is its introduction into crop species. This will enable the generation of genetically identical seed and the ability to maintain superior hybrid genotypes for successive generations. In turn, these developments have the potential to increase global food production. This is important given a growing world population, shrinking environmental resources and the impending impacts of climate change. Given the benefits and the apparent urgency of introducing apomixis into crop species, the next critical step is to determine how this might be achieved. An essential part of this is to develop an understanding of the genes that control apomixis. With this in mind, the focus of this MSc project was to investigate a number of candidate genes associated with a key component of apomixis, the process of parthenogenesis. Three hypotheses were tested, in order to gain insights into the *LOP* locus, which controls parthenogenesis in *Hieracium praealtum*.

4.1.1. Hypothesis One: ‘Alternative Alleles are Allelic to the *LOP* Allele’

The findings from Chapter 2 confirmed that the Gene X and B alternative alleles are acting like alleles to *LOPX* and *LOPB*. This was shown by all polyhaploids (excluding the recombinants) containing the *LOP* allele and only one other alternative allele. On the other hand, the genotypes of Gene H were not allelic to *LOPH*. This was shown by the polyhaploids classified as ‘H2’ which contained *LOPH* as well as two Gene H genotypes, instead of one.

Co-segregation of Genes X and B

The frequency data of Chapter 2 showed that the alleles of Genes X and B were co-segregating into specific clusters or pairs of alleles. Co-segregation of the regions that carry these genes was also reported in a previous study which focused on deletion-based mapping of *LOP*. Co-segregation is typically an indication of genetic linkage, which can arise in order to maintain a

specific genetic function. In the context of this study, this function may be the process of parthenogenesis, which may only operate as a result of epistatic interactions occurring between these allele combinations. Epistatic interactions may be altering the expression or transcription of Genes X and B or may be necessary to induce a specific function at *LOP*. In a similar way, the combined influence of Genes X and B may be required to overcome an expression threshold necessary for parthenogenesis, which is typical of polygenic inheritance.

4.1.2. Hypothesis Two: ‘*LOP* Alleles Display ASD as a Result of Repressed Recombination at *LOP*’

In agreement with the hypothesis, the findings from Chapter 3 showed that the *LOP* alleles of Genes X and B display ASD in relation to their equivalent alternative alleles. As can be seen from the multiple sequence alignments, where the *LOP* alleles varied in length relative to the alternative alleles and contained a number of unique SNPs. These outcomes suggest isolation and divergence of the *LOP* allele over evolutionary time. ASD of the *LOP* allele was also supported by the distance trees where the *LOP* alleles were located on outlying branches, separate from the alternative alleles. Furthermore, the genomic sequence alignments of the flanking regions surrounding these genes also displayed ASD of the *LOP* allele, which had undergone a higher number of insertions and deletions than the alternative alleles. Also consistent with ASD, the number of identified TEs were higher in *LOPX* and *LOPY*. On the other hand, *LOPB* contained fewer TEs than the alternative alleles. The variation in the number of identified TEs is likely caused by differences related to how TEs can act within asexual plants, where some may proliferate while others become extinct.

Recombination

Recombination was also detected with the *LOP* locus, which was inconsistent with the hypothesis which predicted that recombination would be completely repressed at *LOP*. Based on the findings from this study, ASD of the *LOP* alleles was likely caused by a complex interplay between reduced recombination at *LOP* and potential ancient hybridization events which may have originally introgressed the *LOP* locus into *Hieracium praealtum*.

An additional key finding borne out of genotyping the recombinant plants was that *LOPB* was present in all recombinants and in all polyhaploids in both populations. This provided strong evidence that Gene B is essential for parthenogenesis, which is also supported by the reported functions of Gene B. Gene B shows high similarity to a calcium signalling protein annexin D3, which is involved in stalling the elongation of the pollen tubes which ensures that fertilization is avoided, i.e. the end result of parthenogenesis. Based on these findings, it is clear that Gene B warrants further investigation.

4.1.3. Hypothesis Three: ‘All Alleles Originated From an Ancestral Allele’

The analysis of the sequence data in Chapter 3 showed gene synteny between the *LOP* alleles and alternative alleles of Genes X and B, thereby indicating that these alleles shared a common allele ancestor. The results also showed that more recently the *LOP* alleles underwent ASD, indicating that the *LOP* sequence has been through an isolated and independent evolutionary path from that of the alternative alleles.

Based on the observation of tetrasomic inheritance in the alleles, the recombination and genotype data from Chapter 2 also pointed towards an autopolyploid origin associated with *Hieracium praealtum*. Tetrasomic inheritance was evident from the equal distribution of the X1, X2, B1 and B2 alternative alleles of Genes X and B. One of the proposed reasons for the rarity of the X3 and B3 alleles is that they are suspected to negatively impact the viability of the seed.

4.2. Models of the Evolution of *LOP*

One of the most important questions raised by the findings of the study is how has the *LOP* allele underwent ASD when recombination was documented between all alleles which was expected to eliminate divergence of the *LOP* allele?

In a bid to explain ASD and relate it to the findings from the study, a model is proposed and illustrated in Figure 4.1. It is proposed that looking back in time and in agreement with gene synteny documented in Chapter 3, a common ancestral allele gave rise to the original *LOP* allele and alternative alleles. These alleles were likely present within an ancestral *Hieracium* formed through autopolyploidy ($4n$), which is common among the *Hieracium* genus (Chrtek *et al.*, 2009; Krak *et al.*, 2013). Tetrasomic inheritance of these alleles likely originated in this autopolyploid ancestor, as documented in present-day *Hieracium praealtum* in Chapter 2. The ancestral plant then underwent nondisjunction, in which homologous chromosomes failed to separate correctly in meiosis I or meiosis II during anaphase (Snustad & Simmons, 2006). This could have resulted in some progeny ($1n$) which only contained the

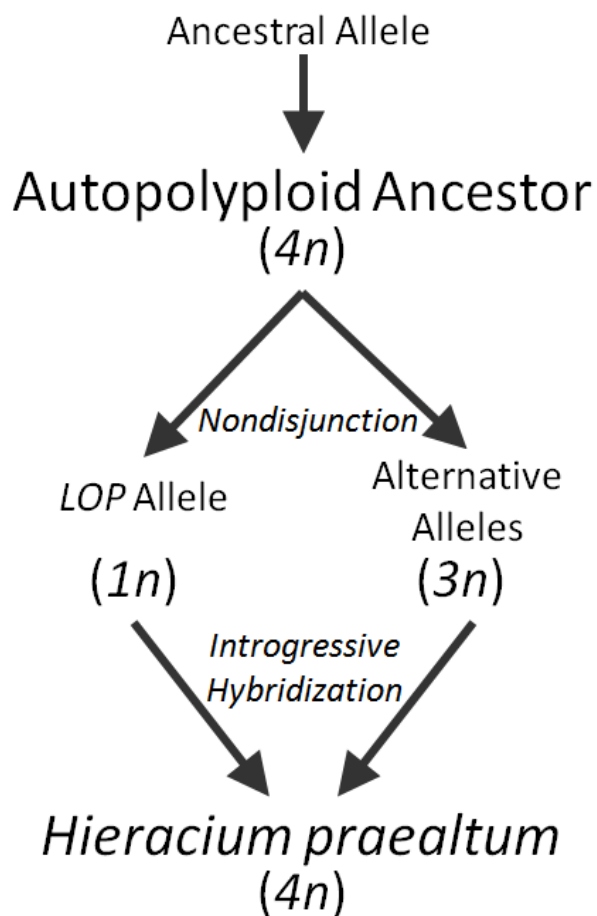


Figure 4.1. Model proposed to explain the occurrence of ASD of the *LOP* allele.

LOP allele while others ($3n$) contained the alternative alleles, see Figure 4.1. In these separate plants, the alternative alleles continued to recombine with one another, while the isolated *LOP* allele diverged and undertook a separate evolutionary path. This is consistent with the findings in Chapter 3 in which the *LOP* alleles varied in sequence composition and length when compared to the alternative alleles and contained a number of unique SNPs not found in the alternative alleles. The four alleles are then thought to have been reunited in the same plant, *Hieracium praealtum*, due to introgressive hybridization. Introgressive hybridization is common among the *Hieracium* genus, and examples include the identification of multiple hybrids types of *Hieracium pilosella* (Morgan-Richards *et al.*, 2004) and the recovery of apomictic hybrids from crosses of *Hieracium caespitosum* and *Hieracium praealtum* (Chapman & Bicknell, 2000). Since this hybridization event, recombination between all four alleles has been restored, as shown by recombinant plants

identified in Chapter 2. In summary, this model proposes that past isolation of the *LOP* alleles from the alternative alleles, caused the present-day ASD of the *LOP* allele, documented in Chapter 3.

Another likely explanation is that the process of apomixis itself could be the cause of ASD of the *LOP* allele, as proposed by Van Dijk *et al.* (2009). Apomixis represses sexual reproduction, which in turn reduces the associated processes of meiotic recombination. Purifying selection will also be reduced, which is expected to cause the mutation load to steadily accumulate at the apomixis loci (Van Dijk *et al.*, 2009). Over time, this will cause the alleles within the apomixis locus, which in this case is the *LOP* allele, to diverge i.e. undergo ASD, known as the Meselson effect (Welch & Meselson, 2000). This will make homologous recombination with the other alleles outside the apomixis locus impossible (Opperman *et al.*, 2004; Li *et al.*, 2006). At that point, the process is self-reinforcing in which the reduction in recombination causes further and continuing divergence (Van Dijk *et al.*, 2009). This process may have continued to such an extent that the *LOP* sequence is evolving independently from the alternative alleles, which may explain the unique SNPs in the *LOP* sequence and difference in sequence composition. This was similarly found in the apomict *Boechera holboellii*, in which the *Het* chromosome associated with apomixis, was found to be evolving independently from the *Del* chromosome in the same plant, based on differences in structure and meiotic dynamics (Kantama *et al.*, 2007). Further research is needed to determine the validity of these models.

4.3. Future Work

An important task to undertake in the future is to generate the sequence data of the complete genetic region of all the alternative alleles of the candidate genes of interest. During the current project, it was not possible to generate the sequence data for the B3, X2, X3, Y2, Y3 alleles and for the Gene H alternative genotypes. One of the many uses of this new sequence data could be to construct new multiple sequence alignments using all the alleles to confirm ASD of the *LOP* allele. Furthermore, the number of TEs, insertions and deletions, could be determined in the alternative allele sequences and compared with *LOP* alleles. Ideally, the equivalent genomic sequences of all these alleles could also be generated in the relatives and ancestors of *Hieracium*. Constructing a

phylogenetic network of these sequences will help to illuminate the origin of the *LOP* allele and look for evidence of ancient autopolyploidy or hybridization.

In order to confirm and build on the genotype data generated in this study, the new sequence data could also be used to design new primers to amplify different regions within the studied candidate genes. A new polyhaploid population could also be generated and genotyped, to see if the distribution of genotypes matches those found in this study and confirm if the same pairs of alleles co-segregate between Genes X and B. Furthermore, new recombinant plants could also be generated and genotyped to confirm the recombination data presented in Chapter 2, in which Gene B did not undergo recombination suggesting that it is most likely essential for parthenogenesis. This potentially pivotal role of Gene B within parthenogenesis could also be the focus of future investigations. These could include studying the regulatory elements within the Gene B sequence, measuring gene expression and mRNA activity during parthenogenesis, and directly analysing and potentially modifying the annexin protein that shows high similarity to Gene B.

4.4. Overall Conclusions

In summary, four major conclusions can be drawn from this study. First, the alternative alleles of Genes X and B are acting as alleles and are segregated from their corresponding *LOP* alleles. Second, the evidence showed *LOPX* and *LOPB* display ASD relative to their equivalent alternative alleles. Third, recombination was detected within the *LOP* locus, which was inconsistent with the hypothesis and inconsistent with the repression of recombination detected in other apomixis loci. Furthermore, Gene B did not undergo recombination which provided strong evidence that it may perform an essential role within parthenogenesis. Fourth, the sequence data suggested that the four detected alleles shared a common allele ancestor, and appear to be undergoing tetrasomic inheritance, indicative of autopolyploidy.

It is suggested that the findings from this project have made a significant contribution towards the future goal of introducing the apomixis trait into crop species and help to target areas for further research.

REFERENCES

- Allen, H.L., Estrada, K., Lettre, G., Berndt, S.I., Weedon, M.N., Rivadeneira, F., Willer, C.J., Jackson, A.U., Vedantam, S., Raychaudhuri, S., et al.** (2010) Hundreds of variants clustered in genomic loci and biological pathways affect human height. *Nature* **467**, 832-838.
- Altschul, S.F., Madden, T.L., Schäffer, A.A., Zhang, J., Zhang, Z., Miller, W., and Lipman, D.J.** (1997) Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Research* **25**, 3389-3402.
- Angers, B., and Bernatchez, L.** (1997) Complex evolution of a salmonid microsatellite locus and its consequences in inferring allelic divergence from size information. *Molecular Biology and Evolution* **14**, 230-238.
- Bednarska, E.** (1989) The effect of exogenous Ca^{2+} ions on pollen grain germination and pollen tube growth. *Sexual Plant Reproduction* **2**, 53-58.
- Bicknell, R.A., and Koltunow, A.M.** (2004) Understanding apomixis: Recent advances and remaining conundrums. *Plant Cell* **16**, S228-S245.
- Bicknell, R.A., Borst, N.K., and Koltunow, A.M.** (2000) Monogenic inheritance of apomixis in two *Hieracium* species with distinct developmental mechanisms. *Heredity* **84**, 228-237.
- Bicknell, R.A., Lambie, S.C., and Butler, R.C.** (2003) Quantification of progeny classes in two facultatively apomictic accessions of *Hieracium*. *Hereditas* **138**, 11-20.
- Bollag, D.M., Edelstein, S.J., and Rozycki, M.D.** (1996). *Protein methods*. (Wiley-Liss New York).
- Budd, M.E., Reis, C.C., Smith, S., Myung, K., and Campbell, J.L.** (2006) Evidence suggesting that Pif1 helicase functions in DNA replication with the Dna2 helicase/nuclease and DNA polymerase δ . *Molecular and Cellular Biology* **26**, 2490-2500.
- Callen, D.F., Thompson, A.D., Shen, Y., Phillips, H.A., Richards, R.I., Mulley, J.C., and Sutherland, G.R.** (1993) Incidence and origin of 'null' alleles in the (AC) $_n$ microsatellite markers. *American Journal of Human Genetics* **52**, 922-927.
- Cantero, A., Barthakur, S., Bushart, T., Chou, S., Morgan, R., Fernandez, M., Clark, G., and Roux, S.** (2006) Expression profiling of the *Arabidopsis* annexin gene family during germination, de-etiolation and abiotic stress. *Plant Physiology and Biochemistry* **44**, 13-24.
- Carman, J.G.** (2007). Do duplicate genes cause apomixis? In *Apomixis: evolution mechanisms and perspectives*. G.U. Hörandl E, van Dijk PJ, Sharbel TF, eds (Rugell: A. R. G. Gantner.), pp. 63-91.
- Castel, S.E., and Martienssen, R.A.** (2013) RNA interference in the nucleus: Roles for small RNAs in transcription, epigenetics and beyond. *Nature Reviews Genetics* **14**, 100-112.
- Catanach, A.S.** (2012). personal communication, R. McGee, ed.
- Catanach, A.S., Erasmuson, S.K., Podivinsky, E., Jordan, B.R., and Bicknell, R.** (2006) Deletion mapping of genetic regions associated with apomixis in *Hieracium*. *Proceedings of the National Academy of Sciences of the United States of America* **103**, 18650-18655.
- Chaparro, C., Guyot, R., Zuccolo, A., Piégu, B., and Panaud, O.** (2007) *RetrOryza*: a database of the rice LTR-retrotransposons. *Nucleic Acids Research* **35**, D66-D70.
- Chapman, H., and Bicknell, R.** (2000) Recovery of a sexual and an apomictic hybrid from crosses between the facultative apomicts *Hieracium caespitosum* and *H. praealtum*. *New Zealand Journal of Ecology* **24**, 81-85.

- Charlson, D.V., Bhatnagar, S., King, C.A., Ray, J.D., Sneller, C.H., Carter Jr, T.E., and Purcell, L.C.** (2009) Polygenic inheritance of canopy wilting in soybean [*Glycine max* (L.) Merr.]. *Theoretical and Applied Genetics* **119**, 587-594.
- Chrtek, J., Zahradníček, J., Krak, K., and Fehrer, J.** (2009) Genome size in *Hieracium* subgenus *Hieracium* (Asteraceae) is strongly correlated with major phylogenetic groups. *Annals of Botany* **104**, 161-178.
- Clough, S.J., and Bent, A.F.** (1998) Floral dip: a simplified method for *Agrobacterium*-mediated transformation of *Arabidopsis thaliana*. *The Plant Journal* **16**, 735-743.
- Conner, J.A., Goel, S., Gunawan, G., Cordonnier-Pratt, M.-M., Johnson, V.E., Liang, C., Wang, H., Pratt, L.H., Mullet, J.E., and DeBarry, J.** (2008) Sequence analysis of bacterial artificial chromosome clones from the apospory-specific genomic region of *Pennisetum* and *Cenchrus*. *Plant Physiology* **147**, 1396-1411.
- Corral, J.M., Piwczynski, M., and Sharbel, T.F.** (2009). Allelic sequence divergence in the apomictic *Boechera holboellii* complex. In *Lost Sex* (Springer), pp. 495-516.
- Darling, A.C., Mau, B., Blattner, F.R., and Perna, N.T.** (2004) Mauve: multiple alignment of conserved genomic sequence with rearrangements. *Genome Research* **14**, 1394-1403.
- Darling, A.E., Mau, B., and Perna, N.T.** (2010) progressiveMauve: multiple genome alignment with gene gain, loss and rearrangement. *PLoS ONE* **5**, e11147.
- Darriba, D., Taboada, G.L., Doallo, R., and Posada, D.** (2012) jModelTest 2: more models, new heuristics and parallel computing. *Nature Methods* **9**, 772-772.
- Delmotte, F., Leterme, N., Gauthier, J.P., Rispe, C., and Simon, J.C.** (2002) Genetic architecture of sexual and asexual populations of the aphid *Rhopalosiphum padi* based on allozyme and microsatellite markers. *Molecular Ecology* **11**, 711-723.
- Dieffenbach, C.W., Lowe, T., and Dveksler, G.S.** (1993) General concepts for PCR primer design. *Genome Research* **3**, S30-S37.
- Distefano, G., Caruso, M., La Malfa, S., Gentile, A., and Wu, S.B.** (2012) High Resolution Melting Analysis Is a More Sensitive and Effective Alternative to Gel-Based Platforms in Analysis of SSR - An Example in Citrus. *PLoS ONE* **7**, e44202.
- Diwan, N., Bouton, J., Kochert, G., and Cregan, P.** (2000) Mapping of simple sequence repeat (SSR) DNA markers in diploid and tetraploid alfalfa. *Theoretical and Applied Genetics* **101**, 165-172.
- Docking, T.R., Saade, F.E., Elliott, M.C., and Schoen, D.J.** (2006) Retrotransposon sequence variation in four asexual plant species. *Journal of Molecular Evolution* **62**, 375-387.
- Dwivedi, S.L., Perotti, E., Upadhyaya, H.D., and Ortiz, R.** (2010) Sexual and apomictic plant reproduction in the genomics era: exploring the mechanisms potentially useful in crop plants. *Sexual Plant Reproduction* **23**, 265-279.
- Ellermeier, C., Higuchi, E.C., Phadnis, N., Holm, L., Geelhood, J.L., Thon, G., and Smith, G.R.** (2010) RNAi and heterochromatin repress centromeric meiotic recombination. *Proceedings of the National Academy of Sciences of the United States of America* **107**, 8701-8705.
- Evsikov, A.V., Graber, J.H., Brockman, J.M., Hampl, A., Holbrook, A.E., Singh, P., Eppig, J.J., Solter, D., and Knowles, B.B.** (2006) Cracking the egg: Molecular dynamics and evolutionary aspects of the transition from the fully grown oocyte to embryo. *Genes and Development* **20**, 2713-2727.
- Feschotte, C., Jiang, N., and Wessler, S.R.** (2002) Plant transposable elements: Where genetics meets genomics. *Nature Reviews Genetics* **3**, 329-341.

- Franklin-Tong, V.E., and Franklin, F.** (2003) The different mechanisms of gametophytic self-incompatibility. *Philosophical Transactions of the Royal Society of London. Series B: Biological Sciences* **358**, 1025-1032.
- Franklin-Tong, V.E., Holdaway-Clarke, T.L., Straatman, K.R., Kunkel, J.G., and Hepler, P.K.** (2002) Involvement of extracellular calcium influx in the self-incompatibility response of *Papaver rhoeas*. *The Plant Journal* **29**, 333-345.
- Gagneux, P., Boesch, C., and Woodruff, D.S.** (1997) Microsatellite scoring errors associated with noninvasive genotyping based on nuclear DNA amplified from shed hair. *Molecular Ecology* **6**, 861-868.
- Garcia-Barrio, M.T., Naranda, T., Vazquez de Aldana, C.R., Cuesta, R., Hinnebusch, A.G., Hershey, J.W.B., and Tamame, M.** (1995) GCD10, a translational repressor of GCN4, is the RNA-binding subunit of eukaryotic translation initiation factor-3. *Genes and Development* **9**, 1781-1796.
- Ge, L.L., Tian, H.Q., and Russell, S.D.** (2007) Calcium function and distribution during fertilization in angiosperms. *American Journal of Botany* **94**, 1046-1060.
- Germanà, M.A.** (2011) Gametic embryogenesis and haploid technology as valuable support to plant breeding. *Plant Cell Reports* **30**, 839-857.
- Gibson, U., Heid, C.A., and Williams, P.M.** (1996) A novel method for real time quantitative RT-PCR. *Genome Research* **6**, 995-1001.
- Goel, S., Chen, Z., Conner, J.A., Akiyama, Y., Hanna, W.W., and Ozias-Akins, P.** (2003) Delineation by fluorescence *in Situ* hybridization of a single hemizygous chromosomal region associated with aposporous embryo sac formation in *Pennisetum squamulatum* and *Cenchrus ciliaris*. *Genetics* **163**, 1069-1082.
- Green, D.M., and Krieg, D.R.** (1961) The delayed origin of mutants induced by exposure of extracellular phage T4 to ethyl methane sulfonate. *Proceedings of the National Academy of Sciences of the United States of America* **47**, 64-72.
- Grimanelli, D., Leblanc, O., Espinosa, E., Perotti, E., González De León, D., and Savidan, Y.** (1998) Mapping diplosporous apomixis in tetraploid *Tripsacum*: One gene or several genes? *Heredity* **80**, 33-39.
- Guindon, S., and Gascuel, O.** (2003) A simple, fast, and accurate algorithm to estimate large phylogenies by maximum likelihood. *Systematic biology* **52**, 696-704.
- Hancks, D.C., and Kazazian, H.H.** (2012) Active human retrotransposons: Variation and disease. *Current Opinion in Genetics and Development* **22**, 191-203.
- Hasegawa, M., Kishino, H., and Yano, T.-a.** (1985) Dating of the human-ape splitting by a molecular clock of mitochondrial DNA. *Journal of Molecular Evolution* **22**, 160-174.
- Hickey, D.A.** (1982) Selfish DNA: a sexuality-transmitted nuclear parasite. *Genetics* **101**, 519-531.
- Hirsch, F.R., Varella-Garcia, M., Bunn, P.A., Di Maria, M.V., Veve, R., Bremnes, R.M., Barón, A.E., Zeng, C., and Franklin, W.A.** (2003) Epidermal growth factor receptor in non-small-cell lung carcinomas: correlation between gene copy number and protein expression and impact on prognosis. *Journal of Clinical Oncology* **21**, 3798-3807.
- Jefferson, R.A., and Bicknell, R.** (1995). The Potential Impacts of Apomixis: A Molecular Genetics Approach. In *The impact of plant molecular genetics*, B.S. Sobral, ed (Birkhäuser Boston), pp. 87-101.
- Jukes, T.H., and Cantor, C.R.** (1969). Evolution of protein molecules. In *Mammalian protein metabolism*, H.N. Munro, ed (New York: Academic Press), pp. 21-123.

- Kantama, L., Sharbel, T.F., Schranz, M.E., Mitchell-Olds, T., de Vries, S., and de Jong, H.** (2007) Diploid apomicts of the *Boechera holboellii* complex display large-scale chromosome substitutions and aberrant chromosomes. *Proceedings of the National Academy of Sciences* **104**, 14026-14031.
- Khush, G., Brar, D., and Bennett, J.** (1998). Apomixis in rice and prospects for its use in heterosis breeding. In *Advances in Hybrid Rice Technology*, S. Virmani, E. Siddiq, and K. Muralidharan, eds (International Rice Research Institute, Manila, Philippines), pp. 297-309.
- Koltunow, A.M.** (1993) Apomixis: Embryo sacs and embryos formed without meiosis or fertilization in ovules. *Plant Cell* **5**, 1425-1437.
- Koltunow, A.M., Hidaka, T., and Robinson, S.P.** (1996) Polyembryony in Citrus. Accumulation of Seed Storage Proteins in Seeds and in Embryos Cultured in Vitro. *Plant Physiology* **110**, 599-609.
- Koltunow, A.M.G., Johnson, S.D., Rodrigues, J.C.M., Okada, T., Hu, Y., Tsuchiya, T., Wilson, S., Fletcher, P., Ito, K., Suzuki, G., et al.** (2011) Sexual reproduction is the default mode in apomictic *Hieracium* subgenus *Pilosella*, in which two dominant loci function to enable apomixis. *Plant Journal* **66**, 890-902.
- Krak, K., Caklová, P., Chrtek, J., and Fehrer, J.** (2013) Reconstruction of phylogenetic relationships in a highly reticulate group with deep coalescence and recent speciation (*Hieracium*, Asteraceae). *Heredity* **110**, 138-151.
- Kühn, C., Thaller, G., Winter, A., Bininda-Emonds, O.R., Kaupe, B., Erhardt, G., Bennewitz, J., Schwerin, M., and Fries, R.** (2004) Evidence for multiple alleles at the DGAT1 locus better explains a quantitative trait locus with major effect on milk fat content in cattle. *Genetics* **167**, 1873-1881.
- Labombarda, P., Busti, A., Caceres, M.E., Pupilli, F., and Arcioni, S.** (2002) An AFLP marker tightly linked to apomixis reveals hemizygoty in a portion of the apomixis-controlling locus in *Paspalum simplex*. *Genome* **45**, 513-519.
- Lander, E.S., and Schork, N.J.** (1994) Genetic dissection of complex traits. *Science* **265**, 2037-2048.
- Lazaris-Karatzas, A., Smith, M.R., Frederickson, R.M., Jaramillo, M.L., Liu, Y.L., Kung, H.F., and Sonenberg, N.** (1992) Ras mediates translation initiation factor 4E-induced malignant transformation. *Genes and Development* **6**, 1631-1642.
- Leblanc, O., Grimanelli, D., Hernandez-Rodriguez, M., Galindo, P.A., Soriano-Martinez, A.M., and Perotti, E.** (2009) Seed development and inheritance studies in apomictic maize-*Tripsacum* hybrids reveal barriers for the transfer of apomixis into sexual crops. *International Journal of Developmental Biology* **53**, 585-596.
- Li, L., Jean, M., and Belzile, F.** (2006) The impact of sequence divergence and DNA mismatch repair on homeologous recombination in *Arabidopsis*. *The Plant Journal* **45**, 908-916.
- Li, Y.C., Korol, A.B., Fahima, T., Beiles, A., and Nevo, E.** (2002) Microsatellites: genomic distribution, putative functions and mutational mechanisms: a review. *Molecular Ecology* **11**, 2453-2465.
- Livak, K.J., and Schmittgen, T.D.** (2001) Analysis of Relative Gene Expression Data Using Real-Time Quantitative PCR and the $2^{-\Delta\Delta C_T}$ Method. *Methods* **25**, 402-408.
- MacKay, J.F., Wright, C.D., and Bonfiglioli, R.G.** (2008) A new approach to varietal identification in plants by microsatellite high resolution melting analysis: Application to the verification of grapevine and olive cultivars. *Plant Methods* **4**.

- Dubeau, M.P., Ghinet, M.G., Jacques, P.E., Clermont, N., Beaulieu, C., and Brzezinski R.** (2009) Cytosine Deaminase as a Negative Selection Marker for Gene Disruption and Replacement in the Genus *Streptomyces* and Other Actinobacteria. *Applied and Environmental Microbiology* **4**, 1211-1214.
- Mahy, G., Bruederle, L.P., Connors, B., Van Hofwegen, M., and Vorsa, N.** (2000) Allozyme evidence for genetic autopolyploidy and high genetic diversity in tetraploid cranberry, *Vaccinium oxycoccos* (*Ericaceae*). *American Journal of Botany* **87**, 1882-1889.
- Mark Welch, D.B., Mark Welch, J.L., and Meselson, M.** (2008) Evidence for degenerate tetraploidy in bdelloid rotifers. *Proceedings of the National Academy of Sciences of the United States of America* **105**, 5145-5149.
- Martin, M.P., Gao, X., Lee, J.-H., Nelson, G.W., Detels, R., Goedert, J.J., Buchbinder, S., Hoots, K., Vlahov, D., and Trowsdale, J.** (2002) Epistatic interaction between KIR3DS1 and HLA-B delays the progression to AIDS. *Nature Genetics* **31**, 429-434.
- Matzk, F., Hammer, K., and Schubert, I.** (2003) Coevolution of apomixis and genome size within the genus *Hypericum*. *Sexual Plant Reproduction* **16**, 51-58.
- McDonald, J.H., and Kreitman, M.** (1991) Adaptive protein evolution at the *Adh* locus in *Drosophila*. *Nature* **351**, 652-654.
- McGee, R.** (2010-2011). Variant sequence analysis of genes governing parthenogenesis in the asexual plant *Hieracium* (Lincoln: Summer Studentship at Plant and Food Research).
- McMenimen, S., and Lubulwa, G.** (1997). Project development assessment: an economic evaluation of the potential benefits of integrating apomixis into hybrid rice, E.E. Unit, ed (Canberra, Australia: ACIAR (Australian Centre for International Agricultural Research)), pp. 1-26.
- Mendel, G.** (1869) *Verh Naturforsch Verein Brünn* **8**, 26-31.
- Morgan-Richards, M., Trewick, S.A., Chapman, H.M., and Krahulcova, A.** (2004) Interspecific hybridization among *Hieracium* species in New Zealand: Evidence from flow cytometry. *Heredity* **93**, 34-42.
- Mortimer, J.C., Laohavisit, A., Macpherson, N., Webb, A., Brownlee, C., Battey, N.H., and Davies, J.M.** (2008) Annexins: multifunctional components of growth and adaptation. *Journal of Experimental Botany* **59**, 533-544.
- Muller, H.J.** (1914) A new mode of segregation in Gregory's tetraploid *Primulas*. *The American Naturalist* **48**, 508-512.
- Nielsen, G.** (1980) Identification of all genotypes in tetraploid ryegrass (*Lolium* spp.) segregating for four alleles in a *Pgi*-enzyme locus. *Heredity* **92**, 49-52.
- Okada, T., Johnson, S.D., Catanach, A.S., Bicknell, R.A., and Koltunow, A.M.G** (2007) An *Hieracium* mutant, *loss of apomeiosis 1 (loa1)* is defective in the initiation of apomixis. *Sexual Plant Reproduction* **4**, 199-211.
- Opperman, R., Emmanuel, E., and Levy, A.A.** (2004) The effect of sequence divergence on recombination between direct repeats in *Arabidopsis*. *Genetics* **168**, 2207-2215.
- Ostenfeld, C.** (1910) Further studies on the apogamy and hybridization of the *Hieracia*. *Zeitschrift für Induktive Abstammungs- und Vererbungslehre* **3**, 241-285.
- Otto, F.** (1990) DAPI staining of fixed cells for high-resolution flow cytometry of nuclear DNA. *Methods in cell biology* **33**, 105-110.
- Ozias-Akins, P.** (2006) Apomixis: Developmental characteristics and genetics. *Critical Reviews in Plant Sciences* **25**, 199-214.

- Perry, G.H., Dominy, N.J., Claw, K.G., Lee, A.S., Fiegler, H., Redon, R., Werner, J., Villanea, F.A., Mountain, J.L., Misra, R., et al.** (2007) Diet and the evolution of human amylase gene copy number variation. *Nature Genetics* **39**, 1256-1260.
- Phillips, P.C.** (2008) Epistasis—the essential role of gene interactions in the structure and evolution of genetic systems. *Nature Reviews Genetics* **9**, 855-867.
- Pinter, S.F., Aubert, S.D., and Zakian, V.A.** (2008) The *Schizosaccharomyces pombe* Pfh1p DNA helicase is essential for the maintenance of nuclear and mitochondrial DNA. *Molecular and Cellular Biology* **28**, 6594-6608.
- Pompanon, F., Bonin, A., Bellemain, E., and Taberlet, P.** (2005) Genotyping errors: Causes, consequences and solutions. *Nature Reviews Genetics* **6**, 847-859.
- Pouchkina-Stantcheva, N.N., McGee, B.M., Boschetti, C., Tolleter, D., Chakrabortee, S., Popova, A.V., Meersman, F., Macherel, D., Hinch, D.K., and Tunnacliffe, A.** (2007) Functional divergence of former alleles in an ancient asexual invertebrate. *Science* **318**, 268-271.
- Purcell, S.M., Wray, N.R., Stone, J.L., Visscher, P.M., O'Donovan, M.C., Sullivan, P.F., Ruderfer, D.M., McQuillin, A., Morris, D.W., O'Gdushlaine, C.T., et al.** (2009) Common polygenic variation contributes to risk of schizophrenia and bipolar disorder. *Nature* **460**, 748-752.
- Quarín, C.L., Norrmann, G.A., and Espinoza, F.** (1998) Evidence for autopolyploidy in apomictic *Paspalum rufum*. *Hereditas* **129**, 119-124.
- Ramsey, J., and Schemske, D.W.** (2002) Neopolyploidy in flowering plants. *Annual Review of Ecology and Systematics*, 589-639.
- Richards, A.** (2003) Apomixis in flowering plants: an overview. *Philosophical Transactions of the Royal Society of London. Series B: Biological Sciences* **358**, 1085-1093.
- Ronald, P.** (2011) Plant genetics, sustainable agriculture and global food security. *Genetics* **188**, 11-20.
- Rosenberg, S.M.** (1907) *Bot Tidsskr* **28**, 143-170.
- Roy, S., Ferreira, M., and Hartl, D.** (2006) Evolution of allelic dimorphism in malarial surface antigens. *Heredity* **100**, 103-110.
- Rozen, S., and Skaletsky, H.** (2000) Primer3 on the WWW for general users and for biologist programmers. *Methods Mol Biol* **132**, 365-386.
- Savidan, Y.** (1999) Apomixis: genetics and breeding. *Plant breeding reviews* **18**, 13-86.
- Savidan, Y., Carman, J.G., and Dresselhaus, T.H.** (2001). The flowering of apomixis: from mechanisms to genetic engineering. (CIMMYT).
- Sherwood, R.T., Berg, C.C., and Young, B.A.** (1994) Inheritance of apospory in buffelgrass. *Crop Science* **34**, 1490-1494.
- Shi, D.-Q., Liu, J., Xiang, Y.-H., Ye, D., Sundaresan, V., and Yang, W.-C.** (2005) *SLOW WALKER1*, essential for gametogenesis in *Arabidopsis*, encodes a WD40 protein involved in 18S ribosomal RNA biogenesis. *The Plant Cell Online* **17**, 2340-2354.
- Shinde, D., Lai, Y., Sun, F., and Arnheim, N.** (2003) Taq DNA polymerase slippage mutation rates measured by PCR and quasi-likelihood analysis: (CA/GT)_n and (A/T)_n microsatellites. *Nucleic Acids Research* **31**, 974-980.
- Smit, A., Hubley, R., and Green, P.** (2013). *Repeatmasker Web Server*. In Version: open-4.0.1, <http://www.repeatmasker.org>.
- Snustad, D.P., and Simmons, M.J.** (2006). *Principles of Genetics*. (Wiley).

- Soltis, D., Soltis, P., and Rieseberg, L.H.** (1993) Molecular data and the dynamic nature of polyploidy. *Critical Reviews in Plant Sciences* **12**, 243-273.
- Southern, E.M.** (1975) Detection of specific sequences among DNA fragments separated by gel electrophoresis. *Journal of molecular biology* **98**, 503-517.
- Taberlet, P., Griffin, S., Goossens, B., Questiau, S., Manceau, V., Escaravage, N., Waits, L.P., and Bouvet, J.** (1996) Reliable genotyping of samples with very low DNA quantities using PCR. *Nucleic Acids Research* **24**, 3189-3194.
- Tiffin, P., and Gaut, B.S.** (2001) Sequence diversity in the tetraploid *Zea perennis* and the closely related diploid *Z. diploperennis*: insights from four nuclear loci. *Genetics* **158**, 401-412.
- Truco, M.J., Ashrafi, H., Kozik, A., van Leeuwen, H., Bowers, J., Wo, S.R.C., Stoffel, K., Xu, H., Hill, T., and Van Deynze, A.** (2013) An ultra-high-density, transcript-based, genetic map of lettuce. *G3 (Bethesda)* **3**, 617-631.
- Tucker, M.R., and Koltunow, A.M.G.** (2009) Sexual and asexual (apomictic) seed development in flowering plants: molecular, morphological and evolutionary relationships. *Functional Plant Biology* **36**, 490-504.
- Valle, C.d., Glienke, C., and Leguizamon, G.** (1994) Inheritance of apomixis in *Brachiaria*, a tropical forage grass. *Apomixis Newsletter* **7**, 42-43.
- Van Dijk, P., de Jong, H., Vijverberg, K., and Biere, A.** (2009). An apomixis-gene's view on dandelions. In *Lost Sex* (Springer), pp. 475-493.
- Van Dijk, P.J., Tas, I.C.Q., Falque, M., and Bakx-Schotman, T.** (1999) Crosses between sexual and apomictic dandelions (*Taraxacum*). I. The breakdown of apomixis. *Heredity* **83**, 715-721.
- Verlhac, M.H., Chen, R.H., Hanachi, P., Hershey, J.W.B., and Derynck, R.** (1997) Identification of partners of TIF34, a component of the yeast eIF3 complex, required for cell proliferation and translation initiation. *EMBO Journal* **16**, 6812-6822.
- Wang, Z., Gerstein, M., and Snyder, M.** (2009) RNA-Seq: A revolutionary tool for transcriptomics. *Nature Reviews Genetics* **10**, 57-63.
- Welch, D.B.M., and Meselson, M.** (2000) Evidence for the evolution of bdelloid rotifers without sexual reproduction or genetic exchange. *Science* **288**, 1211-1215.
- Werle, E., Schneider, C., Renner, M., Völker, M., and Fiehn, W.** (1994) Convenient single-step, one tube purification of PCR products for direct sequencing. *Nucleic Acids Research* **22**, 4354-4355.
- Whitton, J., Sears, C.J., Baack, E.J., and Otto, S.P.** (2008) The dynamic nature of apomixis in the angiosperms. *International Journal of Plant Sciences* **169**, 169-182.
- Winge, Ö.** (1932) On the origin of constant species-hybrids. *Sven Bot Tidskr* **26**, 107-122.
- Wolf, P.G., Soltis, D.E., and Soltis, P.S.** (1990) Chloroplast-DNA and allozymic variation in diploid and autotetraploid *Heuchera grossulariifolia* (*Saxifragaceae*). *American Journal of Botany*, 232-244.
- Wright, S., and Finnegan, D.** (2001) Genome evolution: Sex and the transposable element. *Current Biology* **11**, R296-R299.
- Yu, H.-J., Hogan, P., and Sundaresan, V.** (2005) Analysis of the female gametophyte transcriptome of *Arabidopsis* by comparative expression profiling. *Plant Physiology* **139**, 1853-1869.

APPENDICES

Table A.1 SSR-based primer mixes used to identify BACs containing alternative allele sequences.

Primer Combination	Gene	Primer	Primer sequence	Amplicon size (bp)	TM (°C)
PM 34	B	GeneB_SSR1_for1	TCCTTCTCCTACCCAAGACTCTGAAACT	264	60
		GeneB_SSR1_rev1	CACCTGAAAGCTCAGAACGGAGAGA		
PM 61	Z	GeneZ_1stExon_for2	TGCCAAAGCAACAATGTAAGCCTTTCC	356	60
		GeneZ_1stIntron_rev2	CCAACCTGATGAACACACCAAATTCAGTT		
PM 79	A	GeneA_SSR1_for1	CCACATCACGTCATTGTTCCAGCA	537	60
		GeneA_SSR1_rev2	ACGGCCATAAGGATGTTAGTACTACTACG		
PM 80	X	GeneX_altalleles_for_var2	CCCTTTTTTCTCTCATATATTTATTTATGTA	208	60
		GeneX_VarSpec_rev1	TCACTTCTTCTTTCTACCTTCCTTCTT		

Table A.2. Panel of existing or new SSR primers selected for genotyping. Within ‘Selected for genotyping’ Y refers to yes, N refers to no.

Primer Combination	Gene	Primer	Primer sequence	Amplicon size (bp)	TM (°C)	Selected for genotyping (Y/N)
PM 1	X	GeneX_1stexon_for1 GeneX_2ndexon_rev1	CCATACCGAAGATGTTCTCA TCCACTCTCTGTCCACCAT	346	60	Y
PM 7	A	GeneA_SSR1_for1 GeneA_SSR1_rev1	CCACATCACGTCATTGTTCCAGCA GCTTCTTTCACCATAGCCGCTTCCA	477	60	N
PM 34	B	GeneB_SSR1_for1 GeneB_SSR1_rev1	TCCTTCTCTACCCAAGACTCTGAAACT CACCTGAAAGCTCAGAACGGAGAGA	264	60	Y
PM 54	Z	GeneZ_1stExon_for2 GeneZ_2ndExon_rev1	TGCCAAAGCAACAATGTAAGCCTTTCC TGCCAGCATTGGAATTTCTTTGCCA	415	60	N
PM 81	H	GeneH_2ndExon_for1 GeneH_3rdExon_rev1	TCACTTTCTTAAAAGACGACCAGACCTTTG G ACACGCAGCCTGTGACAGTCGAA	337	60	Y
PM 82	A	GeneA_3rdExon_for1 GeneA_4thExon_rev1	AGGGCGTATCACTCCCTCTTTGAATG CCTTCGTAACGATATGCACTACCAATGC	346	60	Y
PM 83	Y	GeneY_1stExon_for2 GeneY_1stIntron_rev1	TCCAACAATCTTTCAGGAAATAC GCTCATGTCAATTTAGTCATAGTCAC	247	60	Y
PM 84	H	GeneH_3rdExon_for1 GeneH_4thExon_rev2	TTATTCGACTGTACAGGCTGCGT GGCGGTTACCCACAGAAGTGTG	130	60	N
PM 85	B	GeneB_Downstream_for1 GeneB_Reto_rev1	AGCTCTGATGACCTGCCC CATTTGTTTGAAATTC AACAGGAATTC	445	60	N
PM 86	B	GeneB_Downstream_for2 GeneB_Downstream_rev2	TGGGCAGGTCATCAGAGCT CGGTGCTAAAAGTCAACGGG	284	60	N
PM 87	Y	GeneY_Downstream_for1 GeneY_Downstream_rev1	TCCGAACGTTAGATACTGAATGG TGAGGATATAAGATAATGATCAATCCG	229	60	N
PM 88	Y	GeneY_Downstream_for3 GeneY_Downstream_rev3	GCCTGGATCATTATATTGTTCACTAGA CTTTGCCTCAAATAATATAAGACAAATAC	479	60	N

Table A.3. Panel of existing or new HRM primers selected for genotyping. Within ‘Selected for genotyping’ Y refers to yes, N refers to no.

Primer Combination	Gene	Primer	Primer sequence	Amplicon size (bp)	TM (°C)	Selected for genotyping (Y/N)
PM 89	Y	HRM_GeneY_for1	CCTTTACTGAAGATTGCAGGGAGAAGTTG	130	64	N
		HRM_GeneY_var_rev2	ACAATTAATAAACACATTTAACTTTTGGC			
PM 90	Y	HRM_GeneY_for2	AAATACAAGGTATAATTTCTCCTTTACTGA	150	67	N
		HRM_GeneY_var_rev2	ACAATTAATAAACACATTTAACTTTTGGC			
PM 91	Y	HRM_GeneY_for1	CCTTTACTGAAGATTGCAGGGAGAAGTTG	103	68	N
		HRM_GeneY_var_rev1	GGCAAACAAAATTGTGAACTTTTAAACCC			
PM 92	Y	HRM_GeneY_for2	AAATACAAGGTATAATTTCTCCTTTACTGA	346	60	Y
		HRM_GeneY_var_rev1	GGCAAACAAAATTGTGAACTTTTAAACCC			
PM 93	Y	HRM_GeneY_for3	CCTTCCACCAATGGTTAAAAACAC	108	63	N
		HRM_GeneY_rev3	CCAACAAGGATTGCACATGAG			
PM 94	Y	HRM_GeneY_for3	CCTTCCACCAATGGTTAAAAACAC	112	62	N
		HRM_GeneY_rev4	AATCCCAAACAAGGATTGCAC			

Table A.4. Panel of additional *LOP* sequence-specific primers screened on the recombinant plants from the 2011 polyhaploid population.

Primer Combination	Gene	Primer	Primer sequence	Amplicon size (bp)	TM (°C)
PM 2	A	501A19FEfor1	ATCAGGCCGTCAACAATAGG	207	60
		501A19FErev1	GGATTCAGCATCGGAACATC		
PM 3	Z	501A19REfor1	TGAACAGGGGATGAAATGCT	195	60
		501A19RErev1	CGTAATCCCAATCTCCGTTT		
PM 4	X	192I22FEfor8	CTCATTTTGTGATGGCGATG	750	60
		192I22FErev8	TGTTGTTTGACGGACGTTGT		

Table A.5. Genotype data of Genes B, X and H from the 2008 polyhaploid population. Note, the recombinant classes are Gene X genotype 4 and 5, which are highlighted in grey.

Polyhaploid	Gene B Genotype	Allele				Gene X Genotype	Allele				Gene H Genotype	Allele				
		LOPB	B1	B2	B3		LOPX	X1	X2	X3		LOPH	H1	H2	H3	H4
97	2	1	0	1	0	2	1	0	1	0	1	1	1	0	0	0
98	2	1	0	1	0	2	1	0	1	0	1	1	1	0	0	0
99	1	1	1	0	0	1	1	1	0	0	1	1	1	0	0	0
100	3	1	0	0	1	3	1	0	0	1	1	1	1	0	0	0
101	3	1	0	0	1	3	1	0	0	1	1	1	1	0	0	0
103	2	1	0	1	0	2	1	0	1	0	4	1	0	0	0	1
104	3	1	0	0	1	3	1	0	0	1	1	1	1	0	0	0
106	1	1	1	0	0	1	1	1	0	0	1	1	1	0	0	0
107	2	1	0	1	0	2	1	0	1	0	1	1	1	0	0	0
108	3	1	0	0	1	3	1	0	0	1	1	1	1	0	0	0
109	3	1	0	0	1	3	1	0	0	1	1	1	1	0	0	0
110	3	1	0	0	1	3	1	0	0	1	4	1	0	0	0	1
111	2	1	0	1	0	2	1	0	1	0	4	1	0	0	0	1
112	2	1	0	1	0	2	1	0	1	0	1	1	1	0	0	0
113	3	1	0	0	1	3	1	0	0	1	1	1	1	0	0	0
114	2	1	0	1	0	2	1	0	1	0	1	1	1	0	0	0
115	1	1	1	0	0	1	1	1	0	0	1	1	1	0	0	0
116	2	1	0	1	0	2	1	0	1	0	1	1	1	0	0	0
117	1	1	1	0	0	1	1	1	0	0	1	1	1	0	0	0
118	1	1	1	0	0	1	1	1	0	0	1	1	1	0	0	0
119	2	1	0	1	0	2	1	0	1	0	4	1	0	0	0	1
120	2	1	0	1	0	2	1	0	1	0	1	1	1	0	0	0
121	1	1	1	0	0	1	1	1	0	0	4	1	0	0	0	1
122	2	1	0	1	0	2	1	0	1	0	1	1	1	0	0	0
123	1	1	1	0	0	1	1	1	0	0	1	1	1	0	0	0
124	2	1	0	1	0	4	0	0	1	1	1	1	1	0	0	0

125	1	1	1	0	0	1	1	1	0	0	3	1	0	0	1	0
128	3	1	0	0	1	3	1	0	0	1	3	1	0	0	1	0
129	2	1	0	1	0	2	1	0	1	0	1	1	1	0	0	0
130	1	1	1	0	0	1	1	1	0	0	1	1	1	0	0	0
131	1	1	1	0	0	1	1	1	0	0	1	1	1	0	0	0
132	3	1	0	0	1	3	1	0	0	1	4	1	0	0	0	1
133	1	1	1	0	0	1	1	1	0	0	4	1	0	0	0	1
134	2	1	0	1	0	2	1	0	1	0	3	1	0	0	1	0
135	1	1	1	0	0	1	1	1	0	0	1	1	1	0	0	0
136	1	1	1	0	0	1	1	1	0	0	4	1	0	0	0	1
137	2	1	0	1	0	2	1	0	1	0	1	1	1	0	0	0
138	1	1	1	0	0	1	1	1	0	0	4	1	0	0	0	1
139	2	1	0	1	0	2	1	0	1	0	1	1	1	0	0	0
141	3	1	0	0	1	3	1	0	0	1	1	1	1	0	0	0
142	3	1	0	0	1	3	1	0	0	1	1	1	1	0	0	0
143	3	1	0	0	1	3	1	0	0	1	1	1	1	0	0	0
144	2	1	0	1	0	2	1	0	1	0	1	1	1	0	0	0
145	2	1	0	1	0	2	1	0	1	0	1	1	1	0	0	0
146	3	1	0	0	1	3	1	0	0	1	4	1	0	0	0	1
147	2	1	0	1	0	2	1	0	1	0	1	1	1	0	0	0
148	3	1	0	0	1	3	1	0	0	1	1	1	1	0	0	0
150	2	1	0	1	0	2	1	0	1	0	1	1	1	0	0	0
151	1	1	1	0	0	1	1	1	0	0	1	1	1	0	0	0
152	1	1	1	0	0	1	1	1	0	0	1	1	1	0	0	0
153	1	1	1	0	0	1	1	1	0	0	1	1	1	0	0	0
154	3	1	0	0	1	3	1	0	0	1	1	1	1	0	0	0
155	2	1	0	1	0	2	1	0	1	0	1	1	1	0	0	0
156	2	1	0	1	0	2	1	0	1	0	1	1	1	0	0	0
158	2	1	0	1	0	2	1	0	1	0	3	1	0	0	1	0
159	1	1	1	0	0	1	1	1	0	0	1	1	1	0	0	0
160	1	1	1	0	0	1	1	1	0	0	1	1	1	0	0	0

161	2	1	0	1	0	2	1	0	1	0	1	1	0	0	0
162	1	1	1	0	0	1	1	1	0	0	1	1	1	0	0
163	1	1	1	0	0	1	1	1	0	0	1	1	1	0	0
164	2	1	0	1	0	2	1	0	1	0	1	1	1	0	0
165	1	1	1	0	0	1	1	1	0	0	1	1	1	0	0
166	1	1	1	0	0	1	1	1	0	0	1	1	1	0	0
167	1	1	1	0	0	1	1	1	0	0	3	1	0	0	1
168	2	1	0	1	0	2	1	0	1	0	2	1	0	1	0
169	1	1	1	0	0	1	1	1	0	0	1	1	1	0	0
170	2	1	0	1	0	2	1	0	1	0	1	1	1	0	0
171	1	1	1	0	0	1	1	1	0	0	2	1	0	1	0
172	1	1	1	0	0	5	0	1	1	0	1	1	1	0	0
173	1	1	1	0	0	1	1	1	0	0	1	1	1	0	0
174	2	1	0	1	0	2	1	0	1	0	2	1	0	1	0
175	1	1	1	0	0	1	1	1	0	0	4	1	0	0	0
176	2	1	0	1	0	2	1	0	1	0	1	1	1	0	0
177	2	1	0	1	0	2	1	0	1	0	4	1	0	0	0
178	2	1	0	1	0	2	1	0	1	0	1	1	1	0	0
179	2	1	0	1	0	2	1	0	1	0	1	1	1	0	0
181	1	1	1	0	0	1	1	1	0	0	1	1	1	0	0
182	1	1	1	0	0	1	1	1	0	0	1	1	1	0	0
183	2	1	0	1	0	2	1	0	1	0	4	1	0	0	0
184	2	1	0	1	0	2	1	0	1	0	1	1	1	0	0
185	1	1	1	0	0	1	1	1	0	0	1	1	1	0	0
186	1	1	1	0	0	1	1	1	0	0	1	1	1	0	0
187	2	1	0	1	0	2	1	0	1	0	1	1	1	0	0
189	1	1	1	0	0	3	1	0	0	1	1	1	1	0	0
190	2	1	0	1	0	2	1	0	1	0	1	1	1	0	0
191	2	1	0	1	0	2	1	0	1	0	2	1	0	1	0
192	1	1	1	0	0	1	1	1	0	0	1	1	1	0	0
193	2	1	0	1	0	2	1	0	1	0	1	1	1	0	0

194	2	1	0	1	0	2	1	0	1	0	1	1	0	0	0
195	1	1	1	0	0	1	1	1	0	0	1	1	1	0	0
196	1	1	1	0	0	1	1	1	0	0	4	1	0	0	1
197	1	1	1	0	0	1	1	1	0	0	4	1	0	0	1
199	3	1	0	0	1	3	1	0	0	1	1	1	1	0	0
200	1	1	1	0	0	1	1	1	0	0	4	1	0	0	1
201	2	1	0	1	0	2	1	0	1	0	1	1	1	0	0
202	3	1	0	0	1	3	1	0	0	1	1	1	1	0	0
203	2	1	0	1	0	2	1	0	1	0	3	1	0	0	1
204	3	1	0	0	1	3	1	0	0	1	2	1	0	1	0
205	3	1	0	0	1	3	1	0	0	1	2	1	0	1	0
206	1	1	1	0	0	1	1	1	0	0	1	1	1	0	0
207	3	1	0	0	1	3	1	0	0	1	1	1	1	0	0
208	3	1	0	0	1	3	1	0	0	1	2	1	0	1	0
209	3	1	0	0	1	3	1	0	0	1	4	1	0	0	1
210	1	1	1	0	0	1	1	1	0	0	4	1	0	0	1
211	1	1	1	0	0	1	1	1	0	0	4	1	0	0	1
212	3	1	0	0	1	3	1	0	0	1	3	1	0	0	1
213	2	1	0	1	0	2	1	0	1	0	1	1	1	0	0
214	3	1	0	0	1	3	1	0	0	1	4	1	0	0	1
215	2	1	0	1	0	2	1	0	1	0	1	1	1	0	0
216	2	1	0	1	0	2	1	0	1	0	1	1	1	0	0
217	1	1	1	0	0	1	1	1	0	0	1	1	1	0	0
218	1	1	1	0	0	5	0	1	1	0	1	1	1	0	0
219	1	1	1	0	0	1	1	1	0	0	2	1	0	1	0
220	3	1	0	0	1	3	1	0	0	1	1	1	1	0	0
221	1	1	1	0	0	1	1	1	0	0	2	1	0	1	0
222	1	1	1	0	0	1	1	1	0	0	2	1	0	1	0
223	3	1	0	0	1	3	1	0	0	1	1	1	1	0	0
224	2	1	0	1	0	2	1	0	1	0	1	1	1	0	0
225	2	1	0	1	0	2	1	0	1	0	1	1	1	0	0

226	1	1	1	0	0	1	1	1	0	0	3	1	0	0	1	0
227	2	1	0	1	0	2	1	0	1	0	1	1	1	0	0	0
228	2	1	0	1	0	2	1	0	1	0	2	1	0	1	0	0
229	2	1	0	1	0	2	1	0	1	0	1	1	1	0	0	0
230	1	1	1	0	0	1	1	1	0	0	1	1	1	0	0	0
231	1	1	1	0	0	1	1	1	0	0	2	1	0	1	0	0
232	2	1	0	1	0	2	1	0	1	0	2	1	0	1	0	0
233	1	1	1	0	0	1	1	1	0	0	1	1	1	0	0	0
234	1	1	1	0	0	1	1	1	0	0	1	1	1	0	0	0
235	2	1	0	1	0	2	1	0	1	0	1	1	1	0	0	0
236	3	1	0	0	1	3	1	0	0	1	1	1	1	0	0	0
237	1	1	1	0	0	1	1	1	0	0	2	1	0	1	0	0
238	1	1	1	0	0	1	1	1	0	0	1	1	1	0	0	0
239	3	1	0	0	1	3	1	0	0	1	1	1	1	0	0	0
240	3	1	0	0	1	3	1	0	0	1	2	1	0	1	0	0
241	1	1	1	0	0	1	1	1	0	0	1	1	1	0	0	0
242	2	1	0	1	0	2	1	0	1	0	2	1	0	1	0	0
243	1	1	1	0	0	1	1	1	0	0	2	1	0	1	0	0
244	2	1	0	1	0	4	0	0	1	1	2	1	0	1	0	0
245	1	1	1	0	0	1	1	1	0	0	4	1	0	0	0	1
246	1	1	1	0	0	1	1	1	0	0	1	1	1	0	0	0
247	2	1	0	1	0	2	1	0	1	0	1	1	1	0	0	0
248	2	1	0	1	0	2	1	0	1	0	4	1	0	0	0	1
249	2	1	0	1	0	2	1	0	1	0	3	1	0	0	1	0
250	1	1	1	0	0	1	1	1	0	0	2	1	0	1	0	0
251	2	1	0	1	0	2	1	0	1	0	2	1	0	1	0	0
252	1	1	1	0	0	1	1	1	0	0	2	1	0	1	0	0
253	3	1	0	0	1	3	1	0	0	1	1	1	1	0	0	0
254	3	1	0	0	1	3	1	0	0	1	2	1	0	1	0	0
255	1	1	1	0	0	1	1	1	0	0	4	1	0	0	0	1
256	2	1	0	1	0	2	1	0	1	0	3	1	0	0	1	0

257	1	1	1	0	0	1	1	1	0	0	4	1	0	0	0	1
258	1	1	1	0	0	1	1	1	0	0	2	1	0	1	0	0
259	3	1	0	0	1	3	1	0	0	1	1	1	1	0	0	0
260	1	1	1	0	0	1	1	1	0	0	1	1	1	0	0	0
261	1	1	1	0	0	1	1	1	0	0	4	1	0	0	0	1
262	3	1	0	0	1	3	1	0	0	1	1	1	1	0	0	0
263	2	1	0	1	0	2	1	0	1	0	2	1	0	1	0	0
264	1	1	1	0	0	1	1	1	0	0	1	1	1	0	0	0
265	1	1	1	0	0	1	1	1	0	0	1	1	1	0	0	0
266	2	1	0	1	0	2	1	0	1	0	1	1	1	0	0	0
267	1	1	1	0	0	1	1	1	0	0	2	1	0	1	0	0
268	2	1	0	1	0	4	0	0	1	1	1	1	1	0	0	0
269	1	1	1	0	0	1	1	1	0	0	2	1	0	1	0	0
270	1	1	1	0	0	1	1	1	0	0	1	1	1	0	0	0
271	2	1	0	1	0	2	1	0	1	0	1	1	1	0	0	0
273	1	1	1	0	0	1	1	1	0	0	1	1	1	0	0	0
274	1	1	1	0	0	1	1	1	0	0	4	1	0	0	0	1
275	3	1	0	0	1	3	1	0	0	1	1	1	1	0	0	0

Table A.6. Genotype data of Genes B, X, H and Y from the 2011 polyhaploid population. Note, the recombinant classes are Gene X genotype 4 and 5 and Gene H genotype 5, which are highlighted in grey.

Polyhaploid	Gene B Genotype	Allele				Gene X Genotype	Allele				Gene H Genotype	Allele					Gene Y Genotype	Allele			
		LOP B	B1	B2	B3		LOP X	X1	X2	X3		LOP H	H1	H2	H3	H4		LOP Y	Y1	Y2	Y3
301	2	1	0	1	0	2	1	0	1	0	2	1	0	1	0	0	3	1	0	0	1
302	2	1	0	1	0	2	1	0	1	0	1	1	1	0	0	0	1	1	1	0	0
303	2	1	0	1	0	2	1	0	1	0	1	1	1	0	0	0	2	1	0	1	0
304	1	1	1	0	0	1	1	1	0	0	1	1	1	0	0	0	1	1	1	0	0
305	1	1	1	0	0	1	1	1	0	0	2	1	0	1	0	0	1	1	1	0	0
306	1	1	1	0	0	1	1	1	0	0	2	1	0	1	0	0	1	1	1	0	0
307	1	1	1	0	0	1	1	1	0	0	1	1	1	0	0	0	2	1	0	1	0
308	2	1	0	1	0	2	1	0	1	0	1	1	1	0	0	0	2	1	0	1	0
309	2	1	0	1	0	2	1	0	1	0	1	1	1	0	0	0	1	1	1	0	0
310	2	1	0	1	0	2	1	0	1	0	1	1	1	0	0	0	1	1	1	0	0
311	3	1	0	0	1	3	1	0	0	1	2	1	0	1	0	0	2	1	0	1	0
312	2	1	0	1	0	2	1	0	1	0	1	1	1	0	0	0	1	1	1	0	0
313	2	1	0	1	0	2	1	0	1	0	2	1	0	1	0	0	1	1	1	0	0
314	1	1	1	0	0	1	1	1	0	0	4	1	0	0	0	1	1	1	1	0	0
315	2	1	0	1	0	2	1	0	1	0	1	1	1	0	0	0	1	1	1	0	0
316	1	1	1	0	0	1	1	1	0	0	1	1	1	0	0	0	1	1	1	0	0
317	2	1	0	1	0	2	1	0	1	0	3	1	0	0	1	0	2	1	0	1	0
318	1	1	1	0	0	1	1	1	0	0	2	1	0	1	0	0	1	1	1	0	0
319	1	1	1	0	0	1	1	1	0	0	2	1	0	1	0	0	1	1	1	0	0
320	1	1	1	0	0	1	1	1	0	0	1	1	1	0	0	0	1	1	1	0	0
322	2	1	0	1	0	2	1	0	1	0	1	1	1	0	0	0	1	1	1	0	0
323	1	1	1	0	0	1	1	1	0	0	3	1	0	0	1	0	1	1	1	0	0
324	1	1	1	0	0	1	1	1	0	0	1	1	1	0	0	0	1	1	1	0	0
325	1	1	1	0	0	1	1	1	0	0	1	1	1	0	0	0	3	1	0	0	1
326	1	1	1	0	0	1	1	1	0	0	4	1	0	0	0	1	3	1	0	0	1
327	2	1	0	1	0	2	1	0	1	0	1	1	1	0	0	0	1	1	1	0	0

328	1	1	1	0	0	1	1	1	0	0	1	1	0	0	0	1	1	0	0		
329	1	1	1	0	0	1	1	1	0	0	3	1	0	0	1	0	1	0	0		
330	2	1	0	1	0	2	1	0	1	0	1	1	1	0	0	0	2	1	0	1	0
331	2	1	0	1	0	2	1	0	1	0	1	1	1	0	0	0	1	1	1	0	0
332	1	1	1	0	0	1	1	1	0	0	1	1	1	0	0	0	1	1	1	0	0
333	1	1	1	0	0	1	1	1	0	0	1	1	1	0	0	0	1	1	1	0	0
334	2	1	0	1	0	2	1	0	1	0	4	1	0	0	0	1	1	1	0	0	0
335	2	1	0	1	0	2	1	0	1	0	1	1	1	0	0	0	1	1	1	0	0
336	2	1	0	1	0	2	1	0	1	0	4	1	0	0	0	1	2	1	0	1	0
337	1	1	1	0	0	1	1	1	0	0	1	1	1	0	0	0	1	1	1	0	0
338	1	1	1	0	0	1	1	1	0	0	2	1	0	1	0	0	1	1	1	0	0
339	1	1	1	0	0	1	1	1	0	0	1	1	1	0	0	0	1	1	1	0	0
340	2	1	0	1	0	2	1	0	1	0	2	1	0	1	0	0	1	1	1	0	0
341	1	1	1	0	0	1	1	1	0	0	1	1	1	0	0	0	3	1	0	0	1
342	2	1	0	1	0	2	1	0	1	0	1	1	1	0	0	0	1	1	1	0	0
343	2	1	0	1	0	1	1	1	0	0	1	1	1	0	0	0	2	1	0	1	0
344	2	1	0	1	0	2	1	0	1	0	2	1	0	1	0	0	1	1	1	0	0
345	1	1	1	0	0	1	1	1	0	0	1	1	1	0	0	0	2	1	0	1	0
346	1	1	1	0	0	1	1	1	0	0	2	1	0	1	0	0	3	1	0	0	1
347	1	1	1	0	0	1	1	1	0	0	1	1	1	0	0	0	2	1	0	1	0
348	1	1	1	0	0	1	1	1	0	0	1	1	1	0	0	0	1	1	1	0	0
349	1	1	1	0	0	1	1	1	0	0	1	1	1	0	0	0	1	1	1	0	0
350	2	1	0	1	0	2	1	0	1	0	1	1	1	0	0	0	3	1	0	0	1
351	1	1	1	0	0	1	1	1	0	0	4	1	0	0	0	1	2	1	0	1	0
352	2	1	0	1	0	2	1	0	1	0	1	1	1	0	0	0	1	1	1	0	0
353	2	1	0	1	0	2	1	0	1	0	1	1	1	0	0	0	2	1	0	1	0
354	1	1	1	0	0	1	1	1	0	0	1	1	1	0	0	0	2	1	0	1	0
355	2	1	0	1	0	2	1	0	1	0	4	1	0	0	0	1	2	1	0	1	0
356	2	1	0	1	0	2	1	0	1	0	1	1	1	0	0	0	1	1	1	0	0
357	2	1	0	1	0	2	1	0	1	0	1	1	1	0	0	0	3	1	0	0	1
358	1	1	1	0	0	1	1	1	0	0	4	1	0	0	0	1	1	1	0	0	0

359	2	1	0	1	0	2	1	0	1	0	1	1	0	0	0	1	1	1	0	0		
360	1	1	1	0	0	1	1	1	0	0	1	1	0	0	0	1	1	1	0	0		
361	1	1	1	0	0	1	1	1	0	0	3	1	0	0	1	0	1	1	0	0		
362	2	1	0	1	0	2	1	0	1	0	1	1	1	0	0	0	1	1	1	0	0	
363	2	1	0	1	0	2	1	0	1	0	1	1	1	0	0	0	1	1	1	0	0	
364	2	1	0	1	0	2	1	0	1	0	1	1	1	0	0	0	3	1	0	0	1	
365	2	1	0	1	0	2	1	0	1	0	4	1	0	0	0	1	3	1	0	0	1	
366	3	1	0	0	1	2	1	0	1	0	1	1	1	0	0	0	1	1	1	0	0	
367	2	1	0	1	0	2	1	0	1	0	4	1	0	0	0	1	1	1	1	0	0	
368	1	1	1	0	0	1	1	1	0	0	1	1	1	0	0	0	1	1	1	0	0	
369	2	1	0	1	0	2	1	0	1	0	2	1	0	1	0	0	2	1	0	1	0	
370	1	1	1	0	0	1	1	1	0	0	1	1	1	0	0	0	1	1	1	0	0	
371	1	1	1	0	0	1	1	1	0	0	1	1	1	0	0	0	1	1	1	0	0	
372	2	1	0	1	0	2	1	0	1	0	1	1	1	0	0	0	1	1	1	0	0	
373	2	1	0	1	0	2	1	0	1	0	1	1	1	0	0	0	3	1	0	0	1	
374	1	1	1	0	0	1	1	1	0	0	4	1	0	0	0	1	1	1	1	0	0	
375	2	1	0	1	0	2	1	0	1	0	1	1	1	0	0	0	1	1	1	0	0	
376	2	1	0	1	0	2	1	0	1	0	1	1	1	0	0	0	1	1	1	0	0	
377	1	1	1	0	0	1	1	1	0	0	4	1	0	0	0	1	1	1	1	0	0	
378	1	1	1	0	0	1	1	1	0	0	1	1	1	0	0	0	1	1	1	0	0	
379	1	1	1	0	0	1	1	1	0	0	2	1	0	1	0	0	2	1	0	1	0	
380	3	1	0	0	1	4	0	0	1	1	1	1	0	0	0	0	1	0	1	0	0	
381	2	1	0	1	0	2	1	0	1	0	4	1	0	0	0	1	1	1	1	0	0	
382	2	1	0	1	0	2	1	0	1	0	3	1	0	0	1	0	1	1	1	0	0	
383	2	1	0	1	0	2	1	0	1	0	1	1	1	0	0	0	2	1	0	1	0	
384	1	1	1	0	0	1	1	1	0	0	4	1	0	0	0	1	1	1	1	0	0	
385	1	1	1	0	0	1	1	1	0	0	2	1	0	1	0	0	1	1	1	1	0	0
386	3	1	0	0	1	5	0	1	1	0	1	1	1	0	0	0	1	0	1	0	0	
387	2	1	0	1	0	2	1	0	1	0	1	1	1	0	0	0	1	1	1	1	0	0
388	2	1	0	1	0	2	1	0	1	0	1	1	1	0	0	0	1	1	1	1	0	0
389	2	1	0	1	0	2	1	0	1	0	1	1	1	0	0	0	3	1	0	0	1	

390	2	1	0	1	0	2	1	0	1	0	1	1	0	0	0	1	1	1	0	0
391	2	1	0	1	0	2	1	0	1	0	1	1	0	0	0	1	1	1	0	0
392	1	1	1	0	0	5	0	1	1	0	5	0	1	1	0	0	1	0	1	0
393	2	1	0	1	0	2	1	0	1	0	4	1	0	0	0	1	3	1	0	0
394	2	1	0	1	0	2	1	0	1	0	1	1	1	0	0	0	1	1	1	0
395	2	1	0	1	0	2	1	0	1	0	4	1	0	0	0	1	3	1	0	0
396	2	1	0	1	0	2	1	0	1	0	2	1	0	1	0	0	3	1	0	0
397	2	1	0	1	0	2	1	0	1	0	3	1	0	0	1	0	1	1	1	0
398	2	1	0	1	0	2	1	0	1	0	4	1	0	0	0	1	3	1	0	0
399	2	1	0	1	0	2	1	0	1	0	1	1	1	0	0	0	1	1	1	0
400	2	1	0	1	0	2	1	0	1	0	2	1	0	1	0	0	1	1	1	0
401	1	1	1	0	0	1	1	1	0	0	1	1	1	0	0	0	3	1	0	0
402	1	1	1	0	0	1	1	1	0	0	1	1	1	0	0	0	1	1	1	0
403	2	1	0	1	0	2	1	0	1	0	3	1	0	0	1	0	1	1	1	0
404	1	1	1	0	0	1	1	1	0	0	1	1	1	0	0	0	1	1	1	0
405	1	1	1	0	0	1	1	1	0	0	2	1	0	1	0	0	1	1	1	0
406	1	1	1	0	0	1	1	1	0	0	2	1	0	1	0	0	1	1	1	0
407	2	1	0	1	0	2	1	0	1	0	2	1	0	1	0	0	1	1	1	0
408	2	1	0	1	0	2	1	0	1	0	1	1	1	0	0	0	1	1	1	0
409	2	1	0	1	0	2	1	0	1	0	1	1	1	0	0	0	1	1	1	0
410	1	1	1	0	0	1	1	1	0	0	2	1	0	1	0	0	1	1	1	0
411	1	1	1	0	0	1	1	1	0	0	4	1	0	0	0	1	1	1	1	0
412	2	1	0	1	0	2	1	0	1	0	1	1	1	0	0	0	3	1	0	0
413	2	1	0	1	0	2	1	0	1	0	2	1	0	1	0	0	1	1	1	0
414	2	1	0	1	0	2	1	0	1	0	1	1	1	0	0	0	1	1	1	0
415	2	1	0	1	0	2	1	0	1	0	1	1	1	0	0	0	3	1	0	0
416	2	1	0	1	0	2	1	0	1	0	1	1	1	0	0	0	1	1	1	0
417	1	1	1	0	0	1	1	1	0	0	1	1	1	0	0	0	1	1	1	0
418	2	1	0	1	0	2	1	0	1	0	1	1	1	0	0	0	1	1	1	0
419	2	1	0	1	0	2	1	0	1	0	1	1	1	0	0	0	1	1	1	0