INVESTIGATION OF THE GENOMICS OF GENDER REGULATION IN *POPULUS TRICHOCARPA*

by

Nyssa A. Temmel

B.Sc. (Hons) University of Victoria, 2002

A THESIS SUBMITTED IN PARTIAL FULFILLMENT OF

THE REQUIREMENTS FOR THE DEGREE OF

DOCTOR OF PHILOSOPHY

in

THE FACULTY OF GRADUATE STUDIES

(Botany)

THE UNIVERSITY OF BRITISH COLUMBIA

(Vancouver)

December 2011

Abstract

This thesis reports the findings of four projects conducted to study the genomics of gender regulation in *Populus trichocarpa*. Sex-linked markers previously discovered in *Salix vimilanis* were tested to determine if they were also sex-linked in other *Salix* species and *P. trichocarpa.* It was found that the DNA sequence of the SCAR 354 marker, and its position at the 5' end of a gene encoding an Ssu72-like protein, was conserved with some SFP variability in species of *Salix* and *P. trichocarpa*. While this marker may be useful for phylogenetic or population studies in *Salix*, this marker was not sex-linked in the species investigated in this study.

An investigation of genes located on the telomeric end of chromosome 19, the putative sex chromosome in *P. trichocarpa,* was conducted to look for gender-biased SNPs that would indicate recombination suppression in the region on a sex locus. A large variability in the number of SNPs was observed in the gene sequences studied, but no SNPs that segregated with gender were discovered so a genetic marker that could be used to sex *P. trichocarpa* individuals of unknown gender could not be developed.

Using a microarray approach, gender-biased gene-expression was studied in leaf tissue of *P. trichocarpa*. While some gender-biased gene-expression was observed in vegetative tissues the differences observed were statistically insignificant due to biological variation in the samples tested, the small sample size used in this study, and changes in the genome annotation between version 1.1 and 2.0 of the poplar genome. This study could not verify the microarray results using rtPCR in a larger sample of male and female leaf tissue.

MADS-box genes involved in floral development were identified as having gender-biased gene-expression using a microarray approach. Thirteen putative MADS-box genes

that showed gender-biased expression in male and female inflorescences were discovered. Novel expression patterns for nine floral MADS-box genes were identified with this microarray data, and the expression patterns of three of these genes were investigated in further detail using reverse-transcription PCR.

Preface

A version of chapter two has been published. Temmel, Nyssa A., Rai, Hardeep S., and Cronk, Quentin C.B. (2007) Sequence characterization of the putative sex-linked *Ssu72*-like locus in willow and its homologue in poplar. Canadian Journal of Botany 85(11):1092-1097 I wrote the majority of the manuscript and was responsible for all the laboratory work, while Dr. Rai contributed the DNA sequence editing and alignment, and Dr. Cronk assisted with the writing and the research was conducted in his laboratory.

In chapter three and four, Dr. Rick White, the Managing Director of the Statistical Consulting and Research Laboratory at the University of British Columbia, did the statistical analysis of the microarray data.

Collection of the *P. trichocarpa* leaf and floral buds from trees on the UBC Vancouver campus that were used for RNA, DNA and cDNA preparation was assisted by Collin Varner, the Horticulturist/Arborculturist with UBC Plant Operations and Gregg Doughty, the UBC Plant Operations arborist.

Table of Contents

List of Tables

List of Figures

List of Abbreviations

AFLP Amplified Fragment Length Polymorphisms

CAGE Cap Analysis of Gene-expression

MPSS Massively Parallel Signature Sequencing

MSY Male Specific Region

PAR Pseudoautosomal Region

RAPD Random Amplified Polymorphic DNA

SAGE Serial Analysis of Gene-expression

SCAR Sequence Chracterized Amplified Region

SNPs Single Nucleotide Polymorphisms

SSR Simple Sequence Repeat

Acknowledgements

I would like to thank my supervisor, Dr. Quentin Cronk, for his support and guidance throughout my graduate studies, and for encouraging me to explore the field of molecular evolution and development in plants. I would also like to thank my thesis committee, Drs. Sally Aitken, Carl Douglas, and Brian Ellis, for their thoughtful feedback on my research and help in crafting my thesis. I was fortunate to have NSERC support during my degree, so I would like to thank the Natural Sciences and Engineering Research Council of Canada for awarding me grant funding that supported me for four years of my studies.

The UBC Department of Botany was a wonderful place to work and study so I would like to thank the faculty and staff of this department for the many helpful informal conversations that provided assistance, both with my research and with preparing grant applications, applying for TAships and getting all the paperwork of studying in on time.

I was fortunate to work with a great group of laboratory colleagues and fellow graduate students during my studies, so I would like to thank the following people for the assistance they gave me during my degree. Dorothy Cheung and Dr. Hardeep Rai provided excellent technical assistance as I learned new laboratory techniques and worked out research problems, Cindy Sayre and Julia Nowak assisted in collecting leaf samples used for the molecular work in my research, and Hardy Hall helped with the initial statistical analysis of the microarray experiments. Sæmundur Sveinsson helped me perform the phylogenetic analysis of the MADS-box protein sequences, and Dr. Armando Geraldes was instrumental in assisting me with learning to edit DNA sequences and analyze them. Drs. Athena McKown, Ji Yong Yang, Robin Young and Robyn Seipp provided advice and encouragement throughout my degree but especially during the writing of my thesis.

Finally I would like to thank my friends and family, and especially my husband Jason Evans, for reminding me to laugh when my research was not going as well as hoped for, and allowing me to focus on it when it was.

Dedication

This thesis is dedicated to the memory of three people who influenced my decision to pursue a graduate degree. First, to my grandmother, Dr. Ingrid Maria Mathilda Willhelmina Hamburger Temmel, who completed her doctorate in Botany from the University of Graz in July of 1948, and was an inspiration to my pursuing studies in the field of Botany. Also, to my Gr. 9 and 10 science teacher Mr. David Stacey, who taught me that science was fun, and that, if you were lucky, learning was something you never stopped doing. And finally, I'm dedicating my thesis to the memory of my good friend Sandra Patricia Hunter, who passed away too soon, but lived her life with passion, courage and wisdom, and always encouraged me to follow my dreams.

Chapter 1: Introduction

1.1 Sexual systems in plants

Gender-determination in animals and plants is a developmental process that results in the physical separation of female and male gamete production, on the same or on separate individuals. Approximately 90% of the world's angiosperms have hermaphroditic or "perfect" flowers, which contain both male and female reproductive parts (Ainsworth, 2000). The remaining 10% of the flowering plants produce various combinations of male, female or hermaphroditic flowers on the same or separate plants. The most common combinations are, first, monoecy, where male and female flowers are present on the same plant and, secondly, dioecy, which occurs when male and female flowers are produced on separate plants (Ainsworth, 2000). Dioecious species account for 6%, or 14 000 species of all flowering plants (Renner and Won, 2001). Many plant species have evolved other combinations of flower types as well, such as gynodioecy, gynomonoecy, trimonoecy, androdioecy, andromonoecy and trioecy, but these combinations are less common and are usually considered to be intermediate forms in the evolution of unisexual flowers, which contain either male or female reproductive organs (Ainsworth, 2000) (Figure 1.1).

These combinations of sexual systems in plants have been well described morphologically and physiologically, and there is some understanding of the possible ecological reasons for their evolution, such as avoidance of inbreeding depression caused by self-pollination (Charlesworth and Charlesworth, 1979). However, how gender is determined in plants at the molecular level or what the evolutionary pathway(s) for the development of separate genders might have been is a growing field of research

facilitated by the advent of genome sequencing. Research into gender-determination in dioecious plants has generally been focused around plant species that have well-developed heteromorphic sex chromosomes such as *Silene* or *Rumex* (Ainsworth et al., 1995; Filatov et al., 2000).

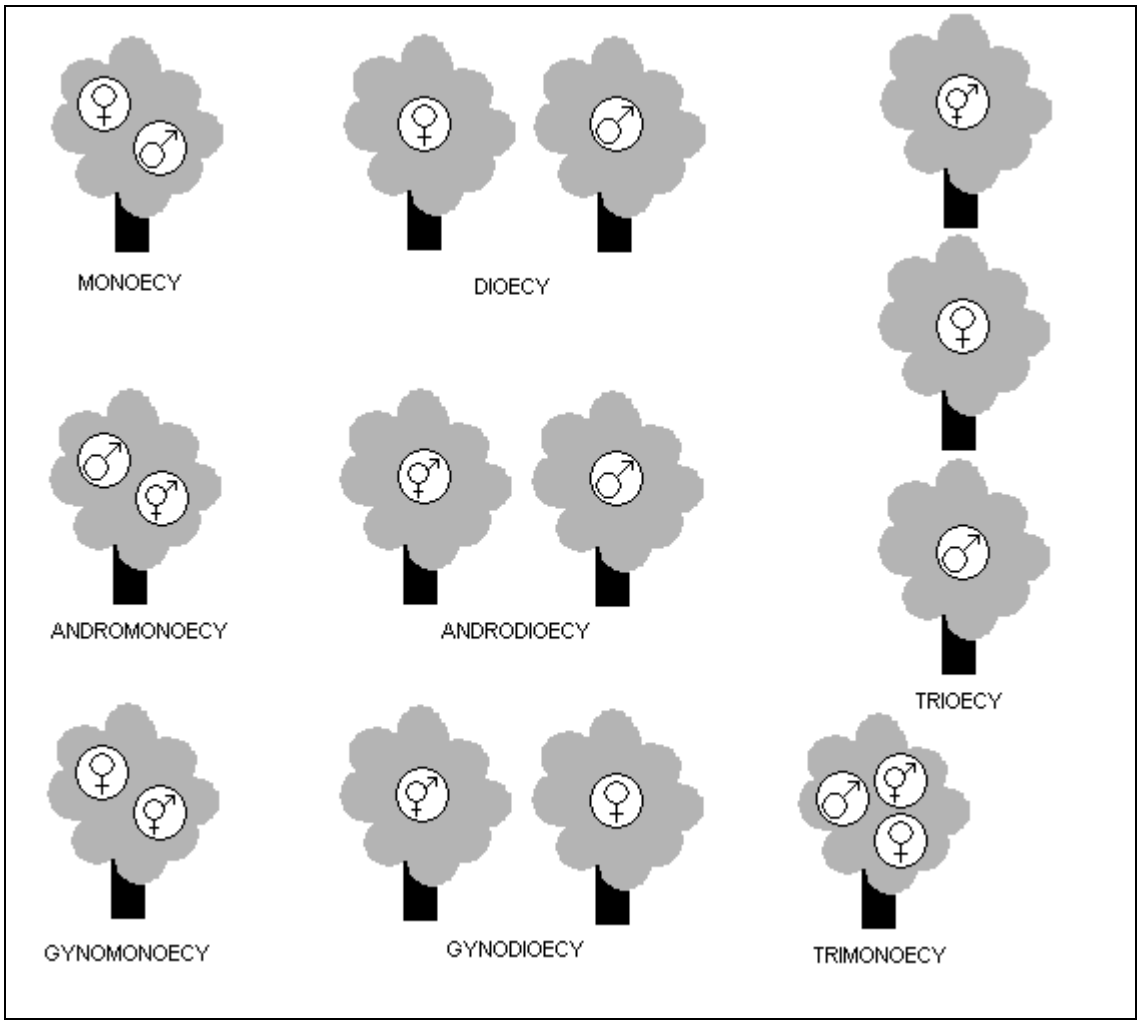

Figure 1.1 Sexual systems in plants.
Circles represent floral forms, and plants that produce only hermaphroditic flowers are not shown.

1.2 Evolution of dioecy

1.2.1 Gynodioecious pathway

Approximately 160 plant families include dioecious species and it is estimated that dioecy has evolved over 100 separate times (Charlesworth and Guttman, 1999). Dioecy is thought to have evolved through several evolutionary pathways. In the gynodioecy pathway, it is theorized that female plants arise by mutations that produce male sterility in hermaphroditic plants. These spread through populations of hermaphroditic plants because they produce out-crossed, and therefore fitter, offspring (Barrett, 2002). Selection then acts on the remaining hermaphroditic plants to reduce female function so that male plants are produced. It is most likely that dioecy evolved via gynodioecy or androdioecy as an intermediate stage because the initial step in this process is a single mutation resulting in either female or male sterility (Charlesworth and Charlesworth, 1978). The probability of dioecy evolving directly from a hermaphrodite is small because it is unlikely that two independent mutations, one for female and one for male sterility, would arise simultaneously and become tightly linked on the genome to ensure a stable dioecious system (Ainsworth, 2000). A species that provides an opportunity to study this is *Carica papaya*, where comparisons of silent mutation rates observed on the Y and Y^h chromosomes indicate that the dioecious and gynodioecious breeding systems in this plant diverged approximately 73 000 years ago (Yu et al., 2008). Recent research with *Fragaria virginiana* Mill., which has both subdioecious and gynodioecious sexual systems, may also provide insight into how the transition from gynodioecy to dioecy occurred (Spigler et al., 2008).

1.2.2 Heterostyly pathway

In a number of species dioecy has probably evolved from heterostyly, whereby male flowers are derived from short-styled flowers and female flowers from long-styled flowers (Bawa, 1980). There are no fewer than 13 plant families where distyly has been observed (Charlesworth and Charlesworth, 1979), and plant species exhibiting this are characterized by pin flowers with long styles and short stamens and thrum flowers having short styles and long stamens (Beach and Bawa, 1980). The two flower types are produced in a population as if they are controlled by alleles at a single locus, where the pin flower phenotype is recessive (Charlesworth and Charlesworth, 1979). The genetic model put forward to explain the evolution of dioecy from heterostyly proposes the existence of a two-gene linkage model whereby a dominant gene for producing thrum flowers is linked to a recessive gene that causes the abortion of female organs (Muenchow and Grebus, 1989). However, the most probable force driving the change from distyly to dioecy is a change in pollinators. Reciprocal pollen transfer ensures the gene flow between the two floral types, so if the pollination system changes, gene flow ceases (Beach and Bawa, 1980). Unidirectional pollen flow (pin to thrum, and thrum to pin only) then facilitates the change from functional unisexuality to structural unisexuality as mutations from male sterility (pin flowers) and female sterility (thrum flowers) occur and spread through the population (Beach and Bawa, 1980). It is unlikely that dioecy evolved from heterostyly in *P. trichocarpa* because the flowers of this species show no evidence of heterostyly, and the genus is wind-pollinated.

1.2.3 Monoecious pathway

Dioecy has also evolved through monoecy where it is thought that disruptive selection acted on existing genetic variation in floral gender ratios in monoecious plants, which resulted in more specialization of the genders (Barrett, 2002). First, male fertility could be reduced through mutations that either reduce pollen production or convert male flowers to female ones (Charlesworth and Charlesworth, 1978). These kinds of mutations would reduce selfing in a population and increase female fertility. Secondly, a mutation that results in reduced fertility in the female plants would have to take place. A second wave of mutations resulting in reduced female fertility is unlikely to spread through the population of remaining monoecious plants unless it is tightly linked to the male fertility locus because such a mutation would be disadvantageous to the females (Charlesworth and Charlesworth, 1978). Dioecy could therefore have evolved from monoecious species through a process of alternating mutations that reduced male and female fertility over time (Charlesworth and Charlesworth, 1978).

1.2.4 Evolution of dioecy in *Populus*

There is a strong association between the occurrence of dioecy and the presence of monoecious species in related genera. This may be because once a group of species has acquired the ability to genetically suppress male or female function in flowers within an individual, the jump to dioecy can occur through changes in floral gender ratios on individual plants (Renner and Ricklefs, 1995). Eventually these genetic differences that cause differentiated expression of gender in individuals become fixed between males and females at “sex loci”.

From the distribution of sexual systems in related genera, it appears that dioecy in *Populus* evolved via the monoecious pathway. The genus *Populus* is classified in the angiosperm Eurosid I clade, in the order Malpighiales (Jansson and Douglas, 2007a). The order Malpighiales contains approximately 16 000 species divided into 39 families (Stevens, 2001), including the family Salicaceae, where *Populus* is placed. The family Salicaceae contains nine genera (*Bennettiodendron*, *Carrierea*, *Idesia*, *Itoa*, *Macrohasseltia*, *Olmediella*, *Poliothyris*, *Populus* and *Salix*), which are predominantly dioecious or monoecious and insect-pollinated except the hermaphroditic *Macrohasseltia*, and the wind-pollinated *Populus* (Cronk, 2005).

1.3 Sex chromosomes in plants

When it comes to gender-determination, gonochorism, or sexual reproduction between unisexual individuals, is the most prevalent system in animals, and is regulated by the presence of sex chromosomes in almost all cases (Janousek and Mrackova, 2010). Sex chromosomes are chromosomes that are involved in the genetic determination of gender in an organism. In the majority of plant species that are dioecious, the genetic mechanism that determines gender is not well understood, and not all dioecious species have identified sex chromosomes. Sex chromosomes in plants appear to be less evolved than those observed in animals, so studying them in plants provides an opportunity to understand the processes involved in their evolution from autosomes (Moore, 2009), as well as the exploring of how gender-determination is controlled genetically in the absence of sex chromosomes.

1.3.1 Identification and description of sex chromosomes

Although sex chromosomes were discovered around the same time in both plants and animals early in the 20th century, knowledge about sex chromosome structure and function in plants has lagged behind our understanding of the role of sex chromosomes in animals such as *Caenorhabditis elegans*, *Drosophila* sp., *Homo sapiens* and *Mus musculus* (Vyskot and Hobza, 2004). As noted above, the fact that many sex chromosomes are heteromorphic from autosomes, and from each other, facilitated their discovery. This heteromorphy is related to the two sex-determination systems that have been observed in most sexually reproducing organisms. In heterogametic sex-determining systems, the gender of an individual is determined by either the presence of a Y chromosome (XY system), or the W chromosome when females are heterogametic (ZW system) (Vyskot and Hobza, 2004). In the second system sex is determined by the ratio between the number of sex chromosomes and the number of autosomes, the X/A ratio (Jamilena et al., 2008). Since sex chromosomes were first described, the basic definition of what constitutes a sex chromosome has expanded to state that sex chromosomes do not recombine during meiosis, along at least a portion, if not the entirety of their length, and that sex-determining genes that control male and female fertility are located on them (Vyskot and Hobza, 2004).

1.3.2 A brief history of the discovery of sex chromosomes

The genetic basis for sex-determination remained a mystery until the early years of the 20th century. One of the first indications that genetic control of sex-determination was associated with chromosomes came in 1891 when Henking, while studying the insect *Pyrrhocorus apterus*, noticed a chromatin body in the spermatocyte that was larger than

the other chromatin bodies, and stained more darkly than the others (Anderson, 2000). These first observations were confirmed by subsequent researchers but no one was yet prepared to state that the sex of progeny was determined exclusively by chromosomes at the time of fertilization, and it wasn't until 1902 that McClung postulated that the "accessory chromosome" identified by Henking could play a part in sex-determination (Brush, 1978). Conclusive evidence for the connection between chromosomes and sex-determination came in 1905 when research on *Tenebrio molitor* (meal worm, n = 20) done by Dr. Nettie Stevens found that sex-determination must be controlled by chromosomes, as female somatic cells contained 20 large chromosomes, and the same cells in males contained one small, and 19 large chromosomes (Brush, 1978). The same year that Stevens discovered the sex-determination system that would become known as XY, where the males (Y) are the heterogametic gender, work on a different insect species, *Anasa tristis*, also inferred that chromosomes were responsible for sex-determination, through a XO sexual determination system, where an X chromosome is present in the females, and males do not have this extra X chromosome (Brush, 1978). Between 1905 and 1909 the terms "heterchromosome", "autosome", "sex chromosomes", and "X" and "Y" chromosomes came into use to describe the newly established link between "accessory chromosomes" and sex-determination in many organisms (Anderson, 2000).

1.3.3 Discovery of sex chromosomes in plants

Conclusive evidence of the existence of sex chromosomes and their role in sex-determination in plants was established in 1917 from research done on two species of liverwort, *Sphaerocarpos donnellii* Aust. and *S. texanus* Aust. (Anderson, 2000). The

haploid gametophyte of these species have eight chromosomes; in females the karyotype consists of seven autosomes and an X chromosome that determines sex, and in males there are seven autosomes and a Y chromosome (Allen, 1917). Heteromorphic sex chromosomes in dicotyledonous plants were first observed in the dioecious species *Lychnis alba* (now named *Silene latifolia*) where evidence for XX females and XY males was reported (Blackburn, 1923). However, plants with separate genders do not usually have sex chromosomes and the mechanism and control of gender identity in the majority of plants remains obscure, despite the numerous studies that have investigated the role of sex chromosomes in determining sex in haploid, diploid and triploid plants (Sakamoto et al., 2005).

1.3.4 Evolution of sex chromosomes

The evolution of sex chromosomes has probably been driven by the need to limit recombination between genes that determine gender (Ainsworth, 2000). In those plants with sex chromosomes, the X-Y system is thought to have evolved more recently than in animals. Evidence suggests that the X-Y system only arose approximately ten million years ago in *Silene* spp., whereas mammals developed sex chromosomes around 166 million years ago (Veyrunes et al., 2008). Birds have a Z-W system of sex chromosomes that evolved independently (Hake and O'Conner, 2008). Of the estimated 14 620 dioecious plant species only 13 species in five families are known to have heteromorphic sex chromosomes (Jamilena et al., 2008; Ming et al., 2007), and an additional 13 plant families containing approximately 16 additional species exhibit homomorphic sex chromosomes (Ming et al., 2007). In mammals the degeneration of gene content and size of sex chromosomes through accumulation of deleterious mutations and subsequent gene

loss due to their long evolutionary history is characteristic, smaller size does not appear to be a genomic feature of young sex chromosomes in plants dues to local duplications and transposable element insertions (Ming and Moore, 2007). However, where they exist, the genetic mechanisms by which sex chromosomes control sex-determination, either via an active Y chromosome, or by the balance of the number of X chromosomes with the number of autosomes, are the same in plants and in animals (Jamilena et al., 2008).

1.3.5 Proposed mechanism for the evolution of sex chromosomes

Sex chromosome evolution is a multiple step process, one of the first processes involved being the suppression of genetic recombination surrounding one or more sex-determining genes at a sex locus (Liu et al., 2004). At some point in the evolutionary history of the plant, a male or female sterility mutation occurs on a nascent sex chromosome, leading to suppression of recombination happens at the locus and its immediate neighbouring region. This process can be observed in *Asparagus officinalis* L. (2n = 20), in which the chromosome pair L5 have been identified as the homomorphic sex chromosomes (Reamon-Büttner et al., 1998). In this species sexual dimorphism is controlled by a dominant gene *M*, with female plants being homozygous for the recessive allele (*mm*), and male plants being either homozygous (*MM*) or heterozygous (*Mm*) at the sex locus (Jamsari et al., 2004). In many organisms that have a X-Y mating system, the sex chromosomes contain a pseudoautosomal region (PAR) where X and Y still recombine (Nicolas et al., 2005). This is the case in the dioecious crop species *Spinacia oleracea* (spinach) (Khattak et al., 2006), and asparagus (Telgmann-Rauber et al., 2007), where both species have homomorphic sex chromosomes that recombine along most of

their length with little or no measurable degeneration of the Y sex chromosome. In *Carica papaya* where a sex locus has also been identified, gender is determined by one chromosome region with three allelic forms, combinations of which result in the male, female and hermaphroditic phenotypes (Liu et al., 2004). The sex locus is distinguished on the homologous autosomes as a region of increased polymorphism with a low level of recombination, known as the male-specific region, or MSY (Liu et al., 2004). A single locus with three alleles also determines gender in *Ecballium elaterium* (Ainsworth, 2000). Species that have heteromorphic chromosomes can also exhibit this kind of sex-determining system, as is the case in *Humulus lupulus* (European cultivated hops), where analysis of sex-linked markers has revealed that suppression of recombination between the X and Y chromosomes is localized to the region around the sex-determining locus (Seefelder et al., 2000).

Once a sex-determining locus has formed on an autosome, the suppression of recombination observed around the sex locus can spread along the length of the autosome, and the proto Y chromosome will slowly accumulate deleterious mutations, along with favorable alleles, which will simultaneously reduce the size and functional gene content of the Y while selecting for male advantageous mutations via "Müller's Ratchet" and genetic hitchhiking (Rice, 1987). In the genus *Silene* three species, *S. latifolia*, *S. dioica* and *S. diclinis,* have a heteromorphic X-Y sex-determination system where the Y chromosome determines maleness (Nicolas et al., 2005). In these *Silene* species the correlation between the synonymous divergence between the X and Y chromosomes and the distance from the PAR on the X chromosome genetic map indicates that suppression of recombination between the sex chromosomes started at an

ancient sex-determining locus and moved towards the current PAR (Nicolas et al., 2005). Now with the identification of eleven Y-linked loci with homologs on the X chromosome, researchers have confirmed that the X and Y sex chromosomes in these species stopped recombining at between ten and twenty million years ago (Bergero et al., 2007). The accumulation of chloroplast DNA sequences on the Y chromosome in *S. latifolia* has also contributed to the degeneration of genes (Kejnovsky et al., 2006).

Rumex acetosa has heteromorphic sex chromosomes, but in this dioecious species sex-determination is controlled by the X/A ratio, where a ratio between numbers of X chromosomes and autosomes controls flower gender (Jamilena et al., 2008). In *Rumex acetosa* female plants have two X chromosomes (2n = 14, XX) and male plants have one X and two different Y chromosomes (2n = 15, XY_1Y_2) (Ming et al., 2007). In this species the non-recombining regions of the Y chromosomes are heterochromatic and may have become specialized due to an accumulation of repetitive DNA sequences (Shibata et al., 1999). Therefore, the Y chromosomes in *Rumex acetosa* represent an advanced stage of sex chromosome evolution (Mariotti et al., 2006).

1.3.6 Degeneration of the Y chromosome

The current theory of sex chromosome evolution is that Y chromosomes shrink in size, and genetic information is lost as they evolve (Ming and Moore, 2007). However, given that sex chromosomes in plants have evolved more recently than those in animals, the smaller size of the Y chromosome is not a universal genomic feature. The accumulation of repetitive DNA sequences, either tandem repeats like those seen in *Rumex* sp., or transposable elements, could account for the increased size of the Y chromosome in plants that have heteromorphic sex chromosomes (Jamilena et al., 2008).

In *Silene latifolia* and *Rumex acetosa*, where sex chromosomes are most highly evolved, the Y chromosomes have not degenerated to a point where non-functional DNA sequences would be lost. Neither of these species have accumulated transposable elements on their respective Y chromosomes (Jamilena et al., 2008), unlike the Y chromosomes of *Cannabis sativa* and *Marchantia polymorpha*. In *C. sativa* the Y is twice the size of the X due in large part to the accumulation of a specific Long INterspersed Element-like (LINE-like) retrotransposon on the terminal end of the long arm of the Y chromosome, which may play a role in suppressing pairing or recombination with the X chromosome (Sakamoto et al., 2000). Unlike in animals where a considerable amount of information on the gene organization on the Y chromosome is available, due in large part to work in some primates and *Drosophila*, data on gene sequence and organization on the Y chromosomes of plants is not as available. In the liverwort *Marchantia polymorpha* the haploid genome consists of eight autosomes and either an X chromosome in females (n = 8 + X) or a Y in males (n = 8 + Y) (Yamato et al., 2007). The X and Y chromosomes do not recombine along their entire length, and 64 genes have been identified on the Y chromosome, 14 of which are expressed in reproductive organs and are only found in the male genome (Yamato et al., 2007). Active Y chromosome specific genes were first reported in this plant species, and while repetitive DNA sequences are a common feature of Y chromosomes in the plant listed above and in animals, the nucleotide sequences that make up these repeats appear to be unique to each species studied, indicating that Y chromosomes have arisen independently in multiple taxa as a sex-determining system (Okada et al., 2001).

1.3.7 The absence of sex chromosomes in *Populus*

There have been no sex chromosomes identified in *Populus trichocarpa*, although a putative sex locus has been mapped by using sex-linked genetic markers on chromosome 19 in the genomes of six poplar species (including *P. trichocarpa*) (Paolucci et al., 2010). The genus *Populus* may have a similar sex-determination system as that found in *Fragaria* sp. (strawberry) as species in this genus do not share the same sex-determination system, they may share a single origin for a sex-determining region in a common ancestor (Goldberg et al., 2010).

1.4 The genetics of gender

The developmental processes that drive the differentiation between males and females are controlled through gene-expression patterns in response to internal or environmental cues, and in plants this results in a large variety of sexual phenotypes. Given the myriad of sexual systems that exist in plant families, and the observation that these systems have evolved multiple times, there appear to be many ways that plants have evolved genetic controls for the development of sexual organs and mating types. Until recently, the lack of genetic data for plant species with diverse sexual systems has prevented a better understanding of how genetics relates to sexual phenotypes. Now that the genes involved in gender-determination have been identified in multiple species, it appears that these genes can be divided into three functional groups: gender-determining genes, gender-differentiating genes, and gender-associated genes (Figure 1.2).

Gender-determining genes are located at a sex locus and are fixed as homozygous in one sex and heterozygous in the other so that they segregate with gender. Gender-differentiating genes can be located throughout the genome and are responsible for the

development of sexual organs and reproductive functions in the plant - primary sexual characteristics. The products of gender-determining genes act upon this group of genes. Gender-associated genes play no direct role in gender function, but may result in secondary sexual characteristics that are associated with one or the other gender.

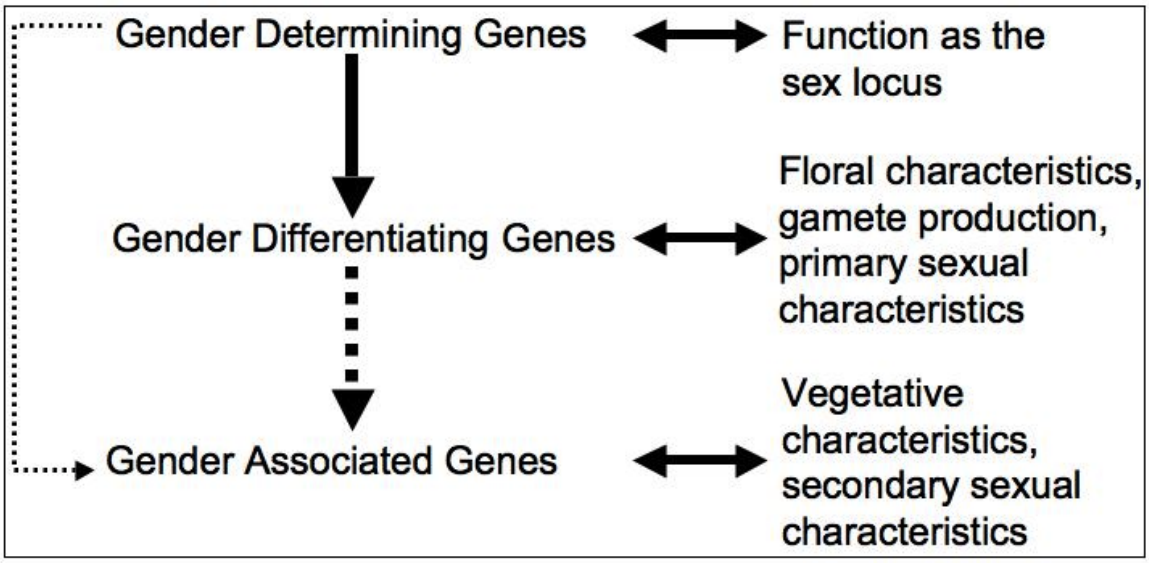

Figure 1.2 Three functional groups of genes are involved in gender-determination in plants.
Solid arrows represent direct relationships between gene groups, and dashed arrows indicate indirect relationships between the gene groups.

1.4.1 Gender-determining genes

Gender-determining genes are regulatory genes, which exist as homozygotes in one sex and heterozygotes in the other sex, segregating at one or more loci in the 1:1 sex ratio observed in *Populus* (Comtois et al., 1986; Farmer, 1964). These genes are distinct from the gender-differentiating genes that result in the differences observed between male and female floral organs. Gender-determining genes presumably initiate a cascade of gender-regulated gene-expression that determines the observed male and female phenotypes. In the liverwort, *Marchantia polymorpha*, the gene content of the sex chromosomes has been investigated, and Y specific genes have been identified, as well as

a gene M2D3.4, located on an autosome, which is expressed only in male gametic cells (Tanurdzic and Banks, 2004). However, none of these genes have been characterized as being gender-determining. Orthologs of the *Arabidopsis thaliana* genes *SHOOT-MERISTEMLESS* (*STMS*) and *CUPSHAPED COTYLEDON* (*CUC*) have been shown to control the establishment and development of the floral meristems in *Silene latifolia*, and therefore may play a role in the gender-determination pathway in this species (Zluvova et al., 2006). However, to date no gene directly involved in gender-determination has been characterized in a dioecious species, while genes that respond to physiological or environmental cues to give rise to unisexual flowers in the monoecious species *Zea mays* and *Cucumis sativus* have been characterized (Jamilena et al., 2008).

1.4.2 Gender-differentiation and development in plants

Genes may be gender-differentiating because their transcriptional regulation is downstream of a gender-determining gene, or because they are gender-linked and therefore differ in copy number between sexes, for instance they may be present in only one sex. An example of this is found in the fly *Drosophila melanogaster*, where sex-determination is controlled by the ratio of numbers of X chromosomes to sets of autosomes, as the X chromosome contains almost one third of the genes in this fly, an extra copy of this chromosome leads to aneuploid conditions and disruption of cellular equilibrium (Hake and O'Conner, 2008). DNA methylation is also important in regulating X-chromosome inactivation (Feng et al., 2010), and in *Carica papaya* methylated heterochromatin structures have been observed associated with the male-specific region of the Y chromosome (MSY), but not with the corresponding region on the X chromosome (Zhang et al., 2008).

An example of gender-differentiating genes can be found in the fern *Ceratopteris richardii*, where two genes are thought to be responsible for producing male (gene *TRA*) or female (gene *FEM1*) gametophyte traits depending on the presence of the hormone A_{CE} (antheridiogen Ceratopteris) which activates a set of genes, *HER*, *NOT1* and *MAN1*, which in turn influence the expression of *TRA* and *FEM1* (Tanurdzic and Banks, 2004). The size and gender composition of the gametophyte population surrounding a germinating spore, and therefore the level of A_{CE}, ultimately control the gender of the developing gametophyte, so this is a good example of hormonal mediation of gender development. A similar situation has been observed in the androdioecious species *Mercurialis annua*, where three unlinked loci (A, B1 and B2) determine gender via changes in levels of the plant hormones auxin and cytokinin (Khadka et al., 2005), but the gender ratio of flowers in hermaphroditic individuals is influenced by the frequency of male plants in the surrounding population (Dorken and Pannell, 2009).

1.4.3 Gender-differentiation in *Zea mays*

One of the species that has been developed as a model system for studying how gender-differentiation occurs in a monoecious plant is *Zea mays*. In this species an interaction of gender-differentiating genes, hormones and environmental factors combine to produce unisexual flowers in the same plant by acting on the immature floral meristem and causing the selective abortion of the inappropriate sex organs (Dellaporta and Calderon-Urrea, 1994). *Z. mays* produces flowers in inflorescences that consist of many individual spikelets, each subtended by two glumes that enclose two florets. At the beginning of development, the florets are bisexual, containing both male and female sexual organs, but when the florets transition to being unisexual, their location in the

plant, being either in a tassel (male) or ear (female) flower, and their position within the spikelet decides the gender that the floret will assume (Calderon-Urrea and Dellaporta, 1999). The abortion of either the stamens in female florets, or pistils in male florets, is regulated by the plant hormones gibberellin (GA) (Banks, 2008), and jasmonic acid (JA) (Acosta et al., 2009).

The model pathway for the differentiation of male florets in *Z. mays* contains three genes (*tasselseed1*, *tasselseed2*, and *silkless1*) that appear to regulate the production of jasmonic acid which plays a role in the programmed cell death in pistil primordia (Browse, 2009; Calderon-Urrea and Dellaporta, 1999). For the development of female florets, it appears that two genes that are cell-cycle regulators (CYCLIN B and WEE1) block the cell-cycle from continuing in stamens, thereby producing pistillate florets (Kim et al., 2007). A second class of mutations that causes dwarf (*d*) plants also results in stamen abortion. *d1*, *d2*, *d3*, and *d5* mutations disrupt steps in the gibberellin biosynthesis pathway which results in lower concentrations of endogenous gibberellin and causes an almost complete conversion of staminate to pistillate florets (Dellaporta and Calderon-Urrea, 1994). Environmental factors such as short day-length and low light levels have been shown to increase gibberellin levels in *Z. mays*, thereby causing the plant to produce predominantly female florets (Dellaporta and Calderon-Urrea, 1994).

1.4.4 Gender-differentiation in *Cucumis sativus* and *C. melo*

The role that environmentally induced changes in hormone levels play in affecting gender-differentiation has been looked into in more detail in two species of the Cucurbitaceae, *Cucumis sativus* L. and *C. melo*. Species in this family exhibit unisexual flowers, within either a dioecious, monoecious, gynoecious, or androdioecious mating,

and phylogenetic data indicates that dioecy is the ancestral mating system (Zhang et al., 2006). In *C. sativus* and *C. melo* gender expression is controlled by a combination of environmental, hormonal and genetic factors (Boualem et al., 2008; Trebitsh et al., 1997). In these two species gibberellin, long days and high temperatures cause the plants to produce a higher proportion of staminate flowers, whereas more pistillate flowers are produced under conditions of low temperature and short days, and in the presence of ethylene and auxin (Knopf and Trebitsh, 2006). Genes involved in the ethylene biosynthesis pathway mediate gender differentiation, but there are differences at the genetic level between the two species. Interestingly, though *C. sativus* and *C. melo* differ in geographic origin and chromosome number, and their most recent common ancestor is estimated to have occurred over 40 MYA, it appears that mutations resulting in reduced activity of 1-aminocyclopropane-1-carboxylic acid synthase (ACS) in both species have produced hermaphroditic, or perfect, flowers in both species independently (Boualem et al., 2009).

In *C. melo* gender-differentiation is governed by two genes, *andromonoecious* (*a*) and *gynoecious* (*g*), and different sexual types arise from different combinations of alleles of these two genes. Andromonoecious (*aaG-*) and monoecious (*A-G-*) genotypes produce male flowers on the main stems with perfect or female flowers on axillary branches, respectively, while gynoecious (*AAgg*) and hermaphroditic (*aagg*) individuals produce only female or perfect flowers (Boualem et al., 2008). The *a* gene has been characterized and codes for a 1-aminocyclopropane-1-carboxylic acid synthase (ACS), now named *CmACS-7*, based on homology to the *Arabidopsis* ACS-7 gene (Boualem et al., 2008). *CmACS-7* is an enzyme thought to catalyze the first rate-limiting step in the ethylene

biosynthesis pathway and in *C. melo* is required for the development of female flowers in monoecious plants, and a reduction in the activity of *CmACS-7* produces perfect flowers on andromonoecious plants (Boualem et al., 2008).

C. sativus L. var *sativus*, the wild-type cucumber, is monoecious and flowers are produced along the stem in a predetermined developmental sequence with staminate flowers occurring first, followed by a region where staminate and pistillate flowers are interspersed, and then at the end of the stems only pistillate flowers occur (Trebitsh et al., 1997). However, in domesticated *C. sativus* there are four different mating types and three major genes determine these: *Female* (*F/f*), *Monoecious* (*M/m*) and *Androecious* (*A/a*). The *F* gene promotes femaleness throughout the entire plant whereas the M gene determines if flowers are unisexual (*M*-) or perfect (*mm*), and when the *F* gene is homozygous recessive the *A* gene influences sex expression (Mibus and Tatlioglu, 2004). The genotypes of the four mating systems in *C. sativus* are as follows: gynoecious (*M-FF*), monoecious (*MMffA*-), hermaphrodite (*mmF*-) and andromonoecious (*mmffaa*) (Mibus and Tatlioglu, 2004). Due to incomplete dominance of the *F* gene, female sex expression can be enhanced by interacting with two additional loci, *Intensifier of female sex expression* (*In-F*) and *gynoecious* (*gy*), though these interactions have not been well characterized (Knopf and Trebitsh, 2006). Two alleles of the *F* gene have been characterized as *CS-ACS1G* (dominant allele) and *CS-ASC1* (recessive allele) and genotypes that are homozygous for the dominant allele (gynoecious and hermaphroditic plants) show increased ethylene concentrations, which results in more pistillate flowers (Mibus and Tatlioglu, 2004).

The *C. sativus* gene *CsACS2* is highly homologous to the *C. melo* gene *CmACS-7*, therefore the *C. sativus* locus *M* may be orthologous to the *C. melo a* locus (Boualem et al., 2009). It appears that in *C. sativus CsACS2* functions in a localized way to sense inhibiting ethylene in stamens and is therefore dedicated to arresting stamen development in floral primordia (Boualem et al., 2009). An *ETHYLENE-INSENSITIVE3*-like genetic sequence is also associated with the *M* locus in *C. sativus*, providing further evidence that gender-differentiation in this species is ethylene mediated (Liu et al., 2008). While further research is needed to identify the roles that the loci *A*, *gy*, and *In-F* play in gender-differentiation in *C. sativus*, it seems evident that the *F* locus controls ethylene concentrations, while the *M* locus mediates differences in ethylene sensitivity in male and female floral primordia (Liu et al., 2008). How gibberellin and auxin influence gender-differentiation in these species, and what relationship temperature and day length have to plant hormone levels that cause floral differentiation, has yet to be determined.

1.4.5 The role of MADS box genes in gender differentiation

Plant genes have been found to be fairly compact and usually grouped together in clusters, surrounded by repetitive DNA sequences, even in large genomes (Kellogg and Bennetzen, 2004). Floral development in hermaphroditic, or perfect flowers, is a complex, multistep process that begins with the establishment of a floral meristem, followed by the differentiation of floral organs that then develop into the structures observed in flowers (Zik and Irish, 2003). The current model for describing the genetic interactions that control flower development is the ABC model. This model defines three regions of the floral meristem, each of which is controlled by A-class of genes – A, B, or C (Coen and Meyerowitz, 1991). Region A comprises the first and second whorls of the

floral meristem, region B includes the second and third whorls and region C contains the third and forth whorls (Coen and Meyerowitz, 1991). Expression of A-class genes alone specifies sepal identity, A + B-class genes specify petal development in the second whorl, B + C-class genes expression results in stamen formation in the third whorl and C-class genes alone specify carpel development in the forth whorl (Zik and Irish, 2003). Many of the ABC model genes that control reproductive development in flowering plants belong to the MADS-box gene family of transcription factors, so called because these genes share a common motif found in the *MCM1*, *AGAMOUS*, *DEFICIENS*, and *SRF* genes (Coen and Meyerowitz, 1991). Further research has indicated that two other classes of genes are also involved in the floral development pathway. D-class genes have been found to be crucial for the determination of ovule identity (Vandenbussche et al., 2003), and E-class genes are required for specifying stamen, petal and carpel identity (Pelaz et al., 2000).

The floral dichotomy observed in dioecious species results from developmental modifications of a perfect flower to a female or male flower via organ suppression (Ainsworth, 2000). In *Silene latifolia* the genes that are differentially expressed in female and male flowers are only involved in male floral development, and neither they, nor the characterized Y-linked genes in this species, control gender-determination (Jamilena et al., 2008). It has been found that gender-determination in *Thalictrum dioicum* is likely to be regulated by genes upstream from organ identity genes because stamen or carpel primordia are not initiated in the unisexual female and male flowers, respectively (Di Stilio et al., 2005). The authors of this study postulated that differentially regulated alleles of B-class genes might be linked to a sex-determining factor. A similar system

may exist in *Populus trichocarpa* whereby a genetic switch at a sex locus controls expression of B and C-class genes, thereby controlling the development of male or female flowers.

MADS-box genes have not been shown to be associated with gender in either *Silene latifolia*, or *Cucumis sativus*, both of which have sex chromosomes (Hardenack et al., 1994; Perl-Treves et al., 1998). However, the C function genes have shown some gender-specific expression in *Rumex acetosa* (Ainsworth et al., 1995) and *Liquidambar styraciflua* (Liu et al., 1999). While the expression of B and C-class floral homeotic genes has been shown to decrease in the aborted organs in unisexual flowers in all species that have been studied, it is not known if this is causal or a consequence of the organ abortion (Tanurdzic and Banks, 2004).

1.4.6 Gender-associated genes

Unlike animals, where secondary sexual characteristics are usually very evident and important in mating habits and the life history of the species, in plants generally little is known about the evolution of sexual dimorphism that would result from the development of secondary sexual characters associated with one or the other gender (Meagher, 1984). Consequently, to date gender-associated genes are not well characterized, though observations in *Chamaelirium luteum* (Meagher, 1984), *Asparagus officinalis* L. (Bracale et al., 1991), and *Populus trichocarpa* (Brunner, 2010), have indicated that male plants can exhibit slightly different growth patterns or flowering times than females. In the homosporous fern *Ceratopteris richardii* the male and hermaphroditic gametophytes can be distinguished from the female gametophyte by the absence of a multicellular meristem (Tanurdzic and Banks, 2004), but it is uncertain that

this difference could be considered a secondary sexual characteristic controlled by gender-associated genes. One limitation to the development of secondary sexual characteristics in plants may be that in order to continue to reproduce, males and females of a species must maintain overlap in their habits and ecological requirements so that they can breed effectively (Meagher, 1984).

1.4.7 Differences in genetic control of gender-determination between monoecy and dioecy

One of the indications that the genetic control of gender-determination differs between monoecious and dioecious mating systems is that the ratio of males and females in a population is determined differently in each system. In dioecious species like *Populus trichocarpa*, the female to male ratio of individuals in a population is thought to be determined by genetic segregation of alleles at one or more loci, whereas in monoecious species the proportion of female to male gametes is influenced by environmental or epigenetic cues (Irish and Nelson, 1989). This would seem to indicate that while both monoecious and dioecious species have gender-differentiation and gender-associated genes, only dioecious species have gender-determining genes (Figure 1.3).

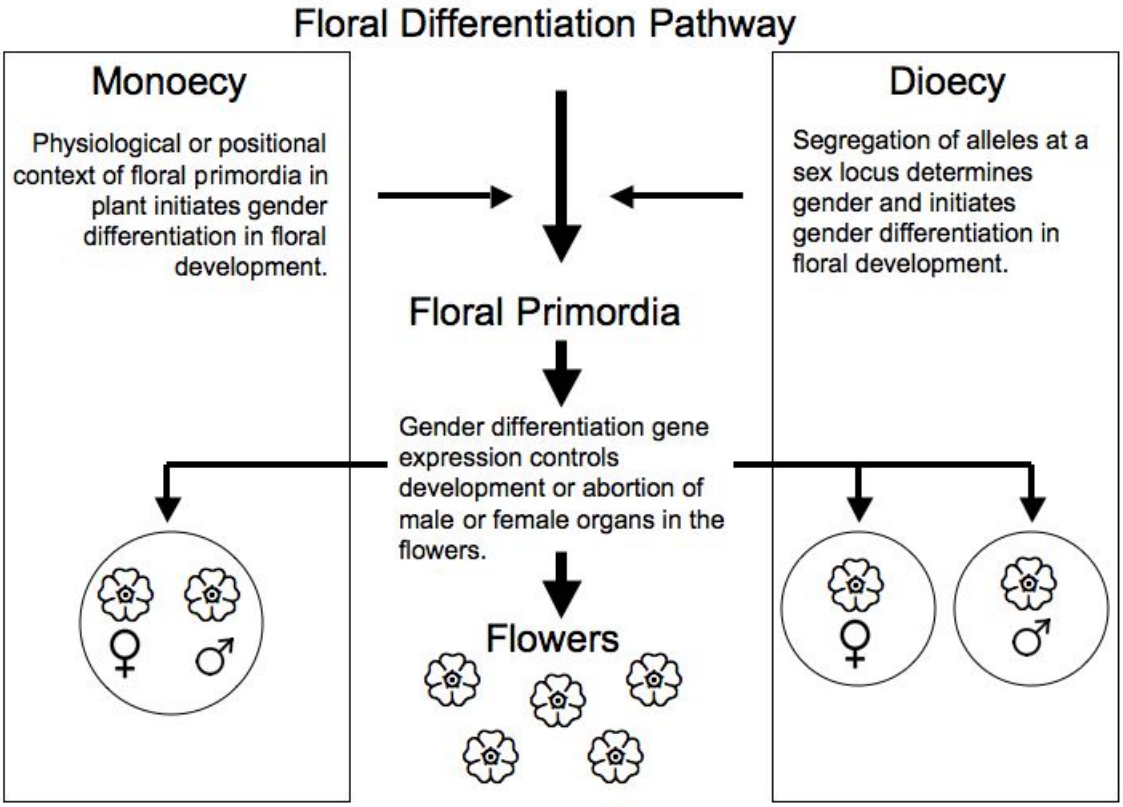

Figure 1.3 Schematic comparing how gender-determining, gender-differentiating and gender-associated genes act in monoecious versus dioecious mating systems.

1.5 *Populus trichocarpa* as a model organism for studying gender-determination in plants

1.5.1 Life history of *Populus trichocarpa*

Populus trichocarpa is a deciduous, paleopolyploid tree native to Western North America where its range extends from California to Alaska (Brunner et al., 2004a). Though it has a broad environmental range, *P. trichocarpa* has fairly specific ecological requirements in that it is generally confined to river flood plains and the riparian zones of alluvial streams (Rood and Polzin, 2003). *Populus trichocarpa* has a juvenile stage that

lasts 5-10 years, during which time the trees are unable to flower and reproduce sexually (Brunner, 2010). The female and male flowers of this species are highly reduced in form, consisting of a floral disk with either 40- 50 stamens or a single pistil (Eckenwalder, 1996). The inflorescences are pendulous catkins that flush three to four weeks before leaves appear in the spring (Boes and Strauss, 1994). *P. trichocarpa* flowers are wind pollinated and after fertilization occurs in the spring, 20-40 capsules/female inflorescence are produced. Each capsule contains approximately 40 seeds that are released and wind dispersed in the early summer (Boes and Strauss, 1994).

1.5.2 Characteristics of *Populus trichocarpa* as a model species

As a member of the genus *Populus*, *P. trichocarpa* is a good model species for the study of gender in plants because of its small genome (approximately 500 million base pairs and 45 000 genes), high growth rates, economic importance of the genus, and the ease with which it is clonally propagated (Wullschleger et al., 2002a). It is already used as a model to study tree morphology, physiology, biochemistry and genetics. *Arabidopsis thaliana*, already a model species for studying most aspects of plant biology, is more related to *Populus* than the majority of dicot species for which genomic resources exist, making it possible to make informative comparisons between the two species' genomes (Jansson and Douglas, 2007). Much of the interest in developing *P. trichocarpa* as a model organism has centered on the importance of secondary xylem formation and the application of this research in the forest industry (Wullschleger et al., 2002b). However, there has been some research done on *P. trichocarpa* genes that could be engineered to render the tree reproductively sterile, which is of interest because of the

importance in limiting gene flow between transgenic and wild populations (Skinner et al., 2003).

The recent sequencing of the female *P. trichocarpa* clone "Nisqually 1" genome (version 1.1 released September 2004, version 2.0 released in January 2010), in concert with developments in gender-determination research in other model species provides exciting opportunities to understand patterns in the genomic architecture relative to gender. *P. trichocarpa* is the first dioecious plant to have had its genome sequenced.

1.5.3 Genomic markers linked to sex-differentiation in *Populus* and *Salix*

Fixed heterozygosity should occur in the region of the sex locus to ensure that there is no recombination, otherwise separate sexes would not be conserved. Research on *Salix viminalis* L. has determined the existence of a sex locus in this species (Gunter et al., 2003b), and as species in the closely related genera *Salix* and *Populus* exhibit stable dioecious sexual systems (Gunter et al., 2003b), it is probable that *Populus* species have a similar system.

A marker co-segregating with gender has been found in *S. viminalis* L. (Gunter et al., 2003b) and it has been postulated that several loci interacting via epistasis may be responsible for sex-determination in this species (Alström-Rapaport et al., 1998). When genetic markers from chromosome 19 for a female *P. alba* and a male *P. nigra* were aligned the linkage maps indicated that the sex morphological trait mapped to two different regions of chromosome 19, providing evidence that at least two loci are involved in sex-determination in *Populus* (Gaudet et al., 2008). If a multilocus sex-determination system is at work in the Salicaceae, I hypothesized that I should be able to use fine-scale mapping of genetic markers linked to gender that, when combined with the

sequenced genome, would enable me to pinpoint the exact location(s) of the *P. trichocarpa* sex locus/loci. RAPD markers associated with sex-determination in *P. tomentosa* have been reported (Hou et al., 2009), however, so far attempts to find markers that reliably co-segregate with gender in *P. trichocarpa* have failed (G. Tuskan, pers. comm.).

1.5.4 Gender-determination and chromosome 19 in *Populus*

In the genus *Populus*, studies suggest that gender is controlled via a genetic mechanism, and that the sex locus is located on chromosome 19 of the linkage map or sequenced genome (Paolucci et al., 2010). Initially it was thought that in *Populus* the females may be the heterogametic gender, and that the males are homogametic at the sex locus. This is because in the closely related genus *Salix*, sex specific markers have been found for females but not for males (Gunter et al., 2003b). However, this is opposite to what has been found in *Asparagus* and many other genera containing dioecious species (Semerikov et al., 2003).

The association between gender-determination and chromosome 19 in the genus *Populus* was first reported in *Populus nigra* (Gaudet et al., 2008) and *P. trichocarpa* (Yin et al., 2008), where detailed genetic maps constructed for these species indicated that a genetic marker segregating with gender existed on chromosome 19. Several features of chromosome 19 in *P. trichocarpa*, such as a region of recombination suppression and segregation distortion extending over a large portion of the chromosome, as well as haplotype divergence observed for this chromosome suggest that it is involved in sex-determination (Yin et al., 2008).

It appears that there is considerable variability across the six species in which gender has been investigated in *Populus* as to the position of the putative sex locus on chromosome 19, and research indicates that either males or females can be the heterogametic gender, depending on species. In *P. trichocarpa*, the sex locus is reported to be located on the peritelomeric portion of chromosome 19, and evidence from distorted segregation of microsatellite markers, suppression of recombination, and haplotype divergence observed only in the maternal parent in pedigreed crosses indicates that the female is the heterogametic gender in this species (Yin et al., 2008). A putative sex locus in *P. nigra* is also located on the terminal end of chromosome 19, though in this species the mapping of AFLP, SSR and SNP genetic markers on the genome indicated that the male was the heterogametic gender (Gaudet et al., 2008).

In two species of *Populus* the sex locus has been mapped to a non-telomeric position on chromosome 19. In *P. tremuloides*, sex as a morphological trait was mapped to a central position on chromosome 19 in the male parent using an interspecific pedigreed of 61 full sibling hybrids of *P. tremula* L. x *P. tremuloides* Michx. (Pakull et al., 2009). The sex locus also maps to the middle of chromosome 19 in *P. alba*, though in this case the heterogametic sex is reported to be the female (Paolucci et al., 2010). In *P. tomentosa*, a RAPD marker has been identified that produces a DNA fragment in the male individuals that were sampled, but not the female individuals, though the reason for this difference in DNA amplification between the genders has not been identified (Hou et al., 2009).

1.6 Research objectives

To further explore the genetics of gender in dioecious species I have *used P. trichocarpa* as a model species to investigate the development of sex-linked markers for *Populus* based on those reported in *Salix viminalis* (chapter two). I conducted a survey of the genes located on the telomeric end of chromosome 19 looking for gender-biased SNPs in gene sequences (chapter three), as this region of the genome in numerous *Populus* species has been reported to be involved in gender-determination in this genus. I also investigated gender-bias in genome wide gene-expression using a microarray approach (chapter four). Finally I identified three MADS box genes that show a gender-biased expression pattern in male and female floral tissues (chapter five).

Chapter 2: Characterization of sex-linked genetic markers developed in *Salix viminalis* in other *Salix* species and *Populus trichocarpa*

2.1 Introduction

The majority of flowering plants produce hermaphroditic flowers, which contain both male (stamens and pollen) and female (style, stigma and ovules) reproductive parts (Ainsworth, 2000). Only approximately ten percent of the world's angiosperms have mating systems that divide male and female function between separate flowers on the same individual (monoecy), or in some cases, separate individuals (dioecy). Of this ten percent, approximately six percent (14 000 species of flowering plants) are dioecious (Renner and Won, 2001).

Gender systems in plants have been characterized morphologically and physiologically, and possible explanations for the ecological basis for their evolution have been studied as well (Ainsworth, 2000; Ming et al., 2007). However, until recently there has been little research on how gender is determined at the molecular level, or what evolutionary pathway(s) development of separate genders could have taken. Many studies on dioecious species have focused on plants that, unlike poplar and willows, have well-developed heteromorphic sex chromosomes such as plants in the genera *Rumex* or *Silene* (Ainsworth et al., 1995; Filatov et al., 2000).

The publication of the *Populus trichocarpa* genome provides an exciting opportunity to explore the genetic underpinning of gender in plants, as it is the first dioecious species to have a sequenced genome. The genus *Populus* is closely related to *Salix*, in the family Salicaceae, and both these genera have 19 chromosomes with no evidence of heteromorphic sex chromosomes (Semerikov et al., 2003) Almost all species

in these two genera are dioecious, but two species (*Populus lasiocarpa* and *Salix martiana*) have been reported to be monoecious (Semerikov et al., 2003), and the mechanism by which sex is determined in this family is not well understood. Due to the strong association between the occurrence of dioecy and monoecy in related genera, it is thought that dioecy evolved via the intermediate step of monoecy in Salicaceae.

Determining how gender is regulated in Salicaceae could increase the understanding of the evolution of sexual systems in plants, but this information also has important economic uses. Many *Populus* and *Salix* species are used in short-rotation plantations for energy production and it has been shown that male *Populus* clones have higher dry weight fiber yields on average than females (Tschaplinski et al., 1994). Gender in dioecious tree species therefore influences their economic value and can affect breeding schemes (Alström-Rapaport et al., 1998). Although *Populus* clones can achieve reproductive maturity in four to six years under specific conditions (Tuskan et al., 2006), the majority of individuals start flowering between five and ten years after germination (Braatne et al., 1996). Because of this, the process of selecting trees with traits that are beneficial to the wood fiber industry, or for use in ecological reclamation or phytoremediation can be a long one. The development of a sex-linked marker that could be used to determine the gender of clones prior to sexual maturity could aid selective breeding programs.

2.1.1 Sex-linked genetic markers in *Salix*

Previous work in *Salix viminalis* has identified a genetic marker that appears to segregate with gender. Alstrom-Rapaport et al. (1998) discovered a decamer primer (UBC 354) which generated a randomly amplified polymorphic DNA (RAPD) product

consisting of a single 560bp band. This marker was shown to be biparentally inherited, and is associated with femaleness in *S. viminalis*. The pattern of gender segregation led the authors to propose that a two-locus epistatic genetic model of sex-determination exists in this species. However, a later study using amplified fragment length polymorphism (AFLP) fragments suggested a single locus determinant of gender in *S. viminalis* (Semerikov et al., 2003). This same study found four AFLP fragments associated with sex in *S. viminalis*, all of which were predominantly present in females, but absent in males (Semerikov et al., 2003). The results from these two studies suggest that females of this species may possess some chromosomal regions not found in the males. However, the sex-linked markers developed in *S. viminalis* did not co-segregate with gender in *Salix caprea* (Semerikov et al., 2003), which suggests that species of the genus *Salix* do not have complete synteny with respect to their sex-determination system.

Sequence characterized amplified region (SCAR) markers developed from the original *S. viminalis* RAPD markers identified by Alstrom-Rapaport et al. (1998) resulted in identification of two sex-linked markers associated with femaleness in this species (Gunter et al., 2003a). These two DNA SCAR markers, SCAR 354 and SCAR AE08, were tested in *Salix eriocephala* Michx. and it was found that the presence or absence of both these markers differed significantly between male and female plants (Gunter et al., 2003b).

2.2 Objectives

It is not known how conserved these SCAR markers for gender are across a broader range of willow species, or if the markers will be associated with sex in other genera in Salicaceae, such as *Populus*. In order to widen the survey of willow species

tested with these markers I identified three objectives for my research. For my first objective I attempted to amplify the SCAR 354 and AE08 markers in 12 willow samples (comprising different species and sexes), and in *Populus trichocarpa* males and females, and sequenced the resulting products. My second objective was to design primers based on the poplar genome sequence in order to expand the amplified SCAR sequence into the associated coding gene regions identified in poplar, to further characterize the sex-linked SCAR marker. The third objective was to sequence the gene-anchored sequences that primers were designed for, and see if differences in the DNA sequences in this region could be identified that would explain why this SCAR marker appears to be sex-linked.

2.3 Methods

2.3.1 Collection and preparation of plant materials

I collected leaf samples from twelve species of *Salix* from the UBC Botanical Gardens, and VanDusen Botanical Gardens, Vancouver B.C., in the fall of 2006, and stored them in silica gel at room temperature to preserve and desiccate the tissues until DNA was extracted from them. The sex of the *Salix* species collected was determined if possible. I collected the *P. trichocarpa* leaf samples used in this work in April of 2006, from trees in natural stands located on the UBC campus that I had previously identified as males or females. The *P. trichocarpa* leaf tissue was collected fresh and stored at -80°C. Collections of *Salix reticulata* and *S. arctica* were made in the spring of 2006 from natural populations near Prince George B.C., and stored in silica gel at room temperature.

Genomic DNA was extracted from leaf tissue of twelve *Salix* species, and two *Populus trichocarpa* individuals using a modified CTAB procedure (Doyle and Doyle,

1987). DNA samples were cleaned up using the QIAquick PCR purification kit (ON, Canada) and stored at minus 20ºC.

2.3.2 PCR conditions and sequencing of samples

SCAR primers were ordered from Integrated DNA Technologies, Inc (Coralville, IA, USA). The fifteen genomic DNA samples were tested for marker presence or absence with SCAR 354 (forward primer: 5'-GAGAGGGAGGGAGATTTAAG-3'; reverse primer: 5'-GCCGTAGCAGATTGTTAATCAC-3') under the following reaction conditions: Each 50 µl reaction contained 5 µl of 10X Taq buffer (100mM Tris-HCl (pH 8.8), 500mM KCl, 0.8% Nonidet P40), 5 µl of 5µM forward primer, 5 µl of 5µM reverse primer, 5 µl of 2mM dNTPs (Fermentas Life Sciences, CA, USA), 3 µl of 25 mM $MgCl_2$, 0.25 µl *Taq* DNA polymerase (Fermentas Life Sciences, CA, USA), 40 ng of genomic DNA and ddH_2O made up to 50 µl. Polymerase chain reactions (PCRs) were run on an Eppendorf Mastercycler gradient thermocycler for 35 cycles under the following reaction conditions: 94˚C 1 min, 55˚C 1 min, and 72˚C 2 mins, with an initial 3 min 95˚C denaturation step and a final 5 min 72˚C polymerization step. Amplification products were visually scored for the presence or absence of marker bands.

Amplified PCR products were then purified using the QIAquick PCR purification kit (ON, Canada), resuspended in ddH_2O, and prepared for sequencing. Each 10 µl sequencing reaction contained 1 µl of 5µM primer (either forward or reverse), 40ng amplified DNA suspended in ddH_2O, 0.5 µl BigDye v3.1 (Applied Biosystems, CA, USA), 2.5 µl 5x BigDye v3.1 buffer, and ddH_2O to 10 µl. Sequencing reactions were run under the following conditions: 96˚C 10 seconds, 55˚C 5 seconds, 60˚C 4 mins, with a 72˚C 1 min extension, and an initial 96˚C denaturation step. The sequenced products

were purified using a Sephadex column procedure (Graham and Olmstead, 2000), and run out on an ABI 377XL machine.

2.3.3 Sequencing alignment for the SCAR sequences amplified in *Salix*

Resulting sequences were aligned using Se-Al 1.d1 Sequence Alignment Editor, available from http://evolve.zoo.ox.ac.uk (Rambaut, 1996) and Sequencher 4.2.2 (Gene Codes Corp., Ann Arbor, MI), and displayed using Boxshade 3.21 (http://www.ch.embnet.org). The sequence for the *Salix viminalis* SCAR amplified fragment sequence (Gunter et al. 2003b; accession number = AY192565) was obtained from the NCBI website (www.ncbi.nlm.nih.gov/Genbank/index.html) and added to the alignment as well. Primer sequences were removed from both ends of the SCAR sequences prior to BLASTing them against the poplar genome on the JGI website (http://genome.jgi-psf.org/Poptr1_1/Poptr1_1.home.html), and the resulting *Populus* sequence that had sequence similarity with the *Salix* sequences was added to the alignment. Sequences were deposited with GenBank (accession numbers EF206296-EF206307). The *Salix* SCAR marker sequences were aligned and compared to the sequenced *Populus trichocarpa* genome using a BLAST search. All the work reported here was completed prior to the publication of version 2.0 of the poplar genome, so gene names and annotations are with respect to version 1.1 of the poplar genome.

2.3.4 Novel primer design based on SCAR sequence and associated coding gene regions in *P. trichocarpa*

Following the comparison of the SCAR sequences to the *P. trichocarpa* genome, I designed primers based on the poplar genome sequence in order to expand the amplified SCAR sequence into the associated coding gene regions identified in poplar. Four new

PCR primers for this project were designed based on the *P. trichocarpa* genome sequence that showed similarity to the *Salix* SCAR sequences (Table 2.2, Figure 2.1). Primers were designed using Primer3 (v. 0.4.0, http://frodo.wi.mit.edu/primer3/), and tested *in silico* to determine the probability of primer dimers or unspecific primer binding with Amplify 3X (v.3.1.4, http://engels.genetics.wisc.edu/amplify). The twelve willow and two poplar genomic DNA samples were then tested, as well as thirteen individuals of *S. arctica* (nine female, and four males), and twelve individuals of *S. reticulata* (eight females and four males), for marker presence or absence with the new primers using the same PCR conditions that were used for SCAR 354 amplification, but with an annealing temperature of 50˚C. PCR products from these reactions were then purified and sequenced following the same protocol as above in section 3.2.

2.3.5 Cloning, sequencing and phylogenetic analysis of *S. arctica* and *S. reticulata* samples

In order to further investigate the relationship between SCAR 354 and gender, I sequenced 13 individuals of *S. arctica* (9 females and 4 males) and 12 individuals of *S. reticulata* (8 females and 4 males). Because *S. arctica* is a tetraploid and *S. reticulata* is a diploid, to get good quality sequences for these willow samples, I cloned the samples using the TOPO TA Cloning Kit Sequ 20 rxn (cat. # K457501, Invitrogen Canada Inc., Burlington, ON), and then confirmed the presence of my target sequence in the plasmid DNA by performing an Eco R1 digestion. The plasmid DNA was cleaned up using the QIAprep Spin Miniprep Kit (50) (cat. # 27104, Qiagen Inc., Toronto, Ont., Canada), and then samples were sequenced following the same protocol outlined above in section 3.2. Once I obtained good quality sequences for all the individuals, they were aligned

following the same method outlined above. I used PAUP (Swofford, 2001) to construct a consensus tree, and then looked at SNP variation using parsimony (DELTRAN), as implemented in MacClade 4.0 (Maddison and Maddison, 2000).

2.4 Results

2.4.1 Amplification and characterization of the SCAR 354 marker

No amplification products were obtained with SCAR AE08 so all the work reported here concerns SCAR 354, and the primers designed are based on SCAR 354 and *Populus trichocarpa* sequences. The SCAR 354 marker was amplified in only five of the *Salix* species tested (*S. caprea*, *S. fargesii*, *S. gracilistyla*, *S. glauca*, and *S. nakamura*) and was not amplified in *Populus trichocarpa*. To the extent that the sex of this material was known, I was able to amplify the SCAR 354 marker in both male and female willow samples (Table 2.1). The sequences appeared to be homologous to the previously determined sequence for SCAR 354 in *S. viminalis* (GenBank AY192565).

When the amplified SCAR 354 sequence was BLASTed against the poplar genome, the part of the sequence that was conserved across *Salix* species had very high similarity to a sequence in *Populus trichocarpa* that maps to chromosome 15 (LG_XV) in the poplar genome (Figure 2.1), very close to a Ssu72-like gene (fgenesh1_pg.C_LG_XV000380). A putative homologue of this poplar gene also occurs on chromosome 1 of *Arabidopsis* (AT1G73820.1, Ssu72-like family protein) with 82% similarity to the gene in poplar. However the upstream region in *Arabidopsis* shows no substantial similarity with the upstream region in *P. trichocarpa* and *Salix* species.

Based on the comparison of the various willow sequences, three distinct regions were identified in the amplified region (Figure 2.1). At the 5' end is a microsatellite-like

region (MSR), which consists of approximately 177bp, and varies considerably in length among the five *Salix* species it was amplified in (Table 2.1) but is not found in poplar.

Adjacent to the MSR is a semi-conserved region (SCR) which shows good similarity among the *Salix* species but which does not completely align with the *P. trichocarpa* sequence. The majority of the SCAR 354 sequence consists of a conserved region that shows a 92% similarity between the *Salix* species and *P. trichocarpa*. This conserved region is 298bp long in *P. trichocarpa*, and 313bp in *S. viminalis*; the discrepancy in sequence length in this region is due to three insertions totaling 15bp in *Salix*. The conserved region is immediately 5' to the Ssu72-like gene (85 bp from the putative start site of the gene). The close proximity to the gene fgenesh1_pg.C_LG_XV000380 (a putative protein involved in transcriptional start site selection) is a possible reason for the high level of sequence conservation across species. Table 2.2 lists the other putative genes that are located in the vicinity of the homologous sequence to SCAR 354 in the poplar genome.

Table 2.1 Species tested for amplification with the SCAR 354 primers.

Species	Accession/Collection number	Sex	Amplification
Salix (Chosenia) arbutifolia	UBC-BG 036493-0126-2002	N/D*	no
S. caprea	VanDusen 0822 1993	female	yes
S. eleagnos	UBC-BG 013854-0013-76	female	no
S. fargesii	22112-027-1982	male	yes
S. glauca	UBC-BG 14540-284-77	N/D	yes
S. gracilistyla	UBC-BG 034625-5555-1999	male	yes
S. hookeriana	UBC-BG 037096-1003-2004	N/D	no
S. lapponum	UBC-BG 013859-0013-1976	female	no
S. lucida	UBC-BG 10512-268-74	N/D	no
S. nakamura	UBC-BG 033409-653-97	male	yes
S. purpurea	UBC-BG 034628-5555-1999	female	no
S. sitchensis	UBC-BG 2538-099-71	N/D	no
Populus trichocarpa	54-524	female	no
P.trichocarpa	50-533	male	no

*N/D indicates that the gender was undetermined.

2.4.2 Gene-anchored amplification in *Salix* and *P. trichocarpa*

Three primer pairs successfully amplified part of the coding region, and the conserved region immediately to the 5' end of the Ssu72-like gene in *Salix* and poplar (Table 2.3). The amplification using 1F/2R was consistent irrespective of species or gender (Table 2.4). Amplification using 1F/5R was less consistent but gave longer sequences for *S. caprea* and *P. trichocarpa* (Figure 2.2). These results confirm that the SCAR marker sequence is adjacent to the Ssu72-like gene in *Salix* species just as in *P. trichocarpa*.

2.4.3 Identifying SNP differences that may explain why this SCAR marker appears to be sex-linked

The SCAR 354 sequence consisted of a length of variable purine-rich repeats, and a region highly conserved between species (Figure 2.1). The conserved region corresponds to the putative promoter region of an Ssu72-like gene (involved in transcriptional start site regulation) on chromosome 15 in poplar. Gene-anchored primers were designed to amplify the conserved region and part of the Ssu72-like coding region in willows and poplar (all genders and species). The conserved region the gene-anchored primers amplified contained numerous single feature polymorphisms (SFPs) both within and between species.

To investigate the relationship between SCAR 354 and gender, 12 individuals of *S. reticulata* (eight females and four males), a diploid willow species (Figure 2.3), and 13 individuals of *S. arctica* (nine females and four males), a tetraploid willow species (Figure 2.4), were sequenced. A consensus tree attained for *S. reticulata* did not indicate that the SNPs segregated with gender, but a bootstrap analysis supported the existence of haplotypes. A consensus tree attained for *S. arctica* did not indicate that the SNPs segregated with gender, but a bootstrap analysis supported the existence of three haplotypes. I detected female biased SNPs in both *S. reticulata* (one SNP) and *S. arctica* (three SNPs) consensus sequences, though the majority of SNPs in both species did not show a gender-bias in their distribution.

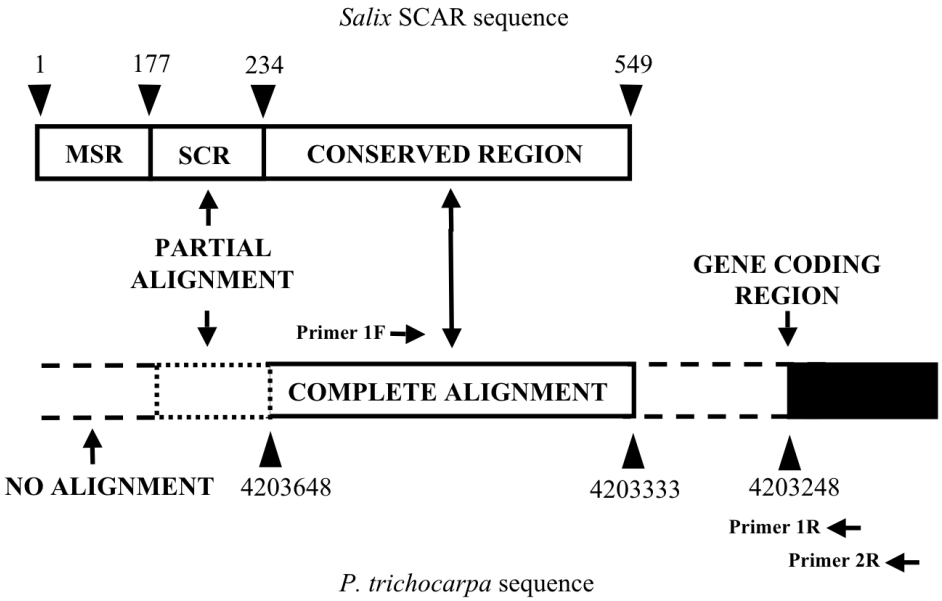

Figure 2.1 Structure of SCAR 354 marker regions in *Salix* species, and homologous region found on the *Populus trichocarpa* genome.
MSR = microsatellite-like region, SCR = semi-conserved region. The numbers above the SCAR indicate the base pair number, 5' to 3'. The numbers below the poplar region are the base pair positions along chromosome 15.

Table 2.2 List of putative genes in the region surrounding the location of the *Populus trichocarpa* sequence on chromosome 15 that show homology to the SCAR 354 *Salix* sequence.

Gene	Name	Location on LG_XV	Putative gene function
- 5	fgenesh1_pg.C_LG_XV000375	4162512-4165784	C-type lectin
- 4	fgenesh1_pg.C_LG_XV000376	4172930-4174370	Predicted transposase
- 3	estExt_fgenesh4_pg.C_LG_XV0383	4176495-4177658	Not known
- 2	eugene3.00150420	4181734-4182456	Not known
- 1	estExt_Genewise1_v1.C_LG_XV1107	4186527-4190414	Ribonucleoprotein complex SRP, Srpl 1 component
0	fgenesh1_pg.C_LG_XV000380	4199692-4203247	Ssu72-like (Protein involved in transcription start site selection)
+ 1	eugene3.00150423	4207845-4210979	Not known
+ 2	fgenesh1_pg.C_LG_XV000390	4214615-4217179	Aldehyde dehydrogenase
+ 3	grail3.0043012001	4229045-4229320	Leucine-rich repeat proteins
+ 4	grail3.0043012101	4229463-4229628	Leucine-rich repeat proteins
+ 5	grail3.0043012201	4229950-4233882	Leucine-rich repeat proteins

Table 2.3 Gene-anchored primers designed to amplify conserved regions 5' of the poplar Ssu72-like gene (fgenesh1_pg.C_LG_XV000380) and the equivalent region in *Salix*.

Primer name	Primer sequence (5'-3')	T_m
Ssu72-like 1F	GAA CAA TTC TCA ATC AAG TTC	49.3
Ssu72-like 1R	ATC ACC GGC ACY GTA CAG	56.0
Ssu72-like 2R	TGT TCT GAT TSG ACG AAC	49.6
Ssu72-like 5R	TAT AAG GGG TGC CAA AGT CG	54.6

Table 2.4 Species tested for amplification with Ssu72-like primers.

Species	Sex	1F + 1R	1F + 2R
Salix (Chosenia) arbutifolia	N/D*	yes	yes
S. caprea	female	yes	yes
S. eleagnos	female	yes	yes
S. fargesii	male	no	yes
S. glauca	N/D	yes	yes
S. gracilistyla	male	no	yes
S. hookeriana	N/D	yes	yes
S. lucida	N/D	no	yes
S. nakamura	male	yes	yes
S. purpurea	female	yes	yes
Populus trichocarpa	female	yes	yes
P. trichocarpa	male	yes	yes

*N/D indicates that the gender was undetermined.

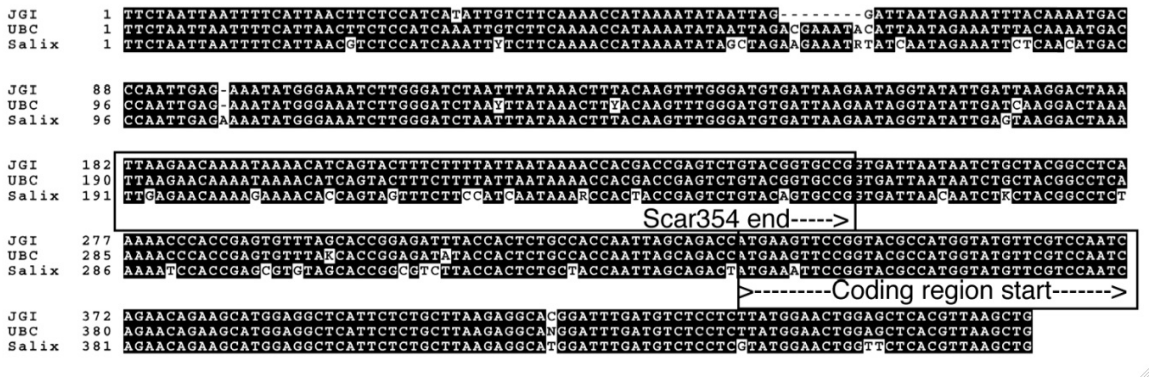

Figure 2.2 Sequence alignment of the SCAR 354 sequences obtained from *Salix caprea* and *Populus trichocarpa* (for the region alignable between the two genera).
JGI indicates the sequence from the genome project and UBC is another poplar sequence (50-533). There is an 8 bp deletion in the genome sequence (JGI) relative to the other sequence (UBC). The relative positions of the SCAR sequence and the coding region are indicated.

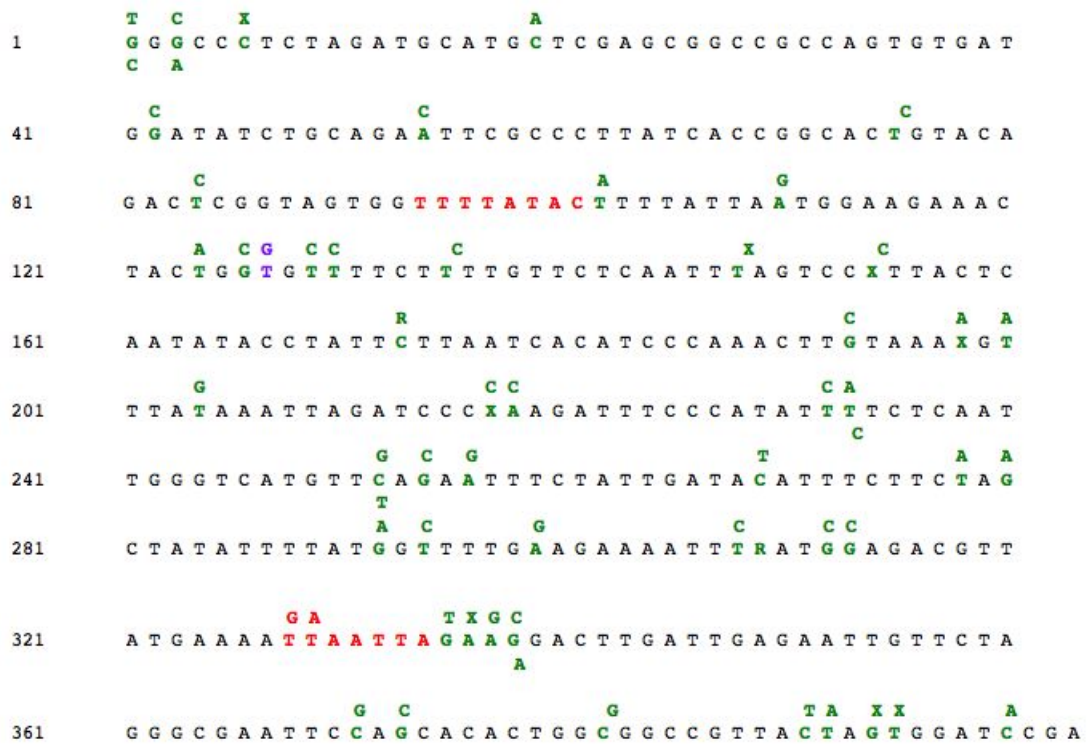

Figure 2.3 Consensus sequence (402bp long) for *S. reticulata* obtained from gene-anchored primers amplifying the SCAR 354 conserved region and a portion of the coding region of the Ssu72-like gene.
Insertions are indicated by red, deletions are indicated by X, SNPs are indicated by green. Female biased SNPs are indicated in purple. In cases where there are two alternate bases for a SNP, the most frequent change is above the line.

Figure 2.4 Consensus sequence (522bp long) for *S. arctica* obtained from gene-anchored primers amplifying the SCAR 354 conserved region and a portion of the coding region of the Ssu72-like gene.
Insertions are indicated by red, deletions are indicated by X, SNPs are indicated by green. Female biased SNPs are indicated in purple. In cases where there are two alternate bases for a SNP, the most frequent change is above the line.

2.5 Discussion

2.5.1 SCAR 354 amplification in *Salix*.

My study, using species other than *S. viminalis*, found that amplification using the original SCAR 354 primers is inconsistent with respect to species and gender. *Salix nakamura*, *S. gracilistyla*, *S. glauca* and *S. caprea* yielded discrete amplification products when tested with the SCAR 354 primers, and of these four species, *S. nakamura* and *S. gracilistyla* have been confirmed as male, and *S. caprea* is a female. For *S. glauca*, the sex is as yet undetermined due to poor growth of the plant. This inconsistency is to be expected, as there is considerable nucleotide divergence between the species sequenced. All the species of willow that successfully amplified are in subgenera *Vetrix* or *Chamaetia*, which are separate from subgenus *Salix* (Azuma et al., 2000).

Comparative sequence alignment reveals different levels of sequence conservation in the SCAR region. The microsatellite-like region (MSR) at the 5' end of the SCAR 354 sequence does not have a counterpart in the poplar genome (Figure 2.1), but is present and variable in length in all *Salix* species tested. The MSR consists of an AG (purine)-rich repeat region with many AGAGG and similar repeats. In the sequences the MSR varies in length, shortest in *S. gracilistyla* and *S. nakamura*, and longest in *S. viminalis* (GenBank AY192565). The conserved region identified between the SCAR 354 sequences in the *Salix* species and the *P. trichocarpa* sequence is located 84 bp from the 5' end of a putative gene on the poplar genome (Figure 2.1). The conserved region also contains a candidate TATA box. Promoters are generally located 50-300 bps away from the beginning of a gene (Cooper et al., 2006), which is consistent with the location of this conserved region.

2.5.2 Amplification of the Ssu72-like gene region in *Salix* and *P. trichocarpa*

Given that there is 92% similarity between the *Salix* and *P. trichocarpa* sequences in the conserved region (Figure 2.2), and that this region exhibits many of the characteristics of a promoter region, it is plausible that, as in the poplar genome, the SCAR marker in *Salix* is also adjacent to this gene. This hypothesis was tested by designing primers based on putatively conserved regions of the poplar gene and the conserved part of the *Salix* SCAR.

The designed primer pairs consistently amplified this region, and the resulting sequences indicate that the association between the SCAR and Ssu72-like gene in the *Salix* species sampled is homologous to *P. trichocarpa* (Figure 2.2). This is not unexpected as *Salix* and *Populus* are closely related, have the same chromosome number and share a whole genome duplication event prior to 65 million years ago (Tuskan et al., 2006).

The Ssu72 protein in yeast consists of 206 amino acids and functions to decrease the elongation rate of RNA polymerase II by balancing elongation and termination of this polymerase activity (Dichtl et al., 2002). In yeast it is thought to be a phosphatase that interacts with CTD kinase Kin28 and CTD phosphatase Fcp1 to terminate the transcription of pre-snoRNA and some pre-mRNAs, which in turn regulates *cis* and *trans*-acting signals in the cell (Ganem et al., 2003).

The poplar Ssu72-like protein has a single-copy homologue in *Arabidopsis*, the gene At1g73820 (chr. 1: bp 27675718-27781654). The putative *Arabidopsis* protein is the same length and highly conserved relative to poplar. The main differences between the genes are (1) a greatly expanded second intron in the poplar gene, and (2) the absence

in *Arabidopsis* of the 87 bp intron between the third and fourth exons (*Arabidopsis* therefore has a three exon gene structure rather than the four exon gene structure of poplar).

2.5.3 Identifying single feature polymorphisms (SFPs) that may explain why this region is sex-linked in some *Salix* species.

Consensus trees attained for *S. arctica* indicated that while the individuals sampled did not segregate with gender, three distinct haplotypes supported by bootstrap analysis seem to exist within this species for the SCAR 354 marker sequence. For *S. reticulata*, two haplotypes were supported by bootstrap analysis of the data. These results could also be explained by sequencing errors or pseudo gene clones, but the existence of difference haplotypes within the individuals sequenced is the best explanation for these results. It appears that these SFPs could be the reason for SCAR marker 354 segregating with gender, as the phylogenetic analysis of the *S. arctica* and *S. reticulata* populations seems to indicate that male and female individuals are segregating according to gender, though the gender-bias is not statistically significant.

Studies of Arctic willow species have suggested that the gender-bias observed in populations of some species could be explained by niche partitioning or gender-biased mortality due to variation in physiology between males and females. In populations of *S. polaris* and *S. herbacea* a 60:40 female to male gender ratio has been documented, and in *S. polaris* it has been observed that females have greater stomatal resistance than males, which perhaps makes females less susceptible to desiccation during the short Arctic growing season (Crawford and Balfour, 1983). In *S. arctica*, 2:1 female to male gender-biases have been observed, and many physiological differences, such as leaf size and

water use efficiency, have been documented between males and females of this species (Dawson and Bliss, 1989). Growth chamber experiments using *S. arctica* seem to indicate that there is a genetic basis for the sex-specific differences between the genders in this species (Dawson and Bliss, 1989), and therefore that the gender-biased distribution of females in wetter habitats and males in dryer habitat is not caused solely by gender specific mortality.

It appears that once the genetic mechanism for determining gender in *Salix* is discovered, it will be possible in some species to link the genetics of secondary sexual characteristics such as leaf size or drought tolerance to gender. In a harsh environment such as the arctic, with short growing seasons, it may be important for the ecological requirements of the genders, and therefore their physiology, to diverge to minimize competition between the genders for limited resources (Cox, 1981).

2.5.4 Potential utility of this region

The consistent amplification of the conserved Ssu72-like gene-associated region in the *Salix* species sampled indicates that this marker may be useful for population and phylogenetic studies in willow. There are numerous SFPs found in the sequences both within and between species. For instance there is an 8 bp deletion in the genome project sequence relative to the two other individuals of *Populus trichocarpa* sequenced (Figure 2.2). Sequence divergence in the conserved region is 0-5.4% within *Salix*, and 7.7-9.9% between *Salix* and *Populus*. In addition to sequence variation between *Salix* species, there are also indications of heterozygous nucleotide positions (from double peaks on the sequencing electropherograms) that are consistent when sequenced in both directions.

This raises the possibility that in the future association studies might be performed between individual SFPs and gender (or other characters).

The conserved region has an apparent homologue in the poplar genome, where it corresponds to the putative promoter region of an Ssu72-like gene (involved in transcriptional start site regulation) on chromosome 15. The poplar genome sequence was used to design gene-anchored primers that consistently amplify this region and part of the Ssu72-like coding region in willows as well as poplars, irrespective of species and gender. The gene-anchored primers amplify a region that, while conserved, has numerous single feature polymorphisms (SFPs) both within and between species. This region could thus be used for population and phylogenetic studies.

2.5.5 Considering this research and version 2.0 of the *Populus trichocarpa* genome.

I completed this research prior to the release of the second version of the *Populus trichocarpa* genome, but as this resource is now available, I revisited the results in order to include the most current gene annotation information available for the sequences I worked with for this project. Notably, the content of Table 2.2 has changed significantly due to the changes in genome assembly and gene model names used between version 1.1 and version 2.0 of the poplar genome.

I performed a BLAST search of the sequences for the genes listed in Table 2.2 against version 2.0 of the poplar genome (http://www.phytozome.net). I was unable to find the gene model fgenesh1_pg.C_LG_XV000375 on version 2.0 of the genome, but all of the other genes had their best BLAST hits on chromosome 15, the same chromosome on which they were located on in version 1.1. The order of the genes was more or less

conserved as well, except that the poplar Ssu72-like gene (fgenesh1_pg.C_LG_XV000380, POPTR_0015s04700 on version 2.0) is not at the centre of the list of putative genes in the region surrounding the location of the *Populus trichocarpa* sequence on chromosome fifteen that shows homology to the SCAR 354 *Salix* sequence (Table 2.5). The positional order of the nine genes is conserved in version 2.0 of the genome, as are the gene functional annotations for the most part (Appendix B). Three of the genes listed on version 1.1 of the genome (grail3.0043012001, grail3.0043012101, grail3.0043012201) are now considered to be one single gene on version 2.0 of the genome (POPTR_0015s04830).

Table 2.5 Comparison of the position of the list of putative genes in the region surrounding the location of the *Populus trichocarpa* sequence on chromosome 15 that shows homology to the SCAR 354 *Salix* sequence in version 1.1 and version 2.0 of the poplar genome.
Genes are listed in the 5' to 3' direction on chromosome 15, and the gene fgenesh1_pg.C_LG_XV000380 / POPTR_0015s04700 encodes for the Ssu72-like protein.

Gene order	Location	Version 1.1 Gene name	Version 2.0 gene name	Location	Gene order
- 5	4162512 - 4165784	fgenesh1_pg.C_LG_XV000375	Not Found	N/A	N/A
- 4	4172930 - 4174370	fgenesh1_pg.C_LG_XV000376	POPTR_0015s04750	4964515 – 4966108	- 4
- 3	4176495 - 4177658	estExt_fgenesh4_pg.C_LG_XV0383	POPTR_0015s04760	4968103 – 4969266	- 3
- 2	4181734 - 4182456	eugene3.00150420	POPTR_0015s04770	4973382 – 4974227	- 2
- 1	4186527 - 4190414	estExt_Genewise1_v1.C_LG_XV1107	POPTR_0015s04780	4978598 – 4982760	- 1
0	4199692 - 4203247	fgenesh1_pg.C_LG_XV000380	POPTR_0015s04700	4923219 – 4926558	- 5
+ 1	4207845 - 4210979	eugene3.00150423	POPTR_0015s04810	4999929 – 5003025	+ 1
+ 2	4214615 - 4217179	fgenesh1_pg.C_LG_XV000390	POPTR_0015s04820	5011059 – 5013623	+ 2
+ 3	4229045 - 4229320	grail3.0043012001	POPTR_0015s04830	5022992 – 5030320	+ 3
+ 4	4229463 - 4229628	grail3.0043012101	POPTR_0015s04830	5022992 – 5030320	+ 3
+ 5	4229950 - 4233882	grail3.0043012201	POPTR_0015s04830	5022992 – 5030320	+ 3

2.6 Conclusions

I was able to identify the position of the SCAR 354 marker on chromosome 15 on the *P. trichocarpa* genome, and amplified the SCAR 354 marker sequence as well as the adjacent gene sequence for a Ssu72-like protein. This indicated that the position of this marker with respect to this gene is conserved in *P. trichocarpa* and the *Salix* spp. I was unable to confirm that the SCAR 354 marker that segregated with sex in *S. viminalis* was also a sex-linked marker in other *Salix* species or in *P. trichocarpa*.

By sequencing the PCR products amplified by the gene-anchored primers, I characterized the SCAR marker sequence, and identified distinct regions of it that varied in the amount that they were conserved between the species I sampled. When I investigated the gene-anchored sequence obtained in males and females of *S. arctica* and *S. reticulata* I found some evidence that gender-biased SNPs do exist in this sequence, which may explain why the SCAR 354 marker segregates with gender in pedigreed families of *S. viminalis* (Alström-Rapaport et al., 1998), though the gender-bias I observed was not statistically significant.

Given that the conserved Ssu72-like gene associated region consistently amplified in the *Salix* species I sampled, this marker may be useful for phylogenetic and population studies of willows. It may be possible to discover species specific single feature polymorphisms in the gene-anchored region my primers amplified, and these SFPs could prove to be useful genetic tools for identifying *Salix* species, which can be difficult to classify based on morphological traits alone.

Chapter 3: Investigation of chromosome 19 of the *Populus trichocarpa* genome to identify molecular genetic markers that can be used to identify gender

3.1 Introduction

3.1.1 Investigation of sex chromosomes and sex loci to identify gender-determining genes

Gender-determining genes initiate the developmental process that results in male and female phenotypes. Plant genes have been found to be fairly compact and usually grouped together in clusters, surrounded by repetitive DNA sequences, even in large genomes (Kellogg and Bennetzen, 2004). When heteromorphic sex chromosomes were first observed in dioecious plant species such as *Silene latifolia* by Blackburn (1923), researchers began investigating these genomic regions looking for gender-determining genes using gene cloning techniques to enrich DNA libraries with sex-linked transcripts or sex chromosome sequences (Shibata et al., 1999). This approach has been used in dioecious species that have well defined sex chromosomes like *Rumex acetosa* and *Silene latifolia*, and it resulted in the identification of a few genes that may be involved in the gender-determination pathway in *S. latifolia* (Zluvova et al., 2006). However, the majority of dioecious plants do not have sex chromosomes, so this approach to investigating gender-determination has not identified the genetic mechanism involved in the majority of plant species studied, despite the numerous studies that have investigated the role of sex chromosomes in determining sex in haploid, diploid and triploid plants (Sakamoto et al., 2005).

3.1.2 Sex-determination in the genus *Populus*

Given the 1:1 sex ratios that have been observed in *Populus* populations, gender-determining genes probably segregate at one locus, or as several closely linked loci on an autosome. In the genus *Populus*, previous studies have indicated that chromosome 19 appears to be consistently involved in the genetics of gender, though there is some discussion as to which gender is homogametic and which is heterogametic, depending on the species studied. There have been four main studies on the genetics of gender in *Populus*. In the first study of this kind, Yin et al. (2008) identified a gender-associated locus that consistently mapped to the peritelomeric region of chromosome 19 in *P. trichocarpa* using microsatellite markers, and their results suggested that the female is the heterogametic gender. In 2008 Gaudet et al. also produced genetic linkage maps of *Populus nigra* L. using AFLPs, SSRs, and SNPs, and found a sex-linked marker that mapped to a terminal position on chromosome 19 in the male parent, suggesting that the male is the heterogametic gender. Genetic linkage maps in aspen (*P. tremula* x *P. tremuloides*) constructed using AFLP and SSR markers showed that a sex-linked marker could be mapped to a non-terminal position on chromosome 19 in the male *P. tremuloides* (Pakull et al., 2009). An AFLP and SSR marker-based genetic map of *Populus alba* L. also located the sex-determining locus at a non-terminal position on chromosome 19 of the female parental map (Paolucci et al., 2010). The results of these four studies are summarized in Figure 3.1. There have also been RAPD markers related to a sex locus reported in *P. tomentosa* by a Chinese research group, but no genetic location has been assigned to these markers (Hou et al., 2009).

As seen in Figure 3.1, while the sex locus in all four species investigated is located on chromosome 19, its position on chromosome 19 (either terminally or centrally located), and which gender is heterogametic seems to vary with species, with evidence that females are heterogametic in two species, and males are heterogametic in the other two species (Paolucci et al., 2010). While these observations may seem contradictory, there is evidence in other species that hybrid zones may play a role in plasticity of sex-determination and the dynamics of how sex chromosomes evolve.

It is well known that many members of the genus *Populus* interbreed and form viable hybrids easily (Hamzeh et al., 2006), so it may be that this ability has allowed *Populus* species to maintain a certain amount of flexibility in their sex-determination mechanism. The Japanese frog, *Rana rugosa*, forms population groups depending on the type of heterogamety and sex chromosome morphology (Hamzeh and Dayanandan, 2004; Janousek and Mrackova, 2010). In this species of frog there are five population groups, three that have a XX/XY sex-determination system, and two where the female is the heterogametic sex (ZZ/WZ) (Ogata et al., 2007). From the phylogenetic data on the species, it appears that the Y and Z sex chromosomes arose in populations from Western Japan, and the X and W sex chromosomes in central Japan in geographical isolation, and subsequently came together again and reciprocally hybridized, resulting in the different sex-determination systems observed within the same species (Janousek and Mrackova, 2010). Experimental breeding programs (Ogata et al., 2003) and theoretical studies (van Doorn and Kirkpatrick, 2007), support the idea that in populations that have a strong sex bias, strong positive selection pressure for the minor sex-favoring gene could reestablish a 1:1 sex ratio in a population. This is because a strong sexual bias will promote sexually

antagonistic genes to accumulate, resulting in heteromorphic sex chromosome formation or even the acquisition of a novel dominant sex-determining locus, and the evolution of a novel sex-determination system (van Doorn and Kirkpatrick, 2007). Given that the first

Pedigree	*P. deltoides* x *P. nigra* BC1	*P. nigra* x *P. nigra*	*P. alba* x *P. alba*	*P. tremula* x *P. tremuloides*
Heterogametic sex	female	male	female	male
No. progeny analyzed	312	93	130	57
No. male : No. female	197 : 115	53 : 40	84 : 46	37 : 20
Segregation ratio	skewed	Mendelian	skewed	skewed
LG XIX map	Consensus	*P. nigra*	*P. alba*	*P. tremuloides*
Reference	Yin et al. 2008	Gaudet et al. 2008	Paolucci et al, 2010	Pakull et al. 2009

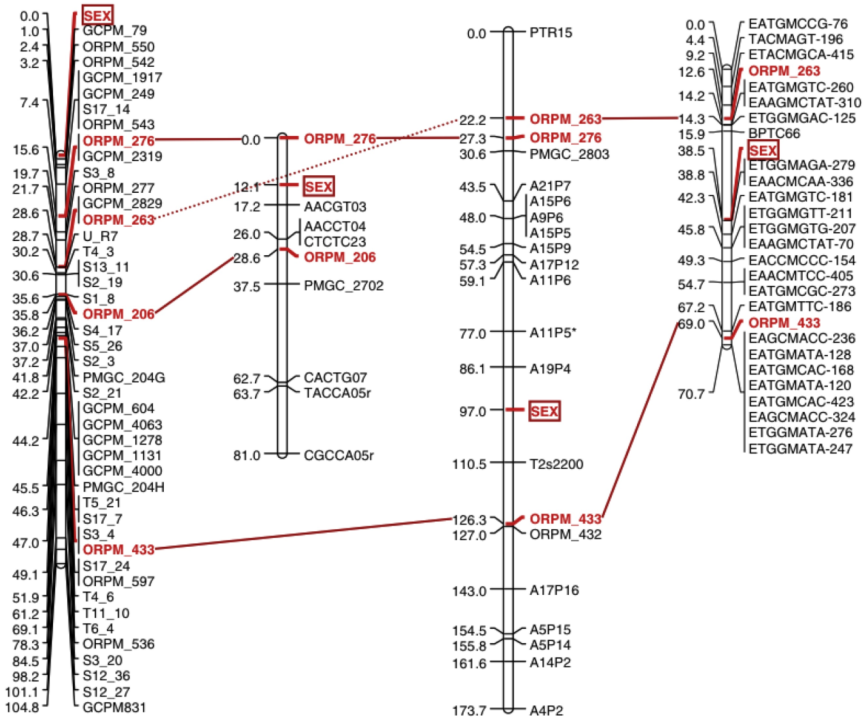

Figure 3.1 Summary of the data available on the position of the putative sex locus on chromosome 19 in the genus *Populus*.
The putative sex locus of each chromosome is indicated by a red box. Markers indicated in red are genomic markers that are shared across all four *Populus* species studied. The numbers on the left of each chromosome indicate the marker positions in centiMorgans. The numbers in black on the right of each chromosome indicate the SSR markers used in mapping the sex loci. (Figure originally published by Paolucci et al., 2010 Copyright Springer-Verlag).

poplar fossils date back to the Eocene period (Cronk, 2005) it is quite plausible that a similar process could have affected the evolution of sex-determination in *Populus*, resulting in the different male or female herterogamy observed among the species studied thus far in the genus.

3.1.3 Features of chromosome 19 that may be involved in gender-determination

Features of chromosome 19 on version 1.1 of the poplar genome such as a marker associated with gender-determination mapped to this chromosome, and a region of recombination suppression on chromosome 19, as well as a haplotype divergence observed for this chromosome, suggest that it is involved in sex-determination in *Populus trichocarpa* (Yin et al., 2008). This research also indicated that a number of previously unassembled scaffolds, principally scaffold 117, seem to overlap with the peritelomeric region of chromosome 19 that appears to be associated with sex-determination in *P. trichocarpa.*

One of the lines of reasoning put forward by Yin et al. (2008) to support the existence and position of a sex locus in *P. trichicarpa* was the evidence for recombination suppression on the telomeric end of chromosome 19. Scaffold 117, a 1MB segment of the *P. trichocarpa* genome containing ~80 gene models, was mapped to the peritelomeric region of chromosome 19 using 27 microsatellite markers in version 1.1 of the *P. trichocarpa* genome, in the region proposed to contain the sex locus (Yin et al., 2008). Fine scale mapping for scaffold 117 indicated no recombination within the upper 706kb region of chromosome 19, while the lower 257Kb region of this scaffold included several recombination positions with chromosome 19. Significant segregation distortion for alleles located on maternal haplotypes was observed between scaffold 117 and

chromosome 19, whereas no distortion within paternal haplotypes was observed between these two genomic regions (Yin et al., 2008).

Based on these results, scaffold 117 potentially represents a divergent haplotype of the telomeric end of chromosome 19, and the distortion of the alleles located on the maternal haplotype provides evidence that the female is the heterogametic sex in *P. trichocarpa.* Following the reasoning that the telomeric end of chromosome 19 of the poplar genome may contain a sex locus, I decided to investigate gene sequences in the region of chromosome 19 that now incorporates the 1MB scaffold 117, the putative heterogametic haplotype of chromosome 19, looking for SNPs that segregate with gender.

3.2 Objectives

In order to investigate how gender is regulated in *Populus trichocarpa*, I identified two objectives based on the availability of the sequenced genome for this species, and previous work done that identified chromosome 19 of this genome as being involved with the genetics of gender differentiation in this species (Yin et al., 2008). My first objective was to identify if the homogametic gender contains two copies of telomeric region of chromosome 19, or two copies of a putatively sex-linked contiguous sequence (contig) labelled scaffold 117. My second objective was to develop a genetic marker that segregates with gender that may be associated with the genetic mechanism of gender-determination, by looking at SNP variation in gene sequences located on the 5' telomeric region of chromosome 19.

3.3 Methods

3.3.1 Collection and preparation of biological materials

I collected plant tissue from leaf buds of male and female *P. trichocarpa* trees in March of 2006 and 2007 from natural stands located on the UBC campus previously identified as males or females by Dr. Cronk. These tissues were stored at –80ºC to ensure that they would be suitable for RNA and DNA extraction. I extracted DNA from 22 individuals (13 female and 9 male) using a modified CTAB extraction protocol (Doyle and Doyle, 1987), and DNA samples were diluted to a concentration of approximately 50ng/µl in nuclease free water, and then stored at -20 ºC.

3.3.2 *In silico* investigation of the gene content of scaffold 117

Given that it is thought that the female is the heterogametic gender *in P. trichocarpa* (Yin et al., 2008*)*, I attempted to determine if the male chromosome 19 consists of two copies of chromosome 19 as it appears in the sequenced genome, or two copies of scaffold 117. The gene content of scaffold 117 was investigated to develop a marker that segregates with gender, and to identify if the homogamete contains two copies of chromosome 19 sequence or two copies of scaffold 117. A series of *in silico* searches were performed to identify all the gene models located on scaffold 117, then all the scaffold 117 gene model sequences were BLASTed against the assembled version 1.1 *P. trichocarpa* genome to look for gene models that could be used to identify scaffold 117 versus chromosome 19, and therefore identify which of these scaffolds was present in two copies in the homogametic gender. Once the genes that were unique to scaffold 117 were identified, primers were designed based on the gene sequence reported in

version 1.1 of the poplar genome, and I attempted to amplify these genes using DNA extracted from the collected leaf tissue samples.

3.3.3 PCR conditions and sequencing of samples

PCR primers for this project were designed using Primer3 (v. 0.4.0, http://frodo.wi.mit.edu/primer3/), and tested *in silico* to determine the probability of primer dimers or unspecific primer binding with Amplify 3X (v.3.1.4, http://engels.genetics.wisc.edu/amplify). Primers were ordered from Integrated DNA Technologies Inc (Coralville, Iowa), and a list of the primers used in this project is included in Appendix A. DNA amplification for this project was performed under the following reaction conditions: Each 25 µl reaction contained 2.5 µl of 10X Taq buffer (100mM Tris-HCl (pH 8.8), 500mM KCl, 0.8% Nonidet P40), 1 µl of 5µM forward primer, 1 µl of 5µM reverse primer, 2.5 µl of 2mM dNTPs (Fermentas Life Sciences, CA, USA), 2.5 µl of 25 mM $MgCl_2$, 0.25 µl *Taq* DNA polymerase (Fermentas Life Sciences, CA, USA), 100 ng of genomic DNA and ddH_2O made up to 25 µl. Polymerase chain reactions (PCRs) were run on an Eppendorf Mastercycler gradient thermocycler for 35 cycles under the following reaction conditions: Lid set to 94˚C, 94˚C 1 min, and 50-57˚C 1 min (temperature range was dependant on primer annealing temperature), with an initial 3 min 94˚C denaturation step and a final 1.5 min 72˚C polymerization step. Amplification products were visually scored for the presence or absence of bands on 1% agarose gels stained with GelRed™ Nucleic Acid Gel Stain, diluted in water (Biotium, Hayward, CA 94545). Amplified PCR products were then prepared for sequencing by Macrogen, in Maryland USA.

3.3.4 Sequence and polymorphism data analysis

SNPs in the gene sequences amplified for all individuals were identified using phred/phrap/consed/polyphred (Ewing and Green, 1998; Ewing et al., 1998; Nickerson et al., 1997). This program was also used to edit and trim the DNA sequences before forming contiguous sequences (contigs) that corresponded with each individual that was sequenced. ClustalX (Larkin et al., 2007) was used to align the DNA sequences obtained for each gene, and then these alignments were edited using BioEdit version 7.0.5.3, (10/28/05) (Hall, 1999), before using DNA Sequence Polymorphism (DNAsp) (Librado and Rozas, 2009) to run a phase analysis. DNAsp uses statistical methods to infer haplotype phase, which allows for haplotype reconstruction from genotypic information (Librado and Rozas, 2009). I used DNAsp to perform calculations of the number of haplotypes represented by the individuals for a given gene, to estimate the amount of recombination between adjacent sites, the % average pairwise nucleotide diversity/site, the % of synonymous nucleotide diversity/site, the % of nonsynonymous nucleotide diversity/site, and the % nucleotide diversity per sequence for all the gene sequences obtained in the *P. trichocarpa* males and females were sampled. Nucleotide diversity values (π values) were calculated according to Nei (1987) and not Jukes and Cantor (1969) so that the values reported by this research would be comparable to those found in the literature for *Populus* species (Nei and Miller, 1990). Network (4.6.0.0, fluxus-engineering.com) was used to construct the haplotype diagrams as visual representations of the haplotype data (Polzin and Daneschmand, 2003).

3.4 Results

3.4.1 Investigation of gene content of scaffold 117 and the telomeric end of chromosome 19

The first gene model of interest was annotated as a putative chloroplast terpene synthase that had three copies located on scaffold 117, and only one copy apparently located anywhere else in the poplar genome, on chromosome 19 (Table 3.1). The second group of genes consists of 10 gene models that are only found on scaffold 117 (Table 3.2). This research investigated the group of 10 genes unique to scaffold 117, as the putative chloroplast terpene synthase genes existed as multiple copies in the genome, making it difficult to develop specific primers that would differentiate between the different copies of these genes.

With the release of the second version of the poplar genome on January 8th, 2010, it was important to update the information on the gene models, and find out if the gene models found on scaffold 117 were still associated with the telomeric region of chromosome 19, or if the content of scaffold 117 had been broken up and included in multiple other places in the genome. On the second version of the poplar genome, it appears that scaffold 117 has been incorporated into the telomeric end of chromosome 19. When the gene models located on scaffold 117 (version 1.1) were BLASTed against version 2.0 of the *P. trichocarpa* genome, the majority of the top hits were on a 1.7MB region of chromosome 19, and the number of gene models dropped from approximately 80 to 43. This is because on version 1.1 of the genome, gene models sometimes indicated that separate exons of the same gene were separate genes. The sequences of the putative chloroplast terpene synthase genes all BLASTed to chromosome 19 of version

2.2 of the poplar genome, in a region between 1075484 and 7014599bp, overlapping with six gene models. More reassuringly, the ten genes that were unique to scaffold 117 were all still located on chromosome 19, though some of them also produced blast hits with lesser amounts of sequence similarity on scaffolds 1, 3, 73, 121, 182 and 879. However, because scaffold 117 was incorporated into the telomeric region of chromosome 19 in the new assembly, it appears that scaffold 117 does not represent a separate gamete of chromosome 19, but is more probably a fragment of chromosome 19 that was not aligned correctly in version 1.1 of the genome.

Primers were designed for 24 genes, just over half of the approximate 40 gene models in the region of chromosome 19 that corresponds to scaffold 117. PCR amplification of the 10 gene models unique to scaffold 117 was done (numbered one though ten, Figure 3.2). These genes were clustered in a 200kb region of scaffold 117, so genes were then chosen at 100kb intervals along the length of the scaffold to get a representative sample of SNP variation in genes on scaffold 117. Eight genes at 100 000bp intervals along the length of scaffold 117 were selected (labeled A though H, Figure 3.2). Given the differences in gene model annotation between version 1.1 and version 2.0 of the poplar genome, it is not surprising that many of the primers that were designed based on gene model sequence from version 1.1 of the genome did not successfully amplify the target sequences, or failed to produce PCR products at all. A list of the putative gene functions and location of all the genes worked with for this project is included in Appendix B.

3.4.2 Analysis of sequence and polymorphism data obtained for eleven genes located on the telomeric region of chromosome 19

After performing a preliminary analysis of the haplotype data based on sequence obtained for using the gene model notation from version 1.1 of the poplar genome (http://genome.jgi-psf.org/Poptr1_1/Poptr1_1.home.html), information on the genes of interest in this study was updated to that available on version 2.0 of the genome, and thereafter version 2.0 of the poplar genome (http://www.phytozome.net) was used as the source of genomic information for designing primers for genes positioned on the telomeric end of chromosome 19. After looking at the SNP recombination rates for genes that sequence data had been collected for, gene SCA_117_10 showed the most interesting pattern of SNPs with regards to gender (Figure 3.3). This is a pattern of SNPs that would be expected in the region of a sex locus where recombination between alleles is suppressed. If these SNPS were sex-linked it would be expected that all the male haplotypes would group together, separate from the female haplotypes and vice versa. Given that this was the only gene that I sampled that showed this haplotype distribution, I decided to sequence genes on either side of this one to determine if they displayed similar patterns in SNP variation and recombination rates between adjacent sites (labeled SCA_19_1 through 5, Table 3.3). Genes that were evenly distributed in a 55 000 bp region on either side of the gene SCA_117_10 were chosen to see if the low recombination rate and distribution of haplotypes specific to males or females occurred in other genes in this region.

After testing the primers for the 24 genes sampled from scaffold 117 for amplification in the sample of 22 individuals (13 females and 9 males), good quality

sequence was obtained for eleven genes (Table 3.3). Then a haplotype analysis was performed to look at SNP variation and see how it relates to sequence differences between gender (Table 3.3). When the estimated recombination between adjacent sites (R-values) was investigated for the genes located in the region of suppressed recombination on the telomeric end of chromosome 19 quite a lot of variability was found.

Table 3.1 Gene models located on scaffold 117 and chromosome 19 annotated as putative chloroplast terpene synthase in *P. trichocarpa*.

Gene name	Location on *P. trichocarpa* genome
fgenesh4_pg.C_scaffold_117000025	Poptr1_1/scaffold_117:421459-422086
eugene3.01170028	Poptr1_1/scaffold_117:428369-431855
fgenesh4_pm.C_scaffold_117000002	Poptr1_1/scaffold_117:539441-542289
grail3.0085006801	Poptr1_1/LG_XIX:2752362-2753129

Table 3.2 Gene models unique to scaffold 117, updated with names and positions on version 2.0 of the *P. trichocarpa* genome.

V 1.1 gene name and assigned code	V 2.0 best match gene name and location
eugene3.01170047(SCA_117_1)	POPTR_0019s01520 sca_19:1339019-1340779
fgenesh4_pg.C_sca_117000045(SC_117_2)	POPTR_0019s01530 sca_19:1347076-1348225
grail3.0117003001(SCA_117_3)	POPTR_0019s01540 sca_19:1348343-1349893
gw1.117.122.1 (SCA_117_4)	POPTR_0019s01560 sca_19:1379318-1382368
fgenesh4_pg.C_sca_117000051(SCA_117_5)	POPTR_0019s01570 sca_19:1391060-1395729
fgenesh4_pg.C_sca_117000053(SCA_117_6)	POPTR_0019s01560 and POPTR_0019s01570
fgenesh4_pg.C_sca_117000054(SCA_117_7)	POPTR_0019s01560 and POPTR_0019s01570
e_gw1.117.150.1(SCA_117_8)	POPTR_0019s01630 sca_19:1438061-1439700
gw1.117.169.1(SCA_117_9)	Between POPTR_0019s01660 sca_19:1460825-1462203 and 1670 sca_19:1463832-1466771
eugene3.01170072(SCA_117_10)	POPTR_0019s01790 sca_19: 1546894-1550419

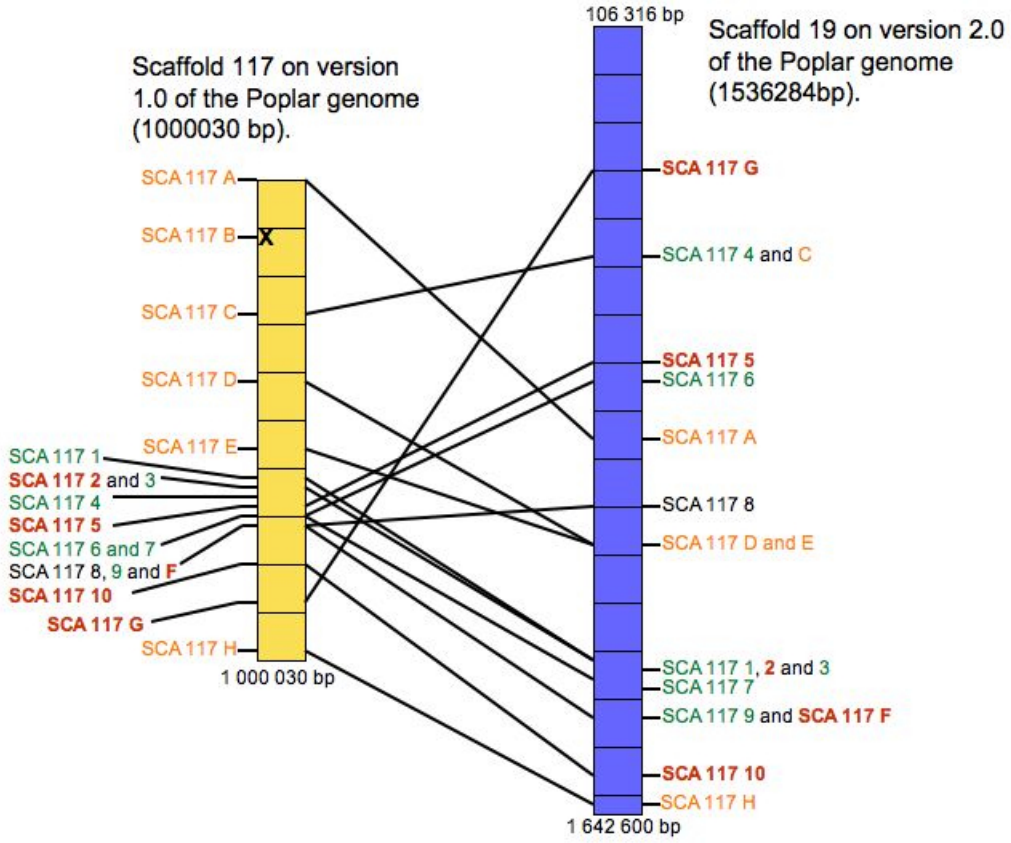

Figure 3.2 Genes sampled along the length of scaffold 117 (version 1.1) and their corresponding positions on chromosome 19 (version 2.0).
Bars indicate 100 000bp increments on scaffolds. Orange indicates no sequence data was obtained for the gene. Green indicates only 200bp of gene was sequenced, and no SNPs of interest were found/no sequence was produced so work with them was discontinued. Red indicates that good sequence for at least 800bp of the ORF of the gene was obtained. The 'X' by gene SCA_117_B indicates that this gene model did not exist in version 2.0 of the poplar genome.

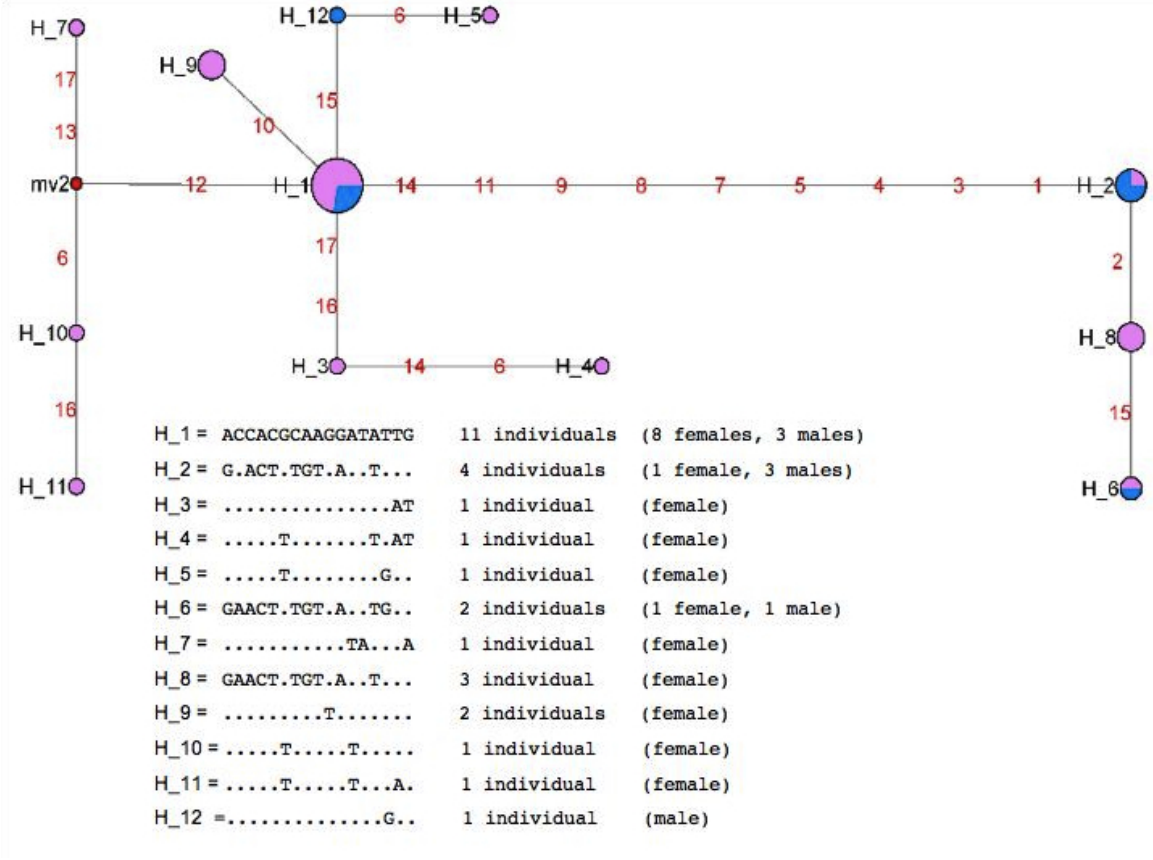

Figure 3.3 Haplotype diagram for gene SCA_117_10 (gene model POPTR_0019s01790) based on 772bp sequenced in ten females and four males.
In this diagram the circles represent nodes where pink indicates the proportion of female haplotypes, blue for male, and size of node is proportional to number of individuals with that haplotype. The length of the lines in not proportional to distance, but the red numbers indicate base pair position of SNPs. The red node indicates a haplotype that must exist for other nodes to be present, but was not observed in the sequences sampled.

Table 3.3 Overview of polymorphism data for genes located on the telomeric end of chromosome 19.
n = number of chromosomes, S = number of variable sites, H = number of haplotypes, R = estimate of recombination between adjacent sites, π/site indicates % average pairwise nucleotide diversity/site, π_{sil} indicates % of synonymous nucleotide diversity /site, π_{syn} indicates % of synonymous nucleotide diversity /site, π_{nonsyn} indicates % of nonsynonymous nucleotide diversity /site, and θ is the % nucleotide diversity per sequence.

Gene	Gene ID on chromosome 19	n	Length*(bp)	S	H	R (%)	π/site (%)	π_{sil} (%)	π_{syn}(%)	π_{nonsyn}(%)	θ/site (%)
SCA_117_A	POPTR_0019s01120	39	847	23	35	13.24	0.713	NA	0.659	0.736	0.643
SCA_117_2	POPTR_0019s01530	31	673	10	11	1.49	0.287	0.292	0.000	0.000	0.394
SCA_117_5	POPTR_0019s01570	31	815	9	2	0.00	0.071	0.061	0.233	0.082	0.276
SCA_117_F	POPTR_0019s01670	45	853	13	15	0.23	0.338	NA	0.643	0.263	0.349
SCA_19_1	POPTR_0019s01700	37	528	29	21	1.31	1.804	3.031	3.248	1.266	1.406
SCA_19_2	POPTR_0019s01740	35	842	22	31	12.49	0.764	1.618	2.059	0.335	0.750
SCA_19_3	POPTR_0019s01780	43	731	12	19	3.03	0.412	0.258	0.190	0.568	0.413
SCA_117_10	POPTR_0019s01790	29	772	17	12	0.22	0.724	0.594	0.000	0.000	0.594
SCA_19_4	POPTR_0019s01830	31	838	14	15	1.98	0.452	0.724	0.623	0.284	0.448
SCA_19_5	POPTR_0019s01850	17	804	17	17	13.63	0.757	1.321	0.392	0.364	0.625
SCA_117_G	POPTR_0019s01880	37	832	23	25	3.16	1.176	1.411	1.743	0.875	0.749
Average:		34.1	775.9	17.2	18.5	4.62	0.682	1.034	0.890	0.434	0.604

*Length: excluding sites with gaps/missing data
*n = this number of chromosomes is calculated by multiplying the number of diploid individuals sequenced for each gene, and adding the single haplotype of the reference sequence "Nisqually 1" from version 2.0 of the poplar genome. Example: For SCA_117_2 15 individuals were sequenced.

3.5 Discussion

3.5.1 Identifying if the homogametic gender of *Populus trichocarpa*

Identifying which gender is the homogametic one with respect to gender *in P. trichocarpa* based on sequence data from chromosome 19 was complicated by the discrepancy in assembly data between version 1.1 and version 2.0 of the poplar genome. This discrepancy in alignment between the two assemblies of the poplar genome could be due to the fact that different software was used to create the two assemblies. Version 2.0 of the poplar genome, accessible at http://www.phytozome.net, was assembled using ARACHNE version 20071016HA, and the assembly covers 403 million base pairs of sequence with an average read depth of 7.45x assembled. ARACHNE is a whole-genome shotgun assembler that analyzes paired forward and reverse sequence reads that are obtained from both ends of plasmid clones, and assembles them with a reported 99% accuracy due to its ability to detect alignment errors caused by sequencing errors or false sequence overlaps caused by repeat DNA sequences in complex genomes (Batzoglou et al., 2002).

Version 1.1 of the poplar genome, accessible at http://genome.jgi-psf.org/Poptr1_1/Poptr1_1.home.html, was assembled using JAZZ, a group of programs developed by the Joint Genome Institute (JGI) designed specifically for working with large sequencing projects (Taylor and Semple, 2002). The first version of the poplar genome was estimated to consist of approximately 485 million base pairs and was sequenced with a read depth of approximately 7.5x (Tuskan et al., 2006). JAZZ uses a multistep process similar to that of ARACHNE to assemble whole genome shotgun

sequences, but has stricter assembly parameters that make it more likely that haplotypes will assemble separately (Shapiro, 2005).

3.5.2 Version 1.1 versus version 2.0 of the poplar genome

The main difference between these two genome assemblers appears to be the way in which they assemble supercontigs (ARACHNE) or scaffolds (JAZZ) which is how the genome assemblers take the shotgun sequences and group them into large contiguous assemblies that match up with the number of chromosomes in the genome. ARACHNE creates supercontigs by using the forward and reverse links from plasmid reads to orient and order unique contigs into longer sequences (Batzoglou et al., 2002). The JAZZ assembler attempts to build a scaffold by iteratively building and breaking sequence contigs, progressively including lower-quality sequence data (Taylor and Semple, 2002).

Because of the discrepancy between the way in which the two versions of the genomes were assembled, it is difficult to know if scaffold 117 really represents an alternate haplotype of chromosome 19, that was detected by the JAZZ assembler (version 1.1) but not the ARACHNE assembler (version 2.0), or if the different way in which ARACHNE bridges gaps in the poplar genome assembly correctly incorporated the smaller scaffold 117 from the first version of the genome into the telomeric end of chromosome 19 in version 2.0 of the poplar genome. The sequence difference between the two gender-determining haplotypes of chromosome 19 may be too small to be detected at this point, even if both copies of chromosome 19 were sequenced since Nisqually-1, the genotype selected to be sequenced for the *Populus trichocarpa* whole genome assembly, was a female (Tuskan et al., 2006), the reported heterogametic gender in this species.

3.5.3 Developing a genetic marker on chromosome 19 that segregates with gender in *Populus trichocarpa*

I investigated 24 genes located on the 5' telomeric end of chromosome 19 that was indicated to have suppressed recombination with scaffold 117 on version 1.1 of the poplar genome, looking for SNPs located in genes in this region that segregate with gender. There was a large amount of variability in the number of SNPs, or variable sites (S), detected in the sequences obtained for the genes that were studied (Table 3.3), however, none of the SNPs that were detected in any of the genes segregated with gender so I was unable to discover a genetic marker that could be used to sex *P. trichocarpa* individuals of unknown gender. Although the number of individuals that were sampled for each gene varied from eight to twenty-two, it appears that neither the length of sequence obtained for each gene, not the number of variable sites or the number of haplotypes observed, were correlated with the number of individuals sampled (Table 3.3).

3.5.4 Investigation of recombination rates in genes located in the telomeric region of chromosome 19 in *P. trichocarpa*

Based on the results of the investigation into the recombination rates of genes located on the 5' telomeric end of chromosome 19, no overall trend of reduced recombination between adjacent SNPs was found in the gene sequences examined. In fact there was considerable variation in R-values even in genes that were adjacent to each other on the genome. An example of this is illustrated by the genes SCA_19_4 and SCA_19_5, which are only 21 134bp apart on chromosome 19, but differ in % R-value by a factor of 10. Three of the 11 genes I looked at, SCA_117_5 (POPTR_0019s01570), SCA_117_F (POPTR_0019s01670), and SCA_117_10 (POPTR_0019s01790), seemed

to show much lower recombination (Table 3.3). A pattern of low recombination between SNPs would be expected in the region of a sex locus where recombination between alleles is suppressed. Suppression of recombination would be expected in the region of a sex locus in order to maintain separate genders, but my examination of genes located on either side of these genes shows that these low R-values are not maintained across this region of chromosome 19.

3.5.5 Sex-linked markers on chromosome 19 of *P. trichocarpa*

It appears that it is possible to develop genetic sex-linked markers that map to chromosome 19 of *P. trichocarpa*, as three SSR markers that were identified in an interspecific cross between *Populus tremula* L. and *Populus tremuloides* Michx. were mapped to a central location on chromosome 19 (Pakull et al., 2011). It is interesting to note that none of the markers developed in the *P. tremula* x *P. tremuloides* cross were mapped to the region of chromosome 19 in *P. trichocarpa* that contained the genes that were investigated in this study. Also, all the fully sex-linked SSR markers that were mapped to a central position on chromosome 19 in *P. trichocarpa* appeared to be inherited from the *P. tremuloides* male parent (Pakull et al., 2011). These results seem to indicate that the male is the heterogametic gender in *P. trichocarpa*, which contradicts the finding of Yin et al. (2008) that provided evidence that in this species the female is the heterogametic gender, and also that a sex locus is located at the telomeric position on chromosome 19. It is possible that the evidence that indicates that different species of *Populus* show differences between which gender is heterogametic and where the sex locus is located on chromosome 19 can be explained by the idea that the evolution of sex paralleled speciation in this genus (Paolucci et al., 2010). However, I think it is more

likely that the emerging genetic evidence supports the theory that gender in *Populus*, and also in its sister genus *Salix*, is controlled by several sex-determining loci (Alström-Rapaport et al., 1998), that in *Populus* appear to be located on chromosome 19.

3.5.6 Nucleotide diversity of genes located on the telomeric end of chromosomes 19 in *P. trichocarpa*

Genetic variation in a species can be measured by looking at nucleotide diversity (π), and measuring genetic variation is important because it is related to most phenotypic variation observed and also can give clues to the evolutionary history of a species (Gilchrist et al., 2006). When I compared the average value of π/site (0.682%) observed for the coding regions of the eleven genes I sampled (Table 3.3) to the average level of nucleotide diversity (0.184%) observed in *P. trichocarpa* according to Gilchrist et al. (2006), I find that my average observed level of nucleotide diversity is higher, though my sample sizes were much smaller than the ones reported by Gilchrist et al. (2006). However, at least one of the genes I investigated, SCA_117_5, showed a π/site value of 0.071% (Table 3.3), which is in the same range of the π/site values for the genes investigated by Gilchrist et al. (2006), and this study also found that nucleotide diversity varies a great deal between genes, which is consistent with my observations.

However, a more recent study of nucleotide diversity found an average π/site value of 3.38% for 3 separate genes, and this study had a sample size of 15 individuals, with similar lengths of sequence analyzed per gene to what I reported (Breen et al., 2009). It would appear that the nucleotide diversity I observed for the genes falls within an acceptable range for *P. trichocarpa*, with π/site values being gene dependant.

There are two likely explanations for my observation of much higher average amount of nucleotide diversity than the values reported by Gilchrist et al. (2006). First, I chose the genes based on their position on chromosome 19 to investigate recombination rates in this region of the poplar genome thought to contain a sex locus, and the genes in this study have functions annotated based only on sequence homology to genes characterized in other species, as yet unconfirmed by research in *P. trichocarpa*. The genes investigated by Gilchrist et al. (2006) were chosen based on data from evolutionary studies from other organisms, or because of their known functions in pathogen defense or wood quality, and therefore are probably under stricter genetic selection to avoid the accumulation of sequence mutations that would affect gene function (Gilchrist et al., 2006). Choosing genes without prior knowledge of their function increases the likelihood that the data from the loci will provide a more accurate and less biased view of genome wide patterns of polymorphism (Ingvarsson, 2008), so it may be that the genes I chose to work with represent a wider pattern of nucleotide diversity than those sampled by Gilchrist et al. (2006).

Secondly, in this study between 8 and 22 individuals were sampled from one stand of trees, while Gilchrist et al. (2006) looked at between 37 and 40 trees, sampled from a wide geographic area, depending on the gene being investigated. Those researchers also sequenced gene fragments that were slightly longer than the sequences I investigated. The smaller tree sample size from a small geographic area, and shorter gene sequences used in this study may mean that the data reported here is not a good estimate for the average nucleotide diversity in the genes studied.

3.5.7 Nucleotide diversity observed in other species of *Populus.*

Nucleotide diversity has been investigated in several other species of *Populus*, notably *P. balsamifera*, which is the sister species to *P. trichocarpa* (Olson et al., 2010) and *P. tremula*, which is a more distant relative. The estimates for nucleotide diversity reported for three genes studied in *P. balsamifera* were similar in range to the π/site values found for the genes in this study, and Breen et al. (2009) had a comparable sample size (between 5 and 18 individuals) with an average sequence length of 609bp, about 100bp shorter than the average length of the gene sequences in this study (775.9bp, Table 3.3). π/site values reported for *P. deltoides* for the same three genes, also from the study by Breen et al. (2009), were also similar in range, and provide comparable values to support this. Several studies of *P. tremula* (Ingvarsson, 2005, 2008) have shown that this species has a high level of genetic variation relative to other *Populus* species. Nucleotide diversity (π/site) is reported at 1.1%, which is higher than my reported π/site of 0.682%, and five times greater that the average π/site values reported for *P. trichocarpa* (Gilchrist et al., 2006) and *P. balsamifera* (Breen et al., 2009).

North American poplar species exhibit approximately 50% less population differentiation and lower nucleotide diversity compared to the values observed *in P. tremula*, but this difference may be explained by the fact that several of the studies of the North American species were conducted over small geographic scales compared to the studies of *P. tremula*, and may not have captured a species-wide estimate of nucleotide diversity (Breen et al., 2009). The lower nucleotide diversity observed in *P. balsamifera* compared to *P. tremula* may also indicate a historically lower effective population size in *P. balsamifera*, caused by population wide bottle necks experienced in response to the

expansion and retraction of ice sheets during the Quaternary period (Hewitt, 2004), followed by a recent population expansion (Keller et al., 2010). The range of π/site values found in my data is comparable to that reported in the literature for *P. trichocarpa* and other *Populus* species, so the π/site values reported in this study for the eleven genes that were sequenced adds to the information available on nucleotide diversity in *P. trichocarpa*.

3.6 Conclusions

I showed that scaffold 117 reported in the version 1.1 genome sequence of *P. trichocarpa* appears to have been a fragment of chromosome 19 in the poplar genome. There was little re-arrangement of the gene order between scaffold 117 of version 1.1, and the telomeric end of chromosome 19 on version 2.0 of the poplar genome. The gene model annotations in this region changed substantially as exons that were considered separate genes on version 1.1 of the genome were annotated as exons of the same gene on version 2.0 of the genome in many cases. I was ultimately unable to identify a way to distinguish the homogametic genotype from the heterogametic genotype in *P. trichocarpa* by using genomic markers in this region.

I investigated 24 genes in the telomeric region of chromosome 19 looking for SNP variation in the genomic sequence, or reduced recombination rates in SNPs between males and females in these genes, which could be associated with genetic markers for gender and a sex locus. I observed large variability in the number of SNPs detected in the gene sequences studied, but was unable to discover a genetic marker that could be used to sex *P. trichocarpa* individuals of unknown gender. I found no overall trend of reduced recombination between adjacent SNPs in the gene sequences that were worked

with, and it appears from the data I collected that low recombination rates between SNPs are not maintained across the telomeric region on chromosome 19. If there is reduced recombination in the region of a sex locus on the telomeric end of chromosome 19, it would appear that it is very localized, as I looked at genes located less than 100kb from each other. Alternatively, it is possible that the sex locus is not in the ~1 000 000bp region I investigated.

My investigation of nucleotide diversity in the telomeric region of chromosome 19 showed that nucleotide diversity varies a great deal between genes, even in the small region on the genome investigated. The π values I observed for the genes investigated fall within the range previously reported for *P. trichocarpa* (Breen et al., 2009; Gilchrist et al., 2006), with π/site values being gene-dependant.

It appears that the genetic mechanism for gender-determination in *P. trichocarpa* is more complicated than previously thought. However, research is now being conducted on gender-determination in multiple *Populus* species. These studies, along with maturing genetic resources such as the re-annotated version 2.0 of the *P. trichocarpa* genome make understanding the genetics of gender-determination in this species only a matter of time.

Chapter 4: Exploration of genome wide gender-biased gene-expression patterns to identify genes involved in gender-determination in *Populus trichocarpa*

4.1 Introduction

In the maintenance of separate genders, a sex locus functions to control differential gene-expression in males and females, as well as gender-biased gene-expression patterns. One approach to investigating gender-determination is to look at differential gene-expression patterns early in the development of male and female inflorescences or in vegetative tissues. Research into gene-expression and its role in gender-differentiation has provided information about the genetic basis for gender-determination in three ways.

The genes involved in controlling the development of male and female reproductive organs can be identified by relating specific gene-expression patterns in the floral meristems to organ position and development (Zik and Irish, 2003). Secondly, investigation of gene-expression differences between males and females in vegetative tissues may reveal genes that are responsible for secondary sexual characteristics. Many dioecious plant species exhibit gender-specific morphological differences in addition to their unisexual flowers, as well as small differences in ecological requirements (Lloyd and Webb, 1977). And thirdly, mapping genes that are differentially expressed between male and female individuals on to the physical map of the *P. trichocarpa* genome may indicate a region containing a sex locus. Clusters of gender-segregating genes may indicate areas of limited recombination in the genome, which are thought to be indicative of the initiation of sex chromosomes (Liu et al., 2004). Sex chromosomes start to evolve

when selection, acting through local suppression of recombination, tightly links all loci involved in gender-determination in a region that does not recombine, thereby ensuring that separate genders are maintained (Scotti and Delph, 2006).

4.1.1 The biology of gender-specific differences in dioecious species

One technique used for identifying genes involved in gender-determination uses homologous genes that are known to function in floral development in hermaphroditic model plants such as *Antirrhinum* and *Arabidopsis* to identify floral development genes that may control organ suppression in unisexual flowers (Ainsworth, 2000). There are many similarities between the genomes of *P. trichocarpa* and *A. thaliana* because they belong to the Eurorosid I clade of angiosperms (Jansson and Douglas, 2007) so it is plausible that extensive gene function is conserved between the two species. An example of this is the *Populus PTD* (*Populus trichocarpa DEFICIENS*) promoter gene, a B-class floral homeotic transcription factor, which is homologous to the MADS box genes *DEFICIENS* and *APETALA3*, that are found in *Antirrhinum* and *Arabidopsis*, respectively (Skinner et al., 2003). However, while there may be an evolutionary advantage for the genes to be inherited together without much recombination, gender-differentiation genes involved in floral development do not need to be to be located at a sex locus as their expression could be *trans*-regulated by genes located at a sex locus. In fact, an in silico survey of the *P. trichocarpa* genome found no poplar MADS box genes located on chromosome 19 (Leseberg et al., 2006), which is the chromosome thought to contain the sex locus in this species (Yin et al., 2008).

4.1.2 Secondary sexual characteristics in plants

Genes that show a gender-biased expression pattern in vegetative tissues may not be directly involved in gender functions in the plant but may produce secondary sexual characteristics that identify male or female phenotypes outside of the floral tissues. There are many examples of secondary sexual characteristics being expressed in plants. Sexual dimorphism between the genders can result in differences in the life histories of males and females in a given species, as seen in *Chamaelirium luteum*, where female plants begin flowering at a greater age than males, flower less frequently and show a higher mortality rate than male plants in the same population (Meagher, 1984). The differences between males and females can also be morphological, as is seen in *Asparagus officinalis* where males usually produce more vigorous growth early in the growing season (Bracale et al., 1991), and in *Cannabis sativa* females of this species do not grow as tall as the males, and have larger root systems and leaf blades (Lloyd and Webb, 1977).

In the genus *Populus*, numerous studies have indicated that there are vegetative morphological differences between males and females, as well as small gender related ecological preferences. Elevation seems to have an influence on the distribution of male and female individuals within populations of *Populus tremuloides* (Einspahr, 1960). Female *P. tremuloides* are more abundant than males at lower elevations, and female growth rates decreased with increasing elevation at more than twice the rate of males (Grant and Mitton, 1979). Though females of this species consistently showed a greater annual radial growth rate at all elevations than males, which contradicts a theory put forth by Lloyd and Webb (1977), sexual reproduction is more costly for females than for males (Grant and Mitton, 1979).

In *P. augustifolia* it was found that females are more flood tolerant than males, and a possible explanation for this is that seed-producing female trees located near streams would facilitate seed dispersal into recruitment areas along river beds (Nielsen et al., 2010). In *P. trichocarpa* it has been shown that male clones have higher dry weight fiber yields on average than females (Tschaplinski et al., 1994), though to date no reliable way has been developed to sex trees prior to their sexual maturity between the ages of five and ten years (Brunner, 2010), so investigating genes that are differentially expressed between males and females in vegetative tissues may yield a genetic marker for gender.

4.1.3 Clustering of differentially expressed genes near the sex locus

Once genes with statistically significant differential expression in males and females have been identified, their locations and linkage groups position can be determined using the sequenced *P. trichocarpa* genome. In *P. trichocarpa* it is thought that chromosome 19 may contain the sex locus (Yin et al., 2008), so investigating gender-biased gene-expression on this chromosome may yield insight into how gender is determined genetically in this species. Genes that are differentially expressed between the genders could be clustered around the sex locus because as suppression of recombination spreads along a chromosome from an initial mutation for female or male sterility (Nicolas et al., 2005), genes not directly involved in gender-determination could also begin to segregate with gender and cause slight phenotypic differences between the genders. Linkage of genes located at a sex locus is necessary in order to maintain separate genders (Charlesworth and Guttman, 1999), as without the suppression of

recombination of the female- or male-sterile mutations, the plant could revert to hermaphroditism (Ming et al., 2007).

The evolution of recombination suppression at the sex locus is thought to be driven by sexually antagonistic genes, whose expression is advantageous in male and deleterious in females, or vice versa (Charlesworth and Mank, 2010). Genetic sexual antagonism has probably had a role in the evolution of dioecy because without it there would be no selective pressure to convert hermaphrodites into males (Bergero and Charlesworth, 2009).

4.1.4 Investigating gender-biased gene-expression using genome wide survey approaches

Gender-determining genes could be identified by investigating genes that are preferentially expressed in males or females using microarrays, or whole transcriptome expression data from next-generation sequencing. The advent of global gene-expression profiling approaches, and the availability of whole genome microarrays for *Populus*, made it possible to use microarrays to look at differential gene-expression between male and female individuals. This approach could identify genes controlling the expression of floral development genes produce separate male and female plants. Of more importance is the identification of genes whose expression is gender-biased in all tissues of the plant as these will provide potential RNA sex-markers and be possible candidates for sex linkage.

The two main types of microarrays used in gene-expression experiments are cDNA arrays and oligonucleotide arrays. cDNA arrays are constructed using spotted cDNA clones that correspond to specific genes based on EST sequence information.

Oligonucleotide arrays are made using 16 to 20 pre-fabricated sequences 25-60 nucleotides long for each gene (Kafadar and Phang, 2003). Oligonucleotide microarrays, which allow examination of the expression of >35000 genes, are available for *P. trichocarpa* through NimbleGen Systems Inc. (USA), and Affymetrix microarrays for *P. trichocarpa* have been available since the fall of 2005. This study used NimbleGen oligonucleotide microarrays, which consisted of 24mer-70mer long nucleotide sequences, with a coverage of 6 to 20 probe sequences/gene, to determine the feasibility of microarray expression profiling to detect gender-specific gene-expression.

In the five years since the microarray experiments were conducted in this study there have been many advancements in the field of collecting genome wide gene-expression and transcriptome data. These include hybridization techniques such as the microarray experiments outlined above, as well as sequence-based approaches, which I will describe briefly here. Initially, Sanger sequencing of cDNA or EST libraries were used to identify genes that were being expressed in given different tissues of *P. trichocarpa* (Sterky et al., 2004), and other organisms. But Sanger sequencing of cDNA libraries is slow, expensive, and not usually quantitative. Tag-based methods such as SAGE, CAGE and MPSS were developed as high-throughput processes to overcome some of the limitations of Sanger sequencing and produce precise information about gene-expression levels (Wang et al., 2009), but these methods also result in only portions of transcripts being analyzed and paralogues of genes are generally indistinguishable from each other.

The development of high throughput, next-generation sequencing techniques such as those available from Illumina IG, Applied Biosystems SOLiD and Roche 454 Life

Science Systems, has allowed for the development of new methods for quantifying and mapping transcriptomes, which are the complete set of DNA transcripts contained in a given cell at a specific physiological or developmental stage (Wang et al., 2009). Currently this approach is being used to investigate SNP variation in *P. trichocarpa* to detect alternative splicing of genes and adaptive evolution in this species (Geraldes et al., 2011).

4.2 Objectives

The goal of this project was to investigate gender-biased gene-expression, and to identify genes that show an expression pattern in vegetative and floral tissues that segregates with gender. I hypothesize that the number of genes that are differentitally expressed between male and female reproductive structures in floral buds to be numerous due to the separate male (stamens) and female (ovules) organs present in the unisexual flowers. For this reason I focused on gene-expression in the leaf tissues to see if there are consistent gene-expression differences between male and female individuals in vegetative tissue. These leaf expressed genes may be the ones responsible for initiating gender-related gene-expression, or they may be linked to such genes.

The first objective was to look at genome wide gene-expression differences between male and female *P. trichocarpa* individuals, and see if any genes showed a statistically significant gene-expression difference between males and females. The second objective was to investigate any genes that showed gender-biased expression patterns, even if they were not statistically significant, with a larger sample size of female and male *P. trichocarpa* individuals using reverse-transcription (rt) PCR, to attempt to identify potential genomic sex-linked markers.

4.3 Methods

4.3.1 Collection of and preparation of biological materials

The plant tissue from floral buds and leaves of male and female *P. trichocarpa* trees used for the first microarray experiment was collected for the lab in May of 2004 from natural stands located near Chilliwack, B.C. Plant tissue from leaf buds of male and female *P. trichocarpa* trees was also collected in March of 2006 from natural stands located on the UBC campus and these samples were used for the second microarray experiment (Figure 4.1).

Collected tissues were stored at –80ºC until RNA was extracted from male and female leaf and floral buds following an RNA extraction protocol used for tree tissues (Kolosova et al., 2004), and samples were cleaned for use in cDNA preparation and microarray experiments with the QIAgen RNeasy MinElute Cleanup Kit (50) (Qiagen Inc., Toronto, Ont., Canada) or the Turbo DNA-free kit (Applied Biosystems/Ambion, Streetsville, Ont., Canada), and then stored at –80ºC to prevent degradation.

4.3.2 cDNA preparation and PCR conditions

cDNA was prepared using the RevertAid™ H minus First Strand cDNA Synthesis Kit (Fermentas Life Sciences, CA, USA). PCR primers for this project were designed using Primer3 (v. 0.4.0, http://frodo.wi.mit.edu/primer3/), and tested in silico to determine the probability of primer dimers or unspecific primer-binding with Amplify 3X (v.3.1.4, http://engels.genetics.wisc.edu/amplify). Primers (Appendix A) were ordered from Integrated DNA Technologies Inc (Coralville, Iowa). DNA amplification for this project was performed as described in chapter two, section 2.3.2 of this thesis. Amplification products were visually scored for the presence or absence of bands on 1%

agarose gels stained with GelRed™ Nucleic Acid Gel Stain, 10,000X in water (Biotium, Hayward, CA 94545). Amplified PCR products were then prepared for sequencing by Macrogen, in Maryland USA.

Figure 4.1 Collecting leaf and floral buds from *P. trichocarpa* individuals.
Trees were located on the University of British Columbia, Vancouver campus near the Triumf parking lot. Gregg Doughty, UBC arborist with plant operations, assisted in collecting.

4.3.3 Microarray gene-expression profiling experiment and statistical analysis of the microarray data

P. trichocarpa cDNA prepared from RNA extracted from male and female floral and leaf tissue was used to probe NimbleGen *P. trichocarpa* Affymetrix oligonucleotide microarray chips. The quality of the mRNA samples used in the microarray experiments was tested for degradation using the Agilent Bioanalyzer at the Centre for Molecular Medicine and Therapeutics (CMMT) facility, Vancouver, BC. Two microarray experiments were designed and these were sent to NimbleGen Systems Inc. (USA) to

perform the microarray hybridization experiments (Figure 4.2). After receiving the microarray expression data file from NimbleGen Systems, a simple Student's t- determined that the expression data from the floral tissues should be studied separately as the unisexual flowers contributed to almost 7000 genes being differentially expressed between the male and the female inflorescences, and these expression value differences in flowers reduced the signal from gender-biased gene-expression in leaf tissues samples to undetectable levels.

The data for the leaf mRNA from the two microarrays were normalized together and the male and females leaf expression were compared to see if there was any difference in gene-expression patterns due to gender. When the leaf samples form both microarrays were added together this resulted in leaf mRNA samples from a total of five female and five male individuals (n = 5 for each gender). Normalization of the data from the arrays was necessary due to experimental variations in sample treatment, dye labeling, efficiency and detection. As a result, the fluorescence intensities of microarrays cannot be compared directly, but must be calibrated or "normalized" (Huber et al., 2002). In this study the microarray expression data was normalized using both Variance Stabilization Normalization (VSN) (Huber et al., 2002) and Quantile Normalization (Bolstad et al., 2003). Two statistical methods were considered for analyzing the microarray data once it had been normalized. A mixed effects model was fit to the data which adjusted for variation within (two probes per gene on the same array) and between arrays along with fixed effects for gender and experiment. Also, a LIMMA software package (www.bioconductor.org), designed to analyze experiments involving

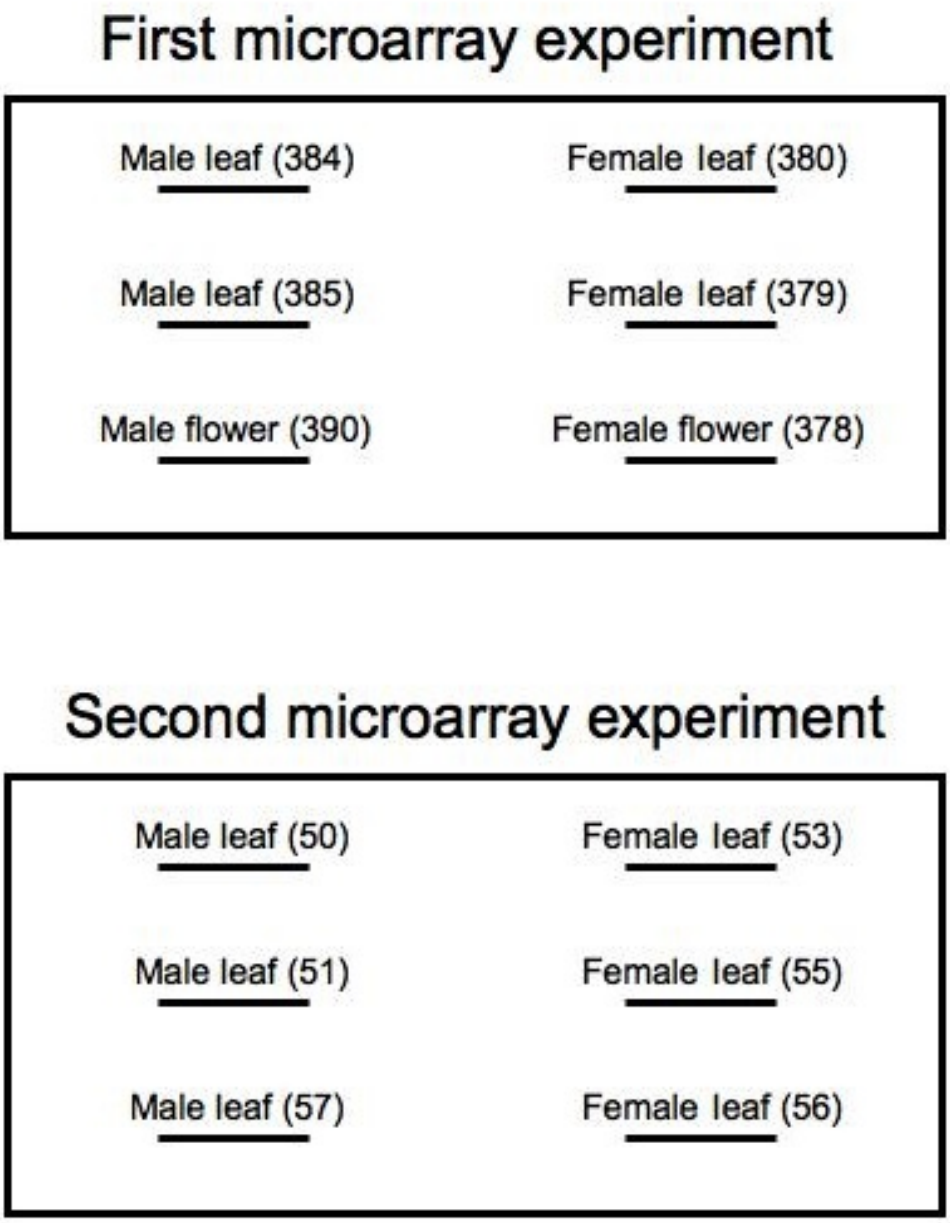

Figure 4.2 Arrangement of *P. trichocarpa* mRNA samples on the NimbleGen Affymatrix oligonucleotide microarray chips.
Numbers in brackets indicate the collection number of the individual tree the MRNA was prepared from.

comparisons of expression data from two colour microarrays, or log-intensity values from one channel Affymetrix arrays (Smyth, 2005) were used to analyze the normalized microarray data. The LIMMA analysis was conducted with an empirical Bayes modified error to compute estimates of differential expression for gender and experiment. After looking at the results from both normalization and stasistical analysis techniques, the Quantile Normalization, followed by the LIMMA analysis was chosen as this analysis identified the most likely candidate genes as having differentially expressed values between male and female individuals.

4.3.4 Investigation of gender-biased gene-expression patterns on chromosome 19

Chromosome 19 is thought to be involved in gender-determination in *P. trichocarpa* (Yin et al., 2008), and given that clustering of genes that are differentially expressed between the genders could occur around the sex locus due to recombination suppression spreading along a chromosome from an initial female or male sterility mutation (Nicolas et al., 2005), gender-biased gene-expression in male and female leaf tissues was investigated by looking at the microarray expression data for genes located on chromosome 19. The log of the ratio of male to female gene-expression values was plotted against the position of genes associated with microarray probe expression values along chromosome 19. Negative log expression values indicated a female-biased gene-expression, and positive log expression values indicated male-biased gene-expression. Genes showing either male- or female-biased expression values were identified and investigated in a larger sample of nine male and thirteen female individuals using reverse-transcription PCR. With the release of version 2.0 of the poplar genome, position and annotation data for the gene content of chromosome 19 was updated from version 1.1

to version 2.0 by performing BLAST searches of gene DNA sequences from version 1.1 against the version 2.0 genome sequence. Reverse-transcription PCR primers were designed based on version 2.0 of the poplar genome.

4.4 Results

4.4.1 Results from the statistical analysis of the *P. trichocarpa* leaf tissue data from both microarray experiments

When the data from both microarray experiments was pooled the LIMMA statistical analysis of the data indicated that 1.17% of genes included on the microarrays were differentially expressed with respect to gender (Figure 4.3). This is a surprisingly small number as some 9.88% of genes are differentially expressed with respect to the different experiments (Figure 4.4). These figures are derived from the distribution of the parametric p-values, calculated by the LIMMA analysis. Given an experiment of this size, the biological noise clearly masks the effects (if any) of gender.

There were two genes that showed a statistically significant difference in gene-expression between male and female *P. trichocarpa* individuals from the microarray experiments (Table 4.1). Reverse-transcription PCR was going to be used to verify if these two genes showed a gender-bias in their expression in a larger number of individuals. The resulting sequences obtained for these two genes were then going to be investigated to see if the sequences from difference individuals were exactly the same, or if they contained SNPs that differed between individuals or genders. By investigating SNP differentiation it may have been possible to determine why there is an expression difference between males and female using the sequence data collected.

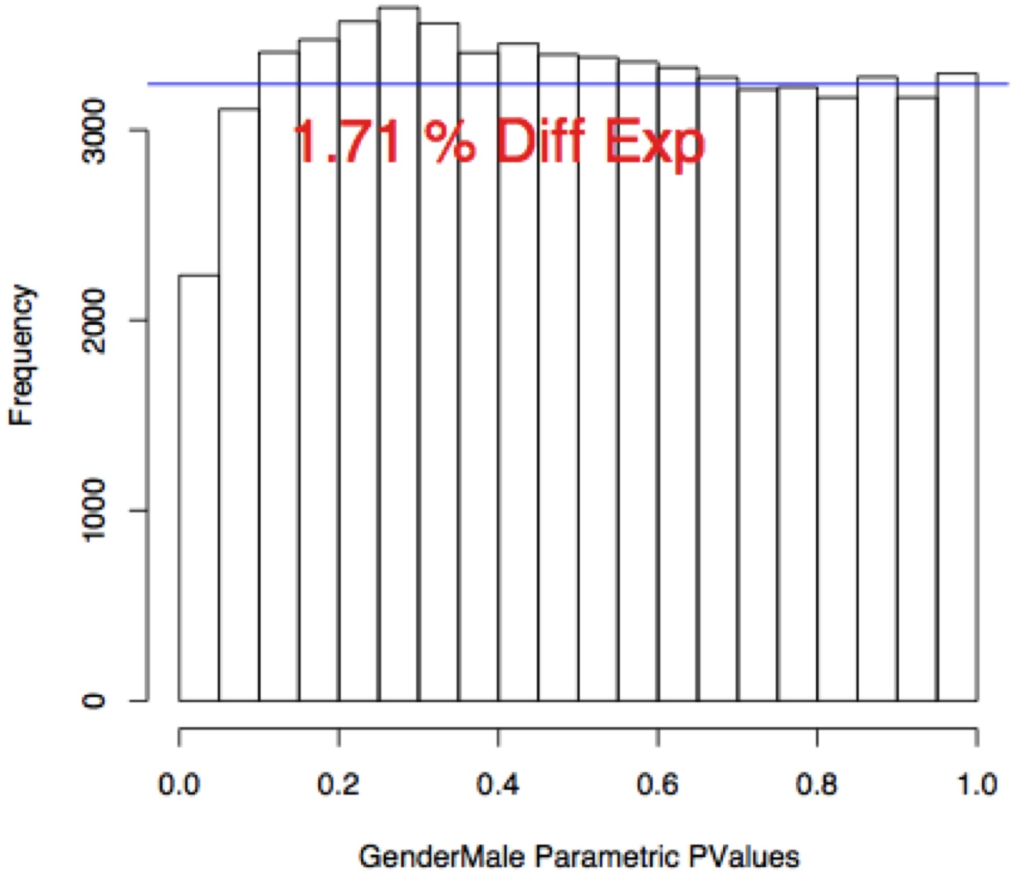

Figure 4.3 Histogram of the parametric p-values plotting gene-expression in male leaf tissue (x axis) against the frequency of gene-expression in female leaf tissue (y-axis). The blue line indicates the threshold above which are the percentage of genes showing differential gene-expression between the genders, N = 5 for both genders.

However, with the release of the second version of the poplar genome, the sequences of these two genes were BLASTed against the new genome and it was discovered that both genes have multiple hits. eugene3.00660277 BLASTs to two genes on scaffold 3 (POPTR_0003s03450 and POPTR_0003s03440), and to intergenetic regions on scaffold 43, 1184 and 19. fgenesh4_pg.C_scaffold_277000004 blasts to four genes on scaffold 17 (POPTR_0017s13140, POPTR_0017s13180, POPTR_0017s13220,

POPTR_0017s13260) (Appendix B). Given that these two genes did not appear to be single copy genes on version 2.0 of the genome, investigating their expression pattern in a greater number of male and female individuals would be problematic as it would be difficult to design primers to amplify specific copies of the gene, and it is unlikely that such a gene would be involved in gender-determination. Therefore work with these two genes was discontinued in favor of focusing research (reported in chapter three of this thesis) on genes located on chromosome 19, the chromosome indicated to contain the sex locus (Yin et al., 2008).

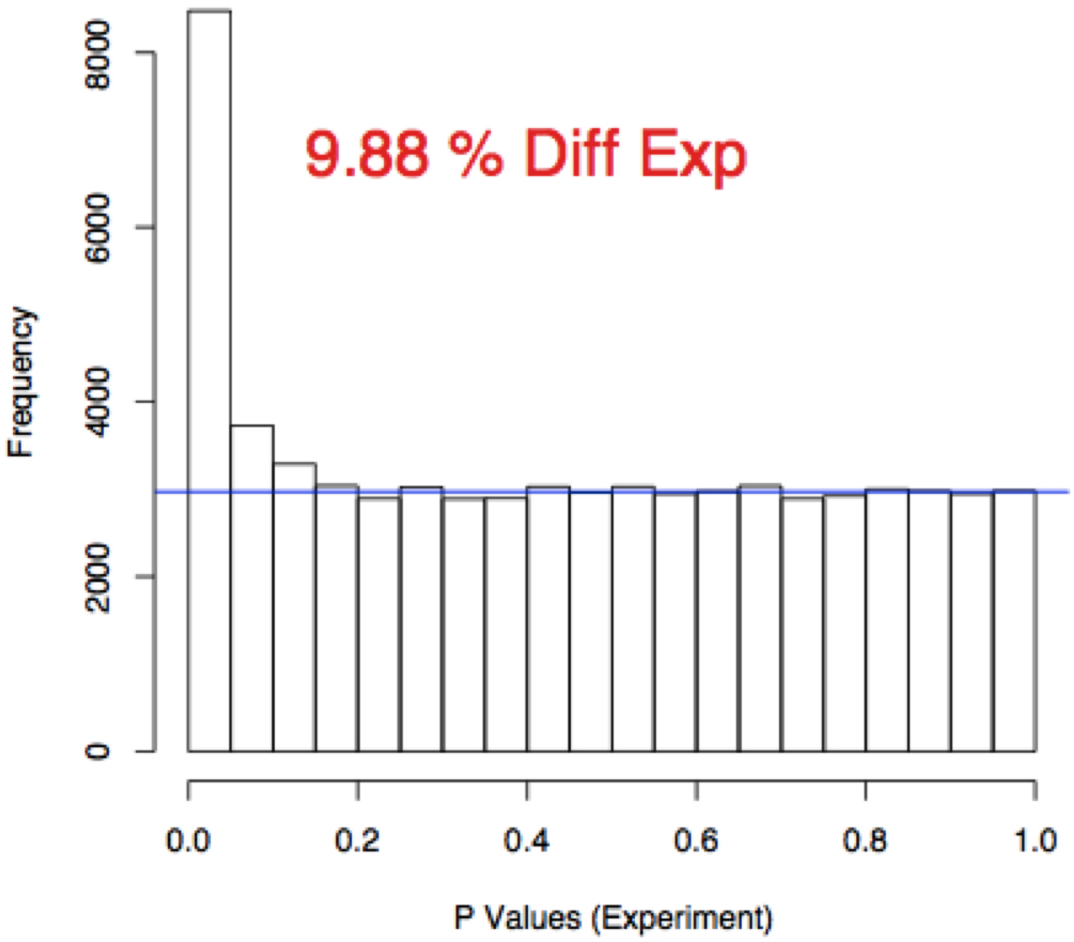

Figure 4.4 Histogram of the parametric p-values comparing gene-expression differences in leaf tissue between the two microarray experiments.
The blue line indicates the threshold above which are the percentage of genes showing differential gene-expression between the two microarray experiments, N = 5 for both male and female mRNA samples.

Table 4.1 Genes showing a statistically significant difference in the gene-expression between males and females in *P. trichocarpa*, on version 1.1 of the genome.

Gene Name	Q-value	Location on poplar genome, V. 1.0
eugene3.00660277	0.2460120	Poptr1_1/scaffold_66:1897328-1898141
fgenesh4_pg.C_scaffold_277000004	0.2460120	Poptr1_1/scaffold_277:17700-18283

4.4.2 Investigation of gender-biased gene-expression patterns on chromosome 19

Four genes that showed a gender-biased expression pattern in my microarray results from working with version 1.1 of the poplar genome translated to genes that were also located on scaffold 19 on version 2.0 of the genome when a BLAST search was performed to update the information from version 1.1 to version 2.0 of the genome (Figure 4.5). Two of the genes, POPTR_0019s03090 (SCA_19_9) and POPTR_0019s03100 (SCA_19_10), highlighted in orange (Figure 4.4), are right next to each other, and fairly close to the region of chromosome 19 where scaffold 117 was incorporated. These two genes show female biased expression, as well as a third gene, located at the other end of the chromosome, POPTR_0019s12360 (SCA_19_11), highlighted in red (Figure 4.4).

POPTR_0019s13010 (SCA_19_7) was the only gene located on chromosome 19 that the statistical analysis of the microarray experiment indicated showed a nearly significant difference in expression values between males and females. It showed a male-biased expression pattern, and is highlighted in blue on Figure 4.5. The gene model from version 1.1 of the genome that showed the gender-biased expression blasted to two gene models on version 2.0 of the genome: POPTR_0019s13010 and POPTR_0019s13020. Primers were designed for both genes, but gene specific primers were only successfully designed for POPTR_0019s13010, so this gene was chosen to verify the microarray results in a larger sample of male and female leaf samples. Information on the position and putative gene functions for all the genes that were worked with is listed in Appendix B.

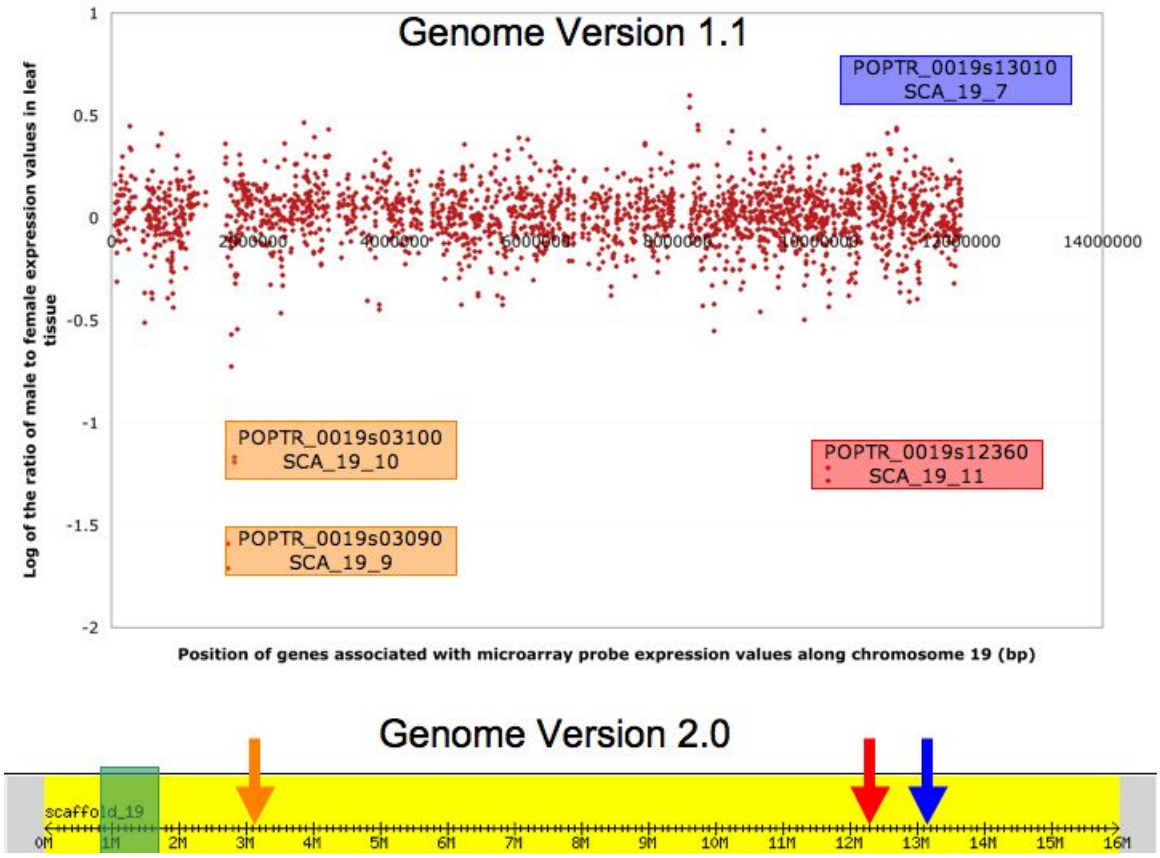

Figure 4.5 Comparison of the gene-expression values from the microarray data for chromosome 19, based on version 1.1 of the poplar genome, to the physical map chromosome 19 from version 2.0 of the poplar genome.
On the y-axis negative values indicate female-biased gene-expression, and positive values indicate male-biased expression. The x-axis indicates the distance along chromosome 19 on both versions of the genome. The green box indicates the region containing the genes sampled on telomeric region of chromosome 19 during the investigation of SNP differences in the research outlined in chapter three of this thesis. Coloured boxes (orange, red and blue) indicate genes showing the most gender-biased expression that were located on chromosome 19, with arrows in corresponding colours indicating their approximate positions on version 2.0 of the chromosome.

After designing specific primers for all four genes located in chromosome 19 that showed gender-biased expression pattern from the microarray experiments, reverse-transcription PCR was performed to confirm this expression pattern in a larger sample of nine male and thirteen female *P. trichocarpa*. All successful PCR products obtained

were sent for sequencing to confirm that expression data for the target genes had been obtained.

For genes POPTR_0019s03090 (SCA_19_9) and POPTR_0019s03100 (SCA_19_10) PCR amplification of bands was successful in a few individuals, both males and females for both genes, and it was confirmed from the good quality sequence obtained from the samples that the target sequence for each gene was being amplified. So it appears that both these genes are expressed in leaf tissue, as indicated by the microarray data, but with no discernable gender-bias. For the gene POPTR_0019s13010 (SCA_19_7) amplified bands were obtained, but it appears that the primers were not specific enough as the sequence obtained was not good quality, and didn't match up well with the reference sequence from version 2.0 of the poplar genome. Also, when the best quality sequence produced for one female sample was BLASTed against version 2.0 of the poplar genome, the best hit was to gene POPTR_00190309, not the target sequence. Amplification products were not obtained using the primers designed for POPTR_0019s12360 (SCA_19_11).

Redesigning primers for SCA_19_7 and SCA_19_11, and testing the primers under various PCR conditions also failed to yield better PCR results for these two genes. The cDNA used in these experiments was tested with a constituently expressed gene, POPTR_0006s19870, an elongation factor the Cronk laboratory uses as a positive control for PCR work with *P. trichocarpa*, and consistent amplification was observed in all samples with this gene, so I am confident that the lack of amplification for the target genes was not due to poor cDNA quality, but due to either ineffective primer binding or that the target gene was not expressed in the tissue sampled.

4.5 Discussion

4.5.1 Sample size on statistical power of microarray experiments

Initially, when flower tissues were included in the microarray analysis a strong gender-bias in gene-expression was detected between males and females, that corresponded to almost 7000 genes showing a positive gender-bias in their expression pattern. This result was expected because given that flowers are unisexual in the dioecious *P. trichocarpa* (Cronk, 2005), and the organs specific to male flowers, and therefore the genes involved in their development, would be more highly expressed in the male flowers than the females flowers, and vice versa. Of more interest from the point of view of developing a genetic sex-linked marker and investigating how gender is controlled genetically in this species would be genes that showed a gender-biased gene-expression pattern in vegetative tissue such as leaves, because ostensibly both males and females would have to express genes involved in determining or regulating gender, though they may be expressed at different levels in males or females. Therefore this investigation was focused on leaf tissue, and the leaf mRNA samples from the two microarray experiments were normalized together to increase the statistical power of this experiment.

When the LIMMA statistical analysis was performed on the expression data for leaf tissue from the two microarrays, it was to shown that 1.17% of the genes included on the microarrays were differentially expressed with respect to gender (Figure 4.3). Given that the microarray experiment conducted consisted of only two arrays, with 5 female and 5 male *P. trichocarpa* leaf samples, it appears that biological variation masked any differential expression between females and males. If there is gender-biased gene-

expression, the effect is very small and a much larger sample size would be necessary to detect this effect. By investigating genes that showed even non-statistically significant gender-biased expression patterns in a larger sample size (13 females and 9 males), if there was differential gene-expression between the genders it should have been detectible. Genes located on chromosome 19 were focused on (Figure 4.5), as research indicates that this chromosome may contain the sex locus in the genus *Populus* (Pakull et al., 2011). Unfortunately it proved difficult to amplify the four genes indicated as having a gender-bias by the microarray experiment analysis. While this research was able to confirm that two of the genes (POPTR_0019s3090/SCA_19_9 and POPTR_0019s3100/SCA_19_10) were expressed in leaf tissue in a larger sample size, these results did not indicate a gender-bias in that expression pattern. When the DNA sequence obtained for these genes was investigated there were no gender-biased SNPs present. Gender effects may be present but a larger experiment will be needed to detect this with any statistical confidence, as the gender effects are clearly subtle in vegetative tissues in *P. trichocarpa*.

4.5.2 Effect of biological variation on interpreting microarray results

One of the reasons I was unable to detect gene-expression differences between the genders was due to the high amounts of biological variation. There is a ~10% difference in gene-expression levels between the two microarray experiments conducted, and the LIMMA statistical analysis indicates that these are real differences, not due to chance (Figure 4.4). The difference in gene-expression levels between the two microarray experiments may have been due to a number of factors. The *P. trichocarpa* individuals

that were sampled in the two experiments came from two different populations, located approximately 100km apart.

The male and female individuals sampled for the first microarray experiment came from groves of *P. trichocarpa* growing in Chilliwack, B.C., and the individuals from both genders sampled for the second microarray experiment were located on the University of British Columbia endowment lands, in Vancouver B.C. While species in the genus *Populus* are all wind pollinated and one would expect a large effective population size (Hamrick et al., 1992), studies of SNP variation in *P. balsamifera* (Keller et al., 2010), and *P. trichocarpa* (Gilchrist et al., 2006), have shown that populations within species in this genus are adapted to local environmental conditions. This ability to adapt could result in variations in gene-expression levels between different stands of trees, even those within 100km of each other. Population substructure arises in *P. trichocarpa* because of dispersal patterns of seed and pollen as well. In fact, using microsatellite markers it has been shown that population substructure can be detected even when sampling populations of *P. trichocarpa* separated by a few hundred meters (Slavov et al., 2010).

Another reason why there may have been a lot of biological variation between the two microarray experiments is that the samples collected from the different populations were collected at different times of year, and in different years. While the samples collected from the Chilliwack individuals were collected from young leaves in May of 2004, the samples UBC individuals were collected in March of 2006 from leaf buds just beginning to open. This means that the samples used for the microarray experiments were at different developmental stages and therefore gene-expression profiles could be

different in the leaf tissue used in the two experiments. Also, clonal field trials in *Populus* species have shown that the timing of spring bud flush is a highly heritable trait, with up to 98% of the variability observed in this trait being under genetic control, with environment influencing only 2% of the variability observed (Frewen et al., 2000).

It appears from the results presented in this study that sampling from two different stands of trees, at different times of the year, and in different years, introduced an amount of biological variation into this experiment that was too large to effectively measure any gene-expression differences between leaf tissue in male and female individuals. With the amount of variation already introduced by using microarrays to look at gene-expression (variations in dye efficiency and detection, or fluorescence intensities, to name a few), it is important to make sure that samples are closely related biologically and environmentally so as to ensure that the gene-expression levels does not vary between samples greatly (Huber et al., 2002). This is much more possible to do when working with model species like *Arabidopsis thaliana*, where inbred lines can be quickly produced and maintained to reduce the amount of biological variation present in the sample individuals prior to conducting a microarray experiment. One such experiment very effectively compared varying amounts of gene-expression variation across seven *Arabidopsis* genotypes by making reciprocal crosses among the inbred lines (Vuylsteke et al., 2005). Unfortunately, controlling for biological variation in this manner is more difficult in *P. trichocarpa* due to its perennial life history, and the fact that sexing individuals is not possible until the individuals flower at between five and ten years of age (Brunner, 2010).

4.5.3 Interpreting gender-biased gene-expression using genome wide survey approaches

Unfortunately, with the release of version 2.0 of the poplar genome, it appears that some of the genes of interest were annotated differently between the two versions, making it problematic to investigate gene function or expression patterns of the genes indicated by the microarray experiment to have had a gender-biased expression in leaf tissues. This is a common problem when working with model organisms that have rapidly developing genomic resources, as gene structural and functional annotations can change as new data sets and more accurate information becomes available (van den Berg et al., 2010). Both version 1.1 and version 2.0 of the poplar genome used Fgenesh+ (http://www.softberry.com/berry.phtml) to predict gene models on the sequenced *P. trichocarpa* genome, but version 1.1 also used three other gene model prediction methods (Fgenesh, EuGene and GrailExp6) (Tuskan et al., 2006), whereas version 2.0 of the genome only used one other gene model predicting program, GenomeScan, if Fgenesh+ did not predict a model at a given locus (http://www.phytozome.net/poplar.php).

Given the differences in how gene models were predicted between the two versions of the genome, it is not surprising that when I searched for the sequences of the two gene models indicated as having a statistically significant difference in gene-expression between males and females, eugene3.00660277 and fgenesh4_pg.C_scafford_277000004, on version 2.0 of the genome, these two gene models were now incorporated into different positions, and also had hits to multiple new gene models, making it challenging to further investigate their expression pattern in leaf tissue, gender-biased or otherwise. Genes involved in gender-differentiation are likely to

be located at the sex locus, and therefore genes with paralogues in multiple places in the genome are unlikely to be involved in maintaining separate genders. This is because current theories on the evolution of sex chromosomes seem to indicate that the evolution of a sex locus begins with a male or female sterile mutation around which genetic recombination becomes suppressed (Ming and Moore, 2007), and it is unlikely that a mutation of this kind would occur in multiple copies of the same gene, spread out on many chromosomes of the genome.

4.5.4 Considerations when working with the developing model system of *Populus trichocarpa*

When this study into gender-determination in *P. trichocarpa* was begun, version 1.1 of the poplar genome had just been released (Tuskan et al., 2006), so the microarray experiments were performed using gene annotation based on the first genome assembly for *P. trichocarpa*. While this first assembly provided a very important tool for developing poplar as a model species for genetic research, once work began with the genomic data, it became apparent just how much data needed to be added to the sequenced genome before it became as comprehensive a source of information as the annotated genomes for other model species such as *Arabidopsis thaliana* and *Drosophila melanogaster*.

Most of the functional genomic research done in *P. trichocarpa*, and other *Populus* species, has focused on genes involved in wood development, and plant anti-pathogen biochemical pathways, as these characteristics are important to the development of poplars as sources of bio-fuel and pulp and paper (Jansson and Douglas, 2007). With version 1.1 of the poplar genome approximately 63.5% of the gene models had at least

some functional data or structural domain annotation (Tuskan et al., 2006), based on either functional studies in *P. trichocarpa*, or sequence homology with genes of known functions in *Arabidopsis thaliana*, or other model plant genomes. Because I was investigating gene-expression based on any detectible gender-biased expression pattern, many of the genes that were of interest in this study were either on genetic scaffolds that were not incorporated into the 19 chromosomes of *P. trichocarpa* (Table 4.1), or had no known function. This is not surprising, as though *P. trichocarpa* has great potential as a model species for plant genome evolution, genetic variation in response to ecological changes, and gender-determination studies (Cronk, 2005), until recently functional studies in *P. trichocarpa* have been few.

The choice to work with the NimbleGen *P. trichocarpa* Affymetrix oligonucleotide microarray chips was based on these being a good option that was available to work on at the time, given that the resources to construct custom microarray chips were not available "in house". However, one draw back to working with commercial microarrays is that they are often annotated prior to being made available for public use, so the structural and functional annotations for the genes they contain may not be the most up to date (van den Berg et al., 2010).

The release of version 2.0 of the poplar genome in 2009 made it necessary for the functional and structural annotations of the genes of interest in this study to be updated, as it was important to work with the most current information available on the *P. trichocarpa* genome in order to be able to draw any valid conclusions from this study. This is a common occurrence for researchers working with newly available genome sequences as updates to the data set happen at a rapid pace and it is a challenge to

correctly interpret an experiment-based data set (such as the one reported here) with developing gene functional annotations (van den Berg et al., 2010).

The rapid pace of development of genetic tools for genome sequencing frequently means that interpreting experimental results is challenging, and requires a lot of data management to ensure that the most current functional and structural gene data is being referenced. Gene re-annotation is becoming an important step in analyzing functional genomic data to improve its quality and quantity (van den Berg et al., 2010). With the availability and cost effectiveness of high throughput, next generation sequencing technologies, the ability to update gene annotations ensures that researchers are deriving conclusions about gene function based on the most up-to-date information, and are developing computational technologies to greatly streamline this process.

4.6 Conclusions

I found that sample size is critical to conducting this kind of genome wide gene-expression experiment as the results reported here showed that there was a greater gene-expression difference between the two microarray experiments performed than between genders. Gender-biased gene-expression may be present in vegetative tissues of *P. trichocarpa*, but these effects are subtle, and therefore a much larger sample size would be needed for a microarray experiment to be able to detect it with any statistical confidence.

I was able to identify differential gene-expression patterns between male and female leaf tissues with the microarray experiments, however the investigation of the genes indicated by the microarray data to have differential expression between males and females using reverse-transcription PCR was unable to confirm the microarray results.

As a result of this I was unable to identify potential genomic sex-linked markers in *P. trichocarpa*.

I detected a large amount of biological variation with the two microarray experiments. This result indicates that there is quite a lot of genetic variation even between two populations that are less than 100km apart. Although the samples used in the microarray experiments were all from young leaves, they were collected in different years, and it was possible to detect differences in gene-expression due to the slight differences in the developmental stage at which the samples are taken. Work with microsatellite markers has shown that population substructure can be detected in populations of *P. trichocarpa* even when sampling from stands of trees within a few hundred meters of each other (Slavov et al., 2010), and this work seems to support this as well.

Another conclusion I can draw from this study is that working with a rapidly developing model system such as *P.trichocarpa* requires that one revisit data and update it frequently to reflect the latest and most accurate gene annotations available. The release of version 2.0 of the poplar genome made much more information available on the genes I was working with, which improved my project greatly, as I was able to more accurately place genes showing gender-biased expression on chromosome 19, the putative sex chromosome in *P. trichocarpa*.

The advances in cost effective, high throughput next generation sequencing technologies that are currently available from Illumina IG, Applied Biosystems SoLiD and Roche 454 Life Sciences Systems (Wang et al., 2009) will make it much easier to generate and manage larger genetic data sets, and to keep the gene annotations up to date.

Data sets of the kind would make it possible to detect with greater accuracy the subtle gender-biased gene-expression differences in vegetative tissues in *P. trichocarpa* that my experiments may indicated exist.

Chapter 5: MADS-box genes in the genus *Populus* and their role in floral development

5.1 Introduction

One of the major discoveries in the field of developmental evolution was that certain gene families control the development of whole organs (Meagher, 2007). In the case of flowers, the MADS-box gene family has been found to be involved in the designation of floral organ identity and development. MADS-box genes that have been identified as having a role in plant organ development encode proteins that have a typical four domain MIKC structure (Alvarez-Buylla et al., 2000). The gene family derives its name from the highly conserved 55 amino acid DNA binding domain called the MADS domain after the genes that it was first discovered in *MINICHROMOSOME MAINTENANCE 1* (*MCM1*), *AGAMOUS* (*AG*), *DEFICIENS* (*DEF*), and *SERUM RESPONSE FACTOR* (*SRF*) (Gramzow and Theißen, 2010). The second most conserved protein domain in this gene family is the keratin-like (K) domain, which is generally 70 amino acids in length with regularly spaced hydrophobic amino acids so it takes on a coiled-coil structure (Alvarez-Buylla et al., 2000). The Intervening (I) domain links the MADS and K domains, and along with the K domain is critical for mediating protein dimerization and specificity of interactions between MIKC-type proteins (Stellari et al., 2004). The fourth protein domain that is part of the MIKC structure is the poorly conserved carboxyl-terminal or C domain which functions as a trans-activation domain (Alvarez-Buylla et al., 2000).

5.1.1 Evolution of MADS-box genes in plants

Given that MADS box genes play a central role in floral development they can provide insight into how flowers have evolved (Meagher, 2007). From phylogenetic analyses it appears that an ancestral MADS-box protein domain duplicated in the common ancestor of the main eukaryotic kingdoms approximately one billion years ago, giving rise to the Type I and Type II MADS-box gene lineages found in animals, fungi and plants today (Alvarez-Buylla et al., 2000). MADS-box genes in plants that have characterized functions are predominantly Type II, and include a K domain, indicating that this plant-specific domain evolved early in the lineage (Alvarez-Buylla et al., 2000). One MADS-box gene has been found in an extant green alga (Gramzow and Theißen, 2010), and research in the moss *Physcomitrella patens* (Henschel et al., 2002) and the fern *Ceratopteris* sp. (Münster et al., 1997) indicate that this gene family is present in all land plant groups, and that the MIKC protein domain structure evolved and has been diversifying in function for over 650 million years (Nam et al., 2003).

5.1.2 MADS-box genes and their role in floral development

The MADS-box gene family has generally been studied in hermaphroditic flower development, by creating homeotic mutants that produce floral organs in inappropriate whorls of the flowers (Lebel-Hardenack and Grant, 1997). Analysis of floral mutants in *Antirrhinum* and *Arabidopsis* led to the formation of the ABC model for floral organ development (Coen and Meyerowitz, 1991; Zik and Irish, 2003). This model defines three regions of the floral meristem, each of which is controlled by a class of genes – A, B, or C (Zik and Irish, 2003). Region A consists of the first and second whorls of the floral meristem, region B comprises the second and third whorls and region C contains

the third and fourth whorls (Coen and Meyerowitz, 1991). Expression of A-class genes such as *APETALA1* (*AP1*) and *APETALA2* (*AP2*) alone specify sepal identity, A + B-class genes specify petal development in the second whorl, B + C-class gene expression results in stamen formation in the third whorl and C-class genes alone specify carpel development in the fourth whorl (Zik and Irish, 2003).

B-class genes include *APETALA3* (*AP3*) and *PISTILLATA* (*PI*) in *Arabidopsis*, and the primary C-class gene is *AGAMOUS* (*AG*) (Kramer et al., 2004). Further research has indicated that two other classes of genes are also involved in the floral development pathway. D-class genes such as *floral binding protein 71* (*FBP71*) have been found to be crucial for the determination of ovule identity in *Petunia* (Colombo et al., 1995), and E-class genes, known as *SEPALLATA1*, *2*, and *3* (*SEP/1/2/3*) which provide further levels of developmental control, are also required for specifying stamen, petal and carpel identity (Pelaz et al., 2000). Further research has indicated that D-function genes are not important in floral development outside of *Petunia*, so the current model for floral development that is generally applicable is the ABCE model (Theißen, 2001).

Research done in *Arabidopsis thaliana* has been instrumental in identifying the role of MADS-box genes in floral development, and the DNA sequences from MADS-box genes in this model system have been used to identify genes with homologous functions and expression patterns in other species of flowering plants (Meagher, 2007), including *Populus* species.

5.1.3 MADS-box gene research in the genus *Populus*.

The molecular genetics of floral development in *Populus* has been studied in a few species, namely *P. deltoides* (Zhang et al., 2009), *P. tremuloides* and *P. tremula*

(Cseke et al., 2005), and *P. trichocarpa* (Brunner and Nilsson, 2004b). *Populus* spp. flowers are unisexual in all but a few cases, and consist of a very reduced perianth cup (Fisher, 1928), that appears as a single ridge encircling the male or female floral organs in some species like *P. deltoides*, or as a true lobate perianth in other *Populus* species (Kaul, 1995).

Trees in the genus *Populus* have a life span of 100 to 200 years, and a long juvenile period, lasting seven to ten years, before the trees mature and begin flowering (Hsu et al., 2006). The transition from this juvenile phase to reproductive competency appears to have a genetic component given the strong developmental pattern of reproductive shoot formation (Yuceer et al., 2003). Studies in *P. trichocarpa*, *P. deltoides* and a *P. tremula* x *P. tremuloides* hybrid have indicated that the cycles of reproductive and vegetative growth in these tree species is coordinated by two genes, *FLOWERING LOCUS T1* (*FT1*) and *FLOWERING LOCUS T2* (*FT2*), which are homologous to the *FT* gene in *Arabidopsis* (Hsu et al., 2011). Over expression of both these genes seems to be associated with rapid shoot to flower conversion in *P. deltoides* (Hsu et al., 2006), and in *P. tremula* x *P. tremuloides* (Böhlenius et al., 2006).

When it comes to floral development, research in *A. thaliana* has shown that the B-class gene *AP1* is a direct target of *FT*, and *AP3* is down-stream from both *FT* and *AP1* in the development of petal and stamen identity (Hsu et al., 2006). Just as two homologs for the *Arabidopsis* gene *FT* have been found in *Populus* species. Two homologs for the *Arabidopsis* floral development genes *AP1*, *PI*, *AP3*, *AG* and *SEEDSTICK* (*STK*) have been identified in *P. trichocarpa* (Leseberg et al., 2006). The *Populus AP1* homologs *PTAP1-1* and *PTAP1-2* are both expressed in the initial stages of developing floral

meristems (Brunner, 2010), but to date the gene-expression patterns of the *P. trichocarpa PI* homologs have not been investigated in detail. The *P. trichocarpa* MADS-box gene *PTD* shows homology with *AP3*, and shows an expression pattern similar to that found for *AP3* and *DEF*, with no detectible expression in vegetative tissue, but spatial and temporal sex-specific expression in the inner whorl of male and female floral meristems (Sheppard et al., 2000). The two *AG* homologs found in *P. trichocarpa*, *PTAG1* and *PTAG2*, show an *AG*-like floral tissue expression pattern, and phylogenetic analysis strongly supports their evolutionary orthology to C-class MADS-box genes (Brunner et al., 2000). Finally, homologs of E-class genes, which are involved in the developmental control of the inner whorls of flowers, have been shown to have similar expression to the *A. thaliana* gene SEPALLATA (SEP), in *P. tremuloides* and *P. deltoides*. In *P. tremuloides, PTM3/4* and *PTM6* genes are expressed at all stages of male and female flower development (Cseke et al., 2005), and *PdMADS2* is expressed in the perianth cup of male inflorescences, and in the ovaries of female inflorescences in *P. deltoides* (Zhang et al., 2009).

5.2 Objectives

I expected to find gender-biased expression patterns in genes involved in floral development in *P. trichocarpa*, as flowers in this species are unisexual. Therefore, it was essential to collect the functional information available for genes involved in floral development in the *Populus* genus, and compare it to the expression data I generated from the microarray experiment (chapter four). I was particularly interested in B and C-class MADS-box genes as these genes play a role in androecium and gynoecium

development (Meagher, 2007), and therefore in dioecious species such as *P. trichocarpa* these genes are good candidates for the separation of male and female functions.

The first objective for this project was to create a comprehensive list of all *Populus* MADS-box genes that included the current annotation data from version 2.0 of the poplar genome, and the functional characterization data from expression studies. Secondly, I wanted to identify which of the poplar MADS-box genes would show gender-biased expression using male and female floral tissue on my microarray experiment. Thirdly, I wanted to relate my data on gender-biased expression pattern of floral MADS-box genes in poplar to the known function and expression patterns of homologous *A. thaliana* floral MADS-box genes. The fourth objective was to confirm the gene-expression bias observed in the microarray experiment in a larger sample size of male and female flower tissue using reverse-transcription PCR (rtPCR), and to add to the expression data available for MADS box genes in *P. trichocarpa* by focusing my work on genes that had little or no expression data available for them.

5.3 Methods

5.3.1 Collection of and preparation of biological materials

The plant tissue from floral buds and leaves of male and female *P. trichocarpa* trees used for the first microarray experiment was collected in May of 2004 from natural stands located near Chilliwack, B.C. Additional flower buds were collected from three male and four female *P. trichocarpa* trees in March of 2006 from natural stands located on the UBC campus previously identified as males or females by myself and Dr. Cronk for cDNA preparation. These tissues were stored at –80ºC for future use. RNA was extracted from male and female floral buds following either an RNA extraction protocol

developed for tree tissues (Kolosova et al., 2004), or a modified protocol using Plant RNA Purification Reagent from Invitrogen (Catalog # 12322-012). Samples were cleaned for use in cDNA preparation and microarray experiments with the QIAgen RNeasy MinElute Cleanup Kit (50) (Qiagen Inc., Toronto, Ont., Canada) or the Turbo DNA-free kit (Applied Biosystems/Ambion, Streetsville, Ont., Canada), and then stored at –80ºC to prevent degradation.

5.3.2 cDNA preparation and PCR conditions

cDNA preparation was done using the RevertAid™ H minus First Strand cDNA Synthesis Kit (Fermentas Life Sciences, CA, USA). PCR primers for this project were designed using Primer3 (v. 0.4.0, http://frodo.wi.mit.edu/primer3/), and tested *in silico* to determine the probability of primer dimers or unspecific primer binding with Amplify 3X (v.3.1.4, http://engels.genetics.wisc.edu/amplify). Primers used in this project were ordered from Integrated DNA Technologies Inc (Coralville, Iowa) (Appendix A). DNA amplification for this project was performed as described in chapter two, section 2.3.2 of this thesis, except that for this project 50ng of DNA was used/reaction. Amplification products were visually scored for the presence or absence of bands on 1% agarose gels stained with GelRed™ Nucleic Acid Gel Stain, 10,000X in water (Biotium, Hayward, CA 94545). Amplified PCR products were sent for sequencing by Macrogen, in Maryland USA.

5.3.3 Microarray gene-expression profiling experiment

In the fall of 2005 *P. trichocarpa* cDNA prepared from RNA extracted from male and female floral and leaf tissue was used to probe NimbleGen *P. trichocarpa* Affymetrix oligonucleotide microarray chips. The extracted mRNA samples were verified to be of

suitable quality for working with microarrays by testing them for degradation using the Agilent Bioanalyzer at the CMMT facility, Vancouver, BC. Two male and two female leaf mRNA samples were sent, as well as one male and one female flower sample (6 samples total) to NimbleGen Systems Inc. (USA) to perform the microarray hybridization. From the data file the floral tissue hybridization data was separated and it was observed that many genes showed a gender-bias, particularly in the male-biased direction, most probably because these genes are involved in pollen production. Gender-biased expression was calculated by taking the log intensity value of the female floral expression for a given gene, and subtracting it from the log intensity value of the male floral expression for the same gene. To narrow the study to genes likely to be directly involved in floral development, the top 15 genes were identified that 1) were annotated as having MADS-box functions and that 2) showed the greatest differences in gender-biased expression values between the male and female flower tissue samples, with log intensity values above a threshold of 0.05 for male inflorescences, and below -0.05 for female inflorescences.

5.3.4 Updating information on MADS-box genes in *P. trichocarpa* from version 1.1 to version 2.0 of the poplar genome

The first comprehensive study of MADS-box genes in *P. trichocarpa* was that of Leseberg et al. (2006), who used an *in silico* approach to identify 105 putatively functional MADS-box genes in version 1.1 of the poplar genome. With the release of version 2.0 of the poplar genome, in order to correctly identify the functional annotations for the MADS-box genes that showed gender-biased expression patterns in the data, the annotation of the MADS-box gene family presented by Leseberg et al. (2006) needed to

be updated to version 2.0 of the poplar genome. To create a dataset of all the putative MADS-box genes on version 2.0 of the poplar genome that corresponded to the PtMADS-box gene model sequences identified by Leseberg et al. (2006), the version 1.1 MADS-box gene sequences were BLASTed against version 2.0 of the poplar genome.

5.3.5 Phylogenetic analysis of MADS-box genes in *P. trichocarpa*

In order to correctly identify sequence homology between the genes showing a gender-biased expression pattern in my experiments and MADS-box genes with characterized functions in *A. thaliana* and other *Populus* species, I performed a phylogenetic analysis by aligning the M, I, and K protein domains from the *A. thaliana* genes *PI*, *AP1*, *AP3*, *AG*, *SHP1*, *SHP2* and *SEP 1/2/3/4* with the genes that showed novel expression data from the microarray and cDNA experiments (M1, M2, M4, M6, F2, SCA19_12, MADS_1 and MADS_2). The MIK protein domains from all of the *Populus* MADS-box genes with characterized function (*PTAG1*, *PTAG2*, *PTAP1-1*, *PTAP1-2*, *PTD*, *PTM/3/4/5* and *PtMADS31*) were also included, as well as all the PtMADS genes from Brunner and Nilsson (2004) for a total of 35 genes. The *Populus trichocarpa* MIK protein sequences were taken from version 2.0 of the poplar genome (http://www.phytozome.net), and the *A. thaliana* MADS-box genes were from the TAIR data base (http://www.arabidopsis.org/).

Protein alignments were initially compiled using the T-COFFEE webserver (available at http://www.tcoffee.org/) (Notredame et al., 2000), and then manually edited using BioEdit version 7.0.5.3, (10/28/05) (Hall, 1999). The alignment included the MADS-box domain of about 60 amino acids, I domain, and the K domain, for a total protein sequence of between 163 and 187 amino acids (Appendix E). The C domain was

excluded from the alignment, as it did not align well across the majority of the MADS-box protein sequences studied. Only MADS-box genes from version 2.0 of the poplar genome that corresponded to a single MADS-box gene annotated on version 1.1 of the poplar genome were included in the alignment. The phylogenetic programs RAxML 7.2.8 (Stamatakis, 2006) with the LG +G model for amino acid substitution, and GARLI 1.0 (Zwickl, 2006), with the WAG amino acid substitution model were used to construct the phylogenetic trees. These amino acid substitution models were the ones that had the lowest Akaike Information Criterion (AIC) score for each program, respectively. These two programs produced similar unrooted phylogenetic trees and bootstrap values, but the RAxML analysis yielded a phylogenetic tree with shorter branch lengths, so this tree was used to represent the MIK protein sequence homology in this study.

5.4 Results

5.4.1 Updating available MADS-box gene research from version 1.1 to version 2.0 of the poplar genomes

To identify the most current MADS-box gene annotations and identify which of these genes have been characterized with expression studies, I updated the entire list of *P. trichocarpa* MADS-box genes (PtMADS genes) identified by Leseberg et al. (2006) to the most current annotation data available for version 2.0 of the poplar genome. Because of the discrepancy between the way in which the two versions of the genome were assembled, a number of genes were identified that had changed position in the genome. Scaffolds that were previously not included in the sequence of chromosomes had been added. Several genes that were single genes on version 1.1 but BLASTed to two separate gene models on version 2.0 were observed (indicated in gray on Figure 5.1). In addition

some genes that were two or more gene models from version 1.1 of the poplar genome were annotated as single gene models on version 2.0 (indicated in yellow on Figure 5.1). Some of the genes previously annotated as MADS-box genes on version 1.1 of the poplar genome now have either no known annotated function, had a best BLAST hit in an intergenic region, or in the case of PtMADS90, did not have a single hit on version 2.0 of the genome. Finally, it had been previously reported that there were no floral MADS-box genes located on chromosome 19 (Leseberg et al., 2006), and I was able to identify two PtMADS-box genes located on that chromosome.

A number of floral MADS-box genes have been studied to confirm their functions and expression patterns in the genus *Populus*, and I was able to ascertain that all of these corresponded to PtMADS-box genes identified by Leseberg et al. (2006) either by direct comparison with the sequences of characterized *P. trichocarpa*, or by identifying the homologous *P. trichocarpa* gene if the MADS-box gene had been studied in another *Populus* species. Of the twelve MADS-box genes that had been previously characterized by expression studies in *Populus*, all twelve were located on the second version of the poplar genome (Figure 5.1). After updating the list of putative MADS-box genes reported in Leseberg et al. (2006) to version 2.0 of the poplar genome the total number of poplar MADS-box genes indicated by the updated data set is now 88, down from 105. The methods used may not have identified all the MADS-box genes in the poplar genome, and given the dynamic nature of genome annotation and the complexity of the poplar genome, it is conceivable that additional MADS-box genes could be detected in *Populus*. However, my data can provide a jumping off point for further studies of this gene family in *Populus*.

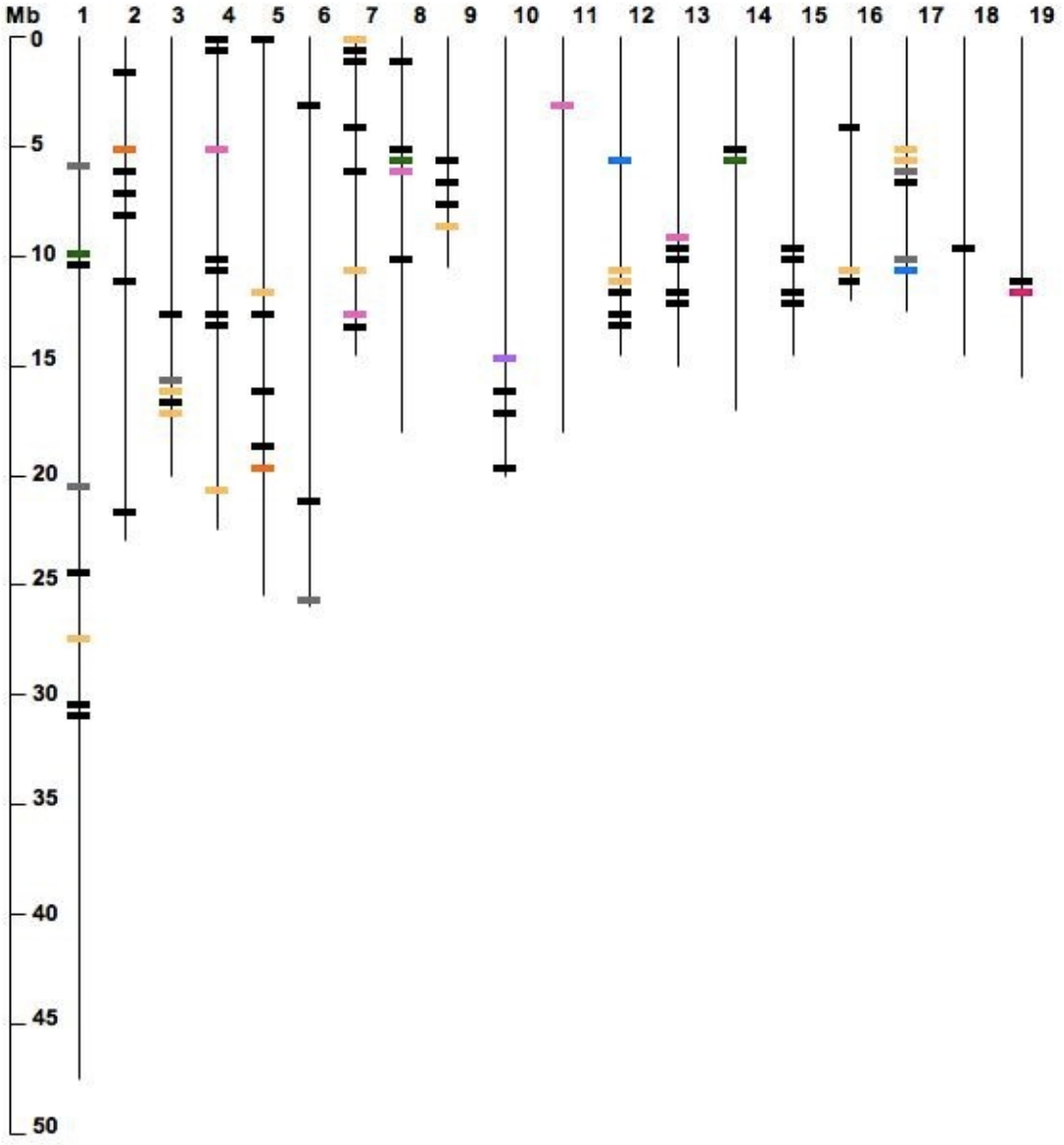

Figure 5.1 Distribution of 88 MADS-box genes on *P.trichocarpa* chromosomes.
The locations of the genes are approximate and not directly proportional to their real physical distances. Genes indicated in black have no expression studies associated with them to date. Genes in gray BLASTed to multiple genes on version 2.0 of the genome. Genes in yellow indicate two or more gene models from version 1.1 of the poplar genome that are now annotated as single gene models on version 2.0. Genes in green have been investigated in other expression studies, but showed no gender-bias expression on the microarray. Genes in light pink show a female-biased expression pattern on the microarray, and genes in blue show a male-biased expression pattern. One gene (purple) was equally expressed in male and female floral tissue. The two genes indicated in orange show male-biased expression, while the one gene indicated in dark pink on chromosome 19 showed female-biased expression from the microarray and cDNA data. Further genomic information on the genes indicated in this figure can be found in Appendix D.

5.4.2 Identifying *P. trichocarpa* MADS-box genes showing gender-biased expression using a microarray approach

The initial microarray experiment indicated that certain MADS-box genes showed a gender-bias in their expression in floral tissue. This result is not unexpected, considering the lack of female organs in male flowers, and vice versa. When male and female floral tissue were hybridized on the microarray, many genes showed male-biased expression, probably because these genes are involved in pollen production. To narrow this study to genes likely to be directly involved in floral development, I identified the top 14 genes that showed gender-biased expression, and were also annotated as MADS-box genes in the microarray file. One of these 14 genes is no longer annotated as a MADS-box gene, but 13 remain identified as such. The sole gene that showed a male gender-biased expression pattern but is not a MADS-box gene was labeled as MALE (M)5 (Table 5.1). It is annotated as a kinesin motor family protein, similar to the AtNACK1 kinesin-like protein in *A. thaliana* on version 2.0 of the poplar genome. As the protein sequence for M5 did not align with the protein sequences for the MADS-box genes in this study, it appears that the current annotation for this gene is correct. Comparatively little work has been done to characterize MADS-box gene-expression patterns in *P. trichocarpa*, but by comparing these results to those available in the literature, I was able to identify which of the genes indicated by the microarray as having gender-biased expression corresponded to *Populus* MADS-box genes that had previously been characterized (Table 5.1).

The genes that showed the most female-biased expression on the microarray were SCA19_12 and FEMALE (F) 2. Neither of these genes had been previously

characterized with expression data, but both of which group with the *A. thaliana STK* gene (Figure 5.2). The remaining four genes that showed female-biased gender expression on the microarray are floral MADS-box genes that have been previously studied in *Populus – PTAG1* and *PTAG2* (Brunner et al., 2000), *PTD* (Sheppard et al., 2000), *PTAP1-1* (Brunner and Nilsson, 2004b) and *PTM2* (Cseke et al., 2005), with *PTAP1-1* from *P. trichocarpa* homologous to *PTM2* in *P. tremuloides*. One gene, characterized as *PTAP1-2* in *P. trichocarpa* and *PTM1* in *P.tremuloides*, was expressed in both male and female floral tissues on the microarray, indicating that this MADS-box gene plays a role in both male and female floral development.

The two genes that showed the most male-biased expression were MADS_1 and MADS_2, and like SCA19_12 and F2, these genes have not been previously studied in detail in *P.trichocarpa* either, so I am reporting novel gender-biased expression for these four genes. The remaining four MADS-box genes showing male-biased expression, M1, M2, M4, M6, were identified as MADS-box genes in *P. trichocarpa* by Leseberg et al. (2006). However, no previous *Populus* expression data appears to be available for any of these six genes, so my research reports novel gender-biased expression for these genes.

Five other MADS-box genes have been studied in *Populus*, *PTM3*, *PTM4*, *PTM5*, *PTM6* (Cseke et al., 2005), and *PtMADS31* (Zhang et al., 2009), but none of these E-class genes showed gender-biased expression in floral tissues in this study, and the expression data from the previous studies indicates that these five genes are expressed in both male and female floral tissues (Cseke et al., 2005; Zhang et al., 2009).

Table 5.1 Relating available information on the expression of MADS-box genes in *Populus* species to MADS-box genes showing gender-biased expression patterns.
All data shown is from this study unless otherwise indicated by superscript.

Characterized Gene Name	Corresponding Gene Model	Gender-bias in this study	Gender-bias from other studies	Position on version 2.0 poplar genome	Species of Gene Characterization
SCA19_12	POPTR_0019s10580	F	N/A	scaffold_19: 12132258-12133799	*P. trichocarpa*
F2	POPTR_0013s09980	F	N/A	scaffold_13: 9955912-9966253	*P. trichocarpa*
PTAG2[1]	POPTR_0011s03140	F	No bias	scaffold_11: 3200188-3207704	*P. trichocarpa*
PTAP1-1[2] [PTM2 [3]]	POPTR_0008s09800	F	N/A	scaffold_8: 6076113-6078596	*P. trichocarpa*
PTAP1-2[2] [PTM1[3]]	POPTR_0010s16380	No bias	N/A	scaffold_10: 15345650-15351172	*P.trich/P. tremuloides*
PTD[4]	POPTR_0007s13660	F	No bias	scaffold_7: 13686094-13687851	*P. trichocarpa*
PTAG1[1]	POPTR_0004s06300	F	No bias	scaffold_4: 4997208-5005680	*P. trichocarpa*
M1	POPTR_0004s11420	M	N/A	scaffold_4: 10038841-10040605	*P. trichocarpa*
M2	POPTR_0012s05960	M	N/A	scaffold_12: 5936482-5942087	*P. trichocarpa*
M4	POPTR_0005s12000	M	N/A	scaffold_5: 8770246-8772014	*P. trichocarpa*
M5	POPTR_0002s02910	M	N/A	scaffold_2: 1789707-1795019	*P. trichocarpa*
M6	POPTR_0017s13410	M	N/A	scaffold_17: 13578276-13585496	*P. trichocarpa*
MADS_1	POPTR_0002s07920	M	N/A	scaffold_2: 5450756 – 5453302	*P. trichocarpa*
MADS_2	POPTR_0005s20480	M	N/A	scaffold_5: 19532615 – 19535235	*P. trichocarpa*
PtMADS31[5]	POPTR_0008s09790	N/A	No bias	scaffold_8: 6063858 – 6070604	*P. deltoides*
PTM3[3] and PTM4[3]	POPTR_0004s11440	N/A	No bias	scaffold_4: 10100959 – 10108181	*P. tremuloides*
PTM5[6]	POPTR_0014s07010	N/A	N/A	scaffold_4: 10100959 – 10108181	*P. tremuloides*
PTM6[3]	POPTR_0001s13650	N/A	No bias	scaffold_1: 10729860 – 10732842	*P. tremuloides*

[1](Brunner et al., 2000) [2](Brunner and Nilsson, 2004b) [3](Cseke et al., 2005) [4](Sheppard et al., 2000) [5](Zhang et al., 2009) [6](Cseke et al., 2007)

5.4.3 Identification of putative gene function based on protein sequence homology for *P. trichocarpa* MADS-box genes

The *Arabidopsis thaliana* genome only contains one copy of the majority of the floral MADS-box genes that are involved in floral development, while studies in *Populus* species have shown many MADS-box genes in this species have two homologs to every one *A. thaliana* MADS-box gene. An example of this is seen with the discovery of *PTAG1* and *PTAG2* in *P. trichocarpa*, which are homologous to *AG* in *A. thaliana* (Brunner et al., 2000), which is likely due to the fact that there has been a whole genome duplication event since the divergence of *Populus* and *Arabidopsis* (Tuskan et al., 2006). It is similarly logical to assume that the *Populus* MADS-box gene homologues have also probably differentiated in function following their origins in genome duplication events in evolutionary history of the genus.

With this in mind, I conducted a phylogenetic analysis that included the eleven of the MADS-box genes involved in floral development in *A. thaliana*, the MADS-box genes that showed gender-biased expression from the microarray experiment (13), and seven PtMADS-box genes from the phylogenetic analysis performed by Brunner and Nilsson (2004), which were included because these genes had been previously identified as being the *P. trichocarpa* homologs of *A. thaliana* MIKC group proteins known to be involved in floral development (Brunner and Nilsson, 2004b).

From the phylogenetic analysis it appears that the female-biased *PTD* gene and the male-biased M4 group closest to the *A. thaliana* gene *AP3*, while MADS_1 and MADS_2 group with *PI*. As expected, *PTAG1* and *PTAG2* group with *AG*, and *PtMADS31*, an E-class gene first characterized in *P. deltoides*, groups with the *SEP* genes

from *A. thaliana*. SCA19_12 and F2 group most closely with *STK*. M1, M2 and M6 group with *PTAP1-2* and *PTAP1-1/PTM1* with the *A. thaliana* gene *AP1*. The overall trend in my data seems to be that most floral MADS-box genes in *A. thaliana*, have two *P. trichocarpa* homologs (Table 5.2). Also, it appears that the *P. trichocarpa* homologs for C-class genes such as *AG* and *STK* show a female bias in their expression pattern, whereas the putative *PI P. trichocarpa* homologs, which are B-class genes, show a male-biased expression pattern.

Table 5.2 *Populus* MADS-box floral genes that have been characterized with expression patterns in floral tissues and their *A. thaliana* homologs.
Homologs were identified based on protein sequence homology of the MADS-box protein domain. All genes listed here showed gender-biased gene-expression in the microarray experiment, and some of these genes have been characterized in other studies, as indicated by superscript. Gene class refers to the ABC model of floral development.

Characterized Gene Name	Corresponding Gene Model	*A. thaliana* MADS-box homolog	A/B/C/E-class gene .
SCA19_12	POPTR_0019s10580	*SEEDSTICK* (*STK*)	C-class
F2	POPTR_0013s09980	*SEEDSTICK* (*STK*)	C-class
PTAG1[1]	POPTR_0004s06300	*AGAMOUS* (*AG*)	C-class
PTAG2[1]	POPTR_0011s03140	*AGAMOUS* (*AG*)	C-class
PTAP1-1[2]	POPTR_0008s09800	*APETALA1* (*AP1*)	A-class
PTAP1-2[2]	POPTR_0010s16380	*APETALA1* (*AP1*)	A-class
PTD[4]	POPTR_0007s13660	*APETALA3 (AP3)*	B-class
M1	POPTR_0004s11420	homology uncertain	N/A
M2	POPTR_0012s05960	homology uncertain	N/A
M4	POPTR_0005s12000	*APETALA3 (AP3)*	B-class
M6	POPTR_0017s13410	homology uncertain	N/A
MADS_1	POPTR_0002s07920	*PISTILLATA* (*PI*)	B-class
MADS_2	POPTR_0005s20480	*PISTILLATA* (*PI*)	B-class .

[1](Brunner et al., 2000) [2](Brunner and Nilsson, 2004b) [3](Sheppard et al., 2000)

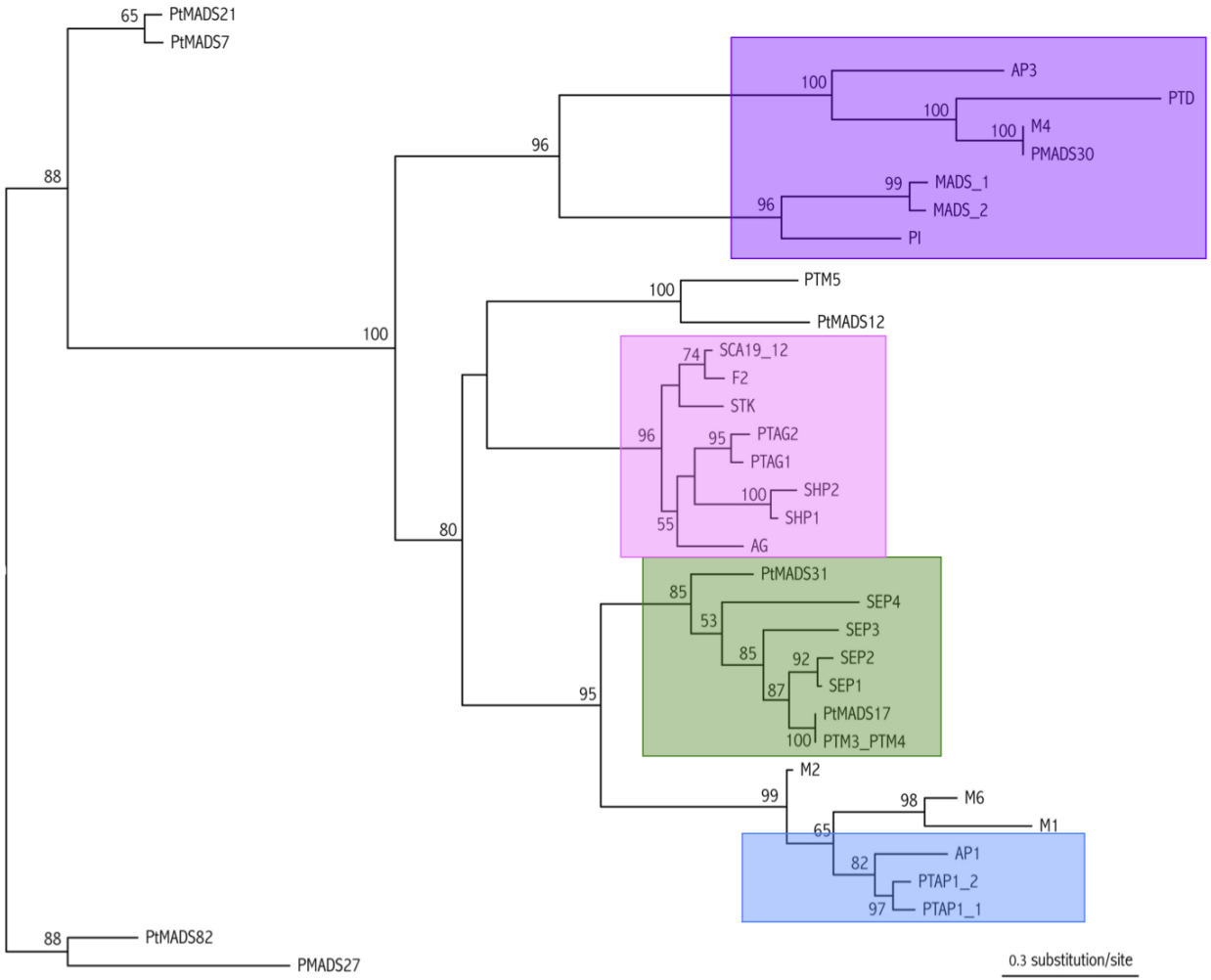

Figure 5.2 Un-rooted phylogenetic tree produced using a Maxmimum Likelihood (ML) analysis to show the homology in protein sequence data between the MADS-box domain from the poplar floral MADS-box genes and *A.thaliana* floral MADS-box genes. Colour shading indicates the class of floral MADS-box genes. Blue = A-class, purple = B-class, pink = C-class, and green = E-class.

5.4.4 Confirming the MADS-box gene microarray results using cDNA expression experiments

I confirmed the gender-biased expression pattern observed on the microarray for one gene showing female expression bias (SCA19_12) and two genes exhibiting male expression bias (MADS_1 and MADS_2) by using reverse-transcription PCR with a sample of floral tissue from four female individuals and three males. I chose to further investigate SCA19_12 because it showed the most female-biased expression value, and it is located on chromosome 19, which is the putative sex chromosome in *P. trichocarpa*. I investigated MADS_1 and MADS_2 because they showed the most male-biased expression values and appear to be the *P. trichocarpa PI* homologs, and *PI* homologs have not previously been investigated in *P. trichocarpa*.

When testing the primers designed to amplify cDNA of SCA19_12, I discovered that this gene is expressed in female, but not male, floral tissues (Figure 5.3). Also, this gene does not appear to be expressed in the leaf tissue of male or female individuals (Figure 5.4). The bands resulting from the cDNA amplification of SCA19_12 were sequenced, and by comparing the PCR product sequence to the sequence for this gene that is annotated on the poplar genome, it was confirmed that these bands represent the target sequence.

The two genes MADS_1 and MADS_2 showed amplification products in male, but not female, floral tissues (Figure 5.3), and like SCA19_12, these genes do not appear to be expressed in the leaf tissues of either gender. The bands observed when amplifying these two genes were confirmed to be the target sequence by sequencing the PCR products, and comparing the sequence to the poplar genome.

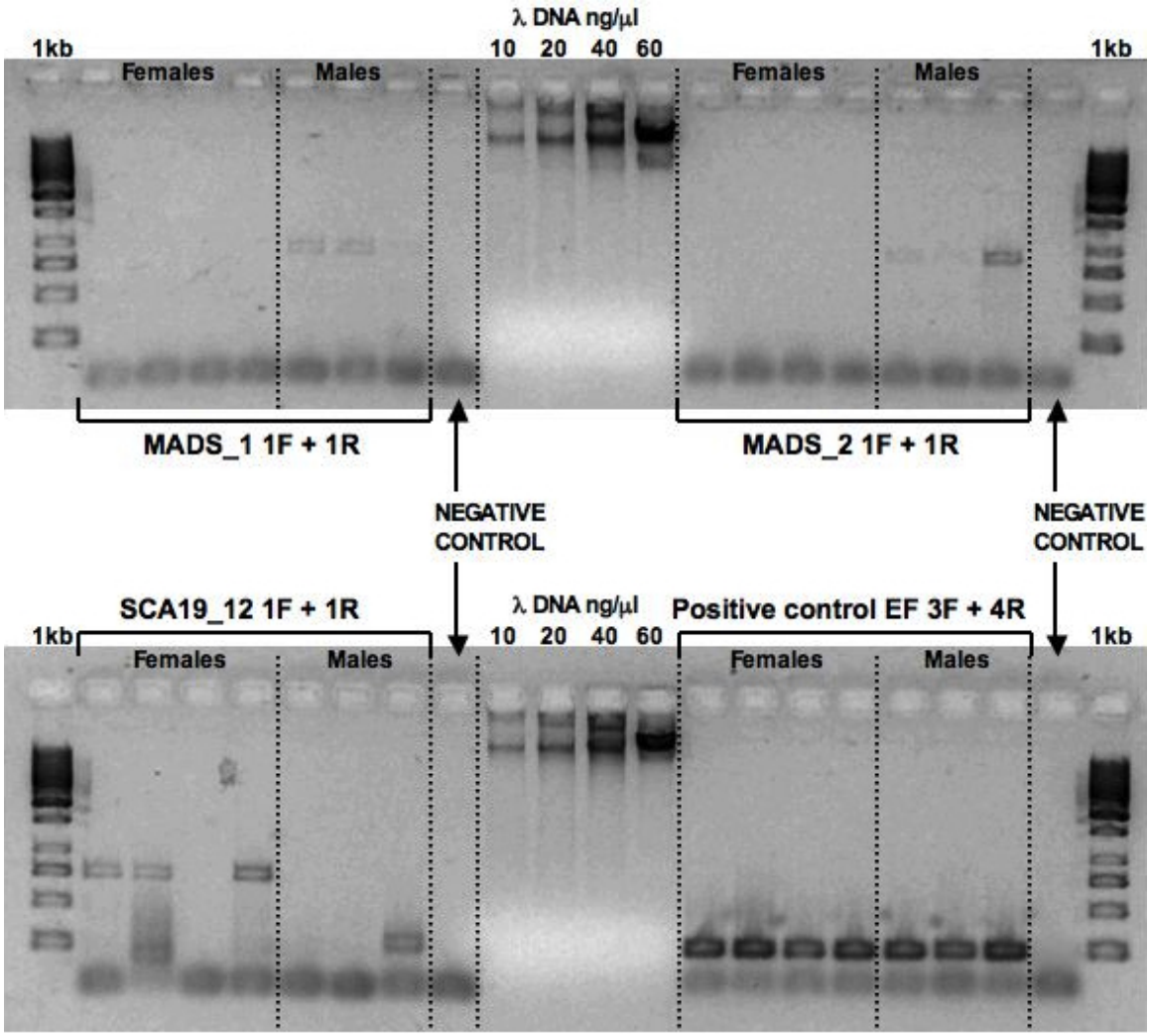

Figure 5.3 Electrophoresis gel photo showing differential expression patterns measured by PCR of cDNA from male and female *P. trichocarpa* flower tissues.
Four female and three male individuals were tested. MADS_1 and MADS_2 are genes expressed in male but not female floral tissues. SCA19_12 is expressed is female floral tissues but not male floral tissues. The positive control for this experiment was EF, a constitutively expressed transcription factor. λ DNA of known concentration was used to give an estimate of cDNA concentration in the bands on the gel, and a 1kb ladder was used to estimate the fragment size in the bands.

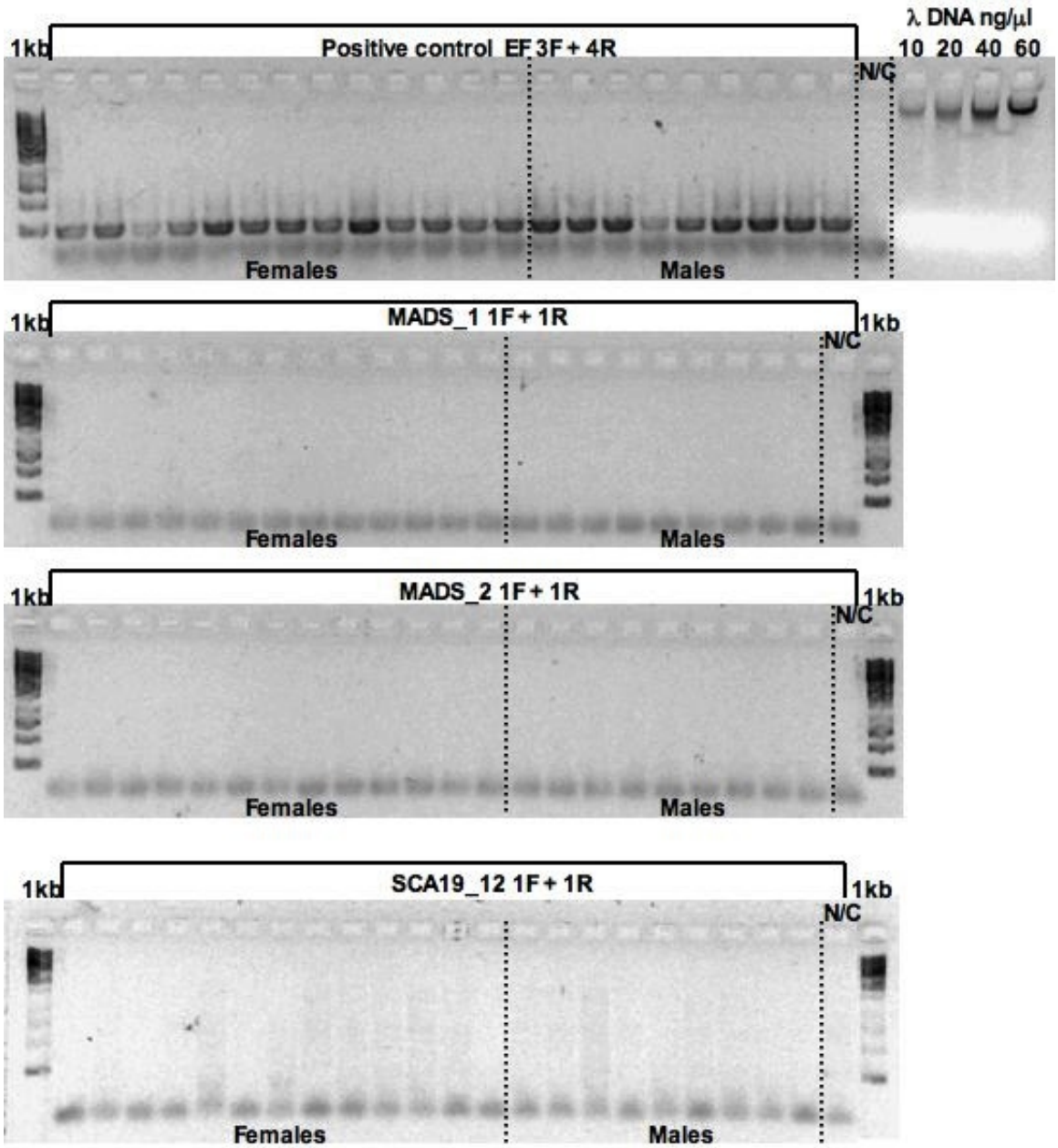

Figure 5.4 Electrophoresis gel photo showing no detectible expression in male and female *P. trichocarpa* leaf tissue of the MADS box genes that had differential expression patterns in male and female floral tissues.
Leaf tissue from thirteen female and nine male individuals was tested with MADS_1, MADS_2 and SCA19_12. The positive control for this experiment was EF, a constitutively expressed transcription factor. λ DNA of known concentration was used to give an estimate of cDNA concentration in the bands on the gel, and a 1kb ladder was used to estimate the fragment size in the bands.

5.5 Discussion

5.5.1 MADS-box genes in the second version of the *P. trichocarpa* genome.

The release of the second version of the *P. trichocarpa* genome in January 2010 provided an opportunity to conduct a survey of the current knowledge on MADS-box genes involved in floral development in *Populus*, and to provide a comprehensive overview of the functional and expression studies available for this gene family in *P. trichocarpa*. Using the genome wide analysis of the MADS-box gene family conducted by Leseberg et al. (2006) as a starting point, and including studies done on floral MADS-box genes from *P. tremula*, *P. tremuloides*, and *P. deltoides* as well as the expression data generated using a microarray approach in *P. trichocarpa* in this study, I created a body of MADS-box gene data that was based on version 2.0 of the poplar genome. To date, all of the published studies of MADS-box genes in *Populus* have relied on version 1.1 of the poplar genome. My work updates the knowledge of these genes to the latest gene position and annotation data available for version 2.0 of the poplar genome.

One of the most interesting findings made during this process is that two floral *P. trichocarpa* MADS-box (PtMADS) genes were located on chromosome 19, when it had been previously reported that there were no floral MADS-box genes located on this chromosome (Leseberg et al., 2006). This is particularly intriguing because chromosome 19 is thought to be the emerging sex chromosome in *P. trichocarpa* and other *Populus* species that have been investigated (Paolucci et al., 2010).

In *Populus*, depending on the species, the putative sex locus appears to map either to a telomeric position, or an internal position (Paolucci et al., 2010). In *P. trichocarpa* the sex locus is thought to occur in a telomeric position (Yin et al., 2008). SCA19_12 and

PtMADS42 both exist in an internal position, so it is unlikely that they are directly involved in gender-determination at the putative sex locus. But if a multi-locus system controls gender-determination in *Populus* as has been proposed (Alström-Rapaport et al., 1998), it is possible that one or both of these genes could be involved in gender-determination. Alternatively, they could be targeted by gender-determining genes located at the site of the putative sex locus at the telomeric end of chromosome 19.

5.5.2 Gender development in poplar flowers

The model species that have been used to study floral development include *Arabidopsis thaliana*, *Antirrhinum majus*, and *Petunia hybrida*, all of which are hermaphroditic species that develop bisexual flowers (Zluvova et al., 2006). The genetics of floral development have also been studied in the dioecious species *Silene latifolia* and *Rumex acetosa,* which both develop unisexual flowers, but appear to abort the male organs in female flowers, and vice versa, early in floral development. In *S. latifolia*, floral bud development is classified into twelve stages, and prior to stage five of development, flowers of both genders develop as bisexual flower primordia. However, once male flowers reach stage six the gynoecium primordia stops differentiating, and in female flowers after stage seven the development of stamen primordia is arrested, so anthers do not develop (Grant et al., 1994) In *Rumex acetosa* four whorls of floral organs are initiated in both male and female flowers, but in each gender development of the non-gender appropriate set of organs (ie: stamens in female flowers and carpels in male flowers) is stopped soon after initiation (Ainsworth et al., 2005). In *P. trichocarpa* however, there is no evidence of the initiation or abortion of inappropriate sets of organs in either the male of female flowers, and the evidence of a reduced perianth suggests that the unisexual flowers in *Populus* are a result of evolution

favoring fewer floral whorls (Boes and Strauss, 1994). This evolution of fewer floral whorls is probably a result of the evolution of dioecy in *Populus* being via the monoecious pathway (chapter one). The monoecious pathway would allow these flowers the long evolutionary time required for the extensive diversification of gene function necessary for the development of unisexual flowers. Therefore, studying the genes involved in floral differentiation in *P. trichocarpa* may provide insight into how the ABC genetic model of floral development works in flowers that are strictly unisexual, and may indicate which floral MADS-box genes have differentiated to have only male or female expression. Given that B and C-class genes are respectively involved in the development of male and female floral organs, these classes of genes seem to be likely candidates to have a role in the gender asymmetry that is found in unisexual flowers (Meagher, 2007).

5.5.3 Comparing the MADS-box genes involved in floral development in *A. thaliana* and *P. trichocarpa*

In *Populus* the study of genes involved in floral development is still a developing field. Using a microarray approach to do a genome wide survey of floral tissue gene expression proved useful in identifying genes that are involved in floral development, and allowed a gender-bias in the expression of some of these genes to be observed. Comparing the MADS-box genes expressed in floral tissues in *P. trichocarpa* to floral MADS-box genes already well characterized in *A. thaliana* provides a valuable tool for inferring floral gene functions in *P. trichocarpa*. This is possible because the *P. trichocarpa* genome has been shown to contain orthologs of the major genes that regulate floral development in *A. thaliana* (Brunner, 2010). *P. trichocarpa* and *A. thaliana* both belong to the rosid clade of the Core

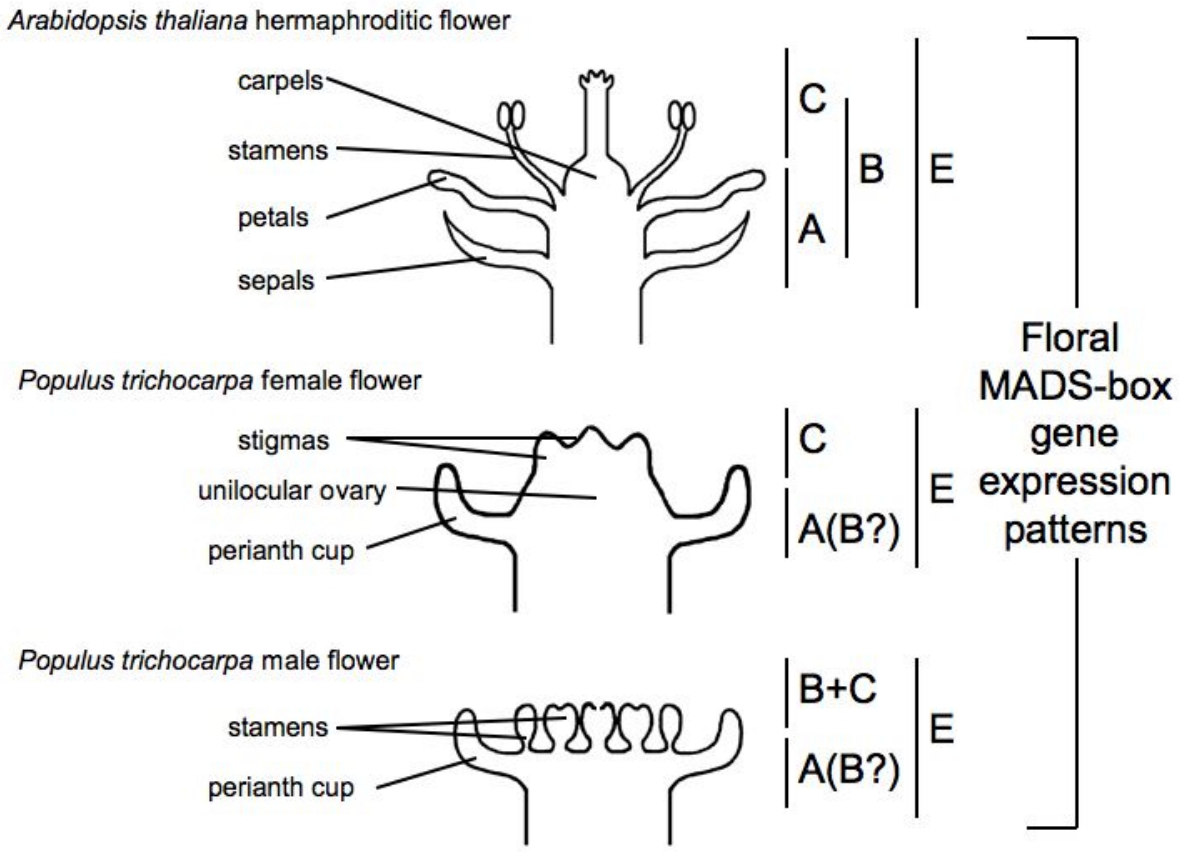

Figure 5.5 Comparing differences in MADS-box gene-expression in floral tissues between *A. thaliana* and *P. trichocarpa*.
The lines on the right next to the labels A, B, C and D indicate the whorls in which these classes of genes are expressed.

Eudicots, which allows for the comparison of gene functions between the two species (Brunner et al., 2004a). Based on the ABC model of MADS-box gene-expression in *A. thaliana* (Coen and Meyerowitz, 1991), and the gene-expression studies done in *Populus* by Brunner et al. (2000), Skinner et al., (2003), Cseke et al. (2005) and Zhang et al. (2010), as well as the data reported in this study, a comparison has been made here between MADS-box gene-expression patterns in the perfect flowers of *A. thaliana*, and the unisexual flowers found in *P. trichocarpa* (Figure 5.5). Comparing the floral MADS-box gene-expression

patterns found in *P. trichocarpa* to those that have been studied in *A. thaliana* provides information about how the ABC model has been altered to result in the development of unisexual flowers.

5.5.4 A-class gene-expression in *P. trichocarpa* inflorescences

In the ABC genetic model of floral development, genes *APETALA1* (*AP1*) and *APETALA2* (*AP2*) are A-class genes, that, when expressed alone specify sepal development, and when expressed with B-class genes form petals (Theißen, 2001). In *A. thaliana*, *AP1* is an A-class gene that is initially expressed in the entire floral meristems, but later its expression is limited to specifying sepal and petal identity (Zik and Irish, 2003). The microarray data presented in this study indicates that *PTAP1-1* has a female-biased expression pattern, which is interesting because if this gene is an A-class gene like *AP1*, it should be expressed in the first whorl of floral organs (Theißen, 2001), which corresponds to the sepals, or perianth cup in *Populus* flowers, which is found in both male and female flowers (Boes and Strauss, 1994). The *PTAP1-2* gene was expressed in both male and female floral tissues according to the microarray data, indicating that this MADS-box gene plays a role in both male and female floral development.

AP2, the second A-class gene, was not included in the phylogenetic analysis in this study because while it plays a role in floral meristem identification and is initially expressed throughout the floral meristems in *A. thaliana* (Zik and Irish, 2003), it is not a MADS-box gene but rather a ethylene responsive element binding protein (EREBP), a family of at least 12 proteins in *A. thaliana* that are characterized by a specific *AP2* protein domain (Okamuro et al., 1997).

Three genes indicated as having male-biased expression from the microarray results, M1, M2 and M6, were identified as being MADS-box genes in *P. trichocarpa* by Leseberg et al. (2006), however, no previous *Populus* expression data appears to be available for these three genes, so this research introduces novel gender-biased gene-expression data for M1, M2 and M6. These three genes grouped most closely with *PTAP1-1* and *PTAP1-2* in the phylogenetic analysis (Figure 5.2). It is possible that *PTAP1-1*, *PTAP1-2*, M1, M2 and M6 have similar roles in floral development in *P. trichocarpa*, with *PTAP1-1* and *PTAP1-2* having specific female functions, and with M1, M2 and M6 having specific male functions. Further investigation of the putative A-class genes in *P. trichocarpa* is required to determine their function in floral development, and confirm the gender-biased expression that was observed in this study.

5.5.5 B-class gene-expression in *P. trichocarpa* inflorescences

B-class function in the ABC model of floral development is contributed to by two genes, *APETALA3* (*AP3*) and *PISTILLATA* (*PI*), which together with A-class genes specify petal development, or when paired with C-class genes are involved in stamen development (Theißen, 2001). The two MADS-box genes that showed the highest male-biased expression pattern in the microarray data were MADS_1 and MADS_2. When the sequence of the MADS-box protein domain for these two genes was aligned with MADS-box genes from *P. trichocarpa* and *A. thaliana*, both of these genes were positioned most closely in the tree with the *A. thaliana* gene *PI* (Figure 5.2). cDNA expression patterns confirmed that MADS_1 and MADS_2 are expressed in male, but not in female floral tissues, and that there was no detectible expression of these two genes in leaf tissues (Figures 5.3 and 5.4). In *A. thaliana, PI* is necessary for the specification of organ identity in the second and third whorls of

developing perfect flowers, and is therefore involved in designating petal and stamen identity (Goto and Meyerowitz, 1994). As flowers in *P. trichocarpa* do not have petals, it is most likely that these two genes are involved in stamen development, which would also explain why they have a male-biased expression pattern in floral tissue.

In *A. thaliana*, *AP3* plays a role in determining the development of the second and third whorls of organs in perfect flowers, the petal and stamens (Krizek and Meyerowitz, 1996). Two genes grouped most closely with AP3 according to the phylogenetic analysis performed in this study, M4 and *PTD* (Figure 5.2). *PTD* is already identified as the *P. trichocarpa* homolog of *AP3* (Sheppard et al., 2000) and showed a female-biased expression according to the microarray data in this study. This is an interesting result because *AP3* is a B-class gene that is required to specify petal and stamen identity (Krizek and Meyerowitz, 1996), and female *P. trichocarpa* flowers have neither petals, nor stamens. Sheppard et al. (2000) reported that *PTD* is expressed initially in the inner whorls of both male and female flowers, but once reproductive primordia began to form the spatial expression pattern became gender-specific, with *PTD* expression continuing in the stamen primordia, but excluded from the carpel primordia. M4 exhibited male-biased expression according to the microarray data, and this study reports novel expression data for this gene. Flowers of *P. trichocarpa* do not have petals (Boes and Strauss, 1994), therefore it may be plausible that the male-biased expression pattern that was observed for the gene M4, along with its protein sequence similarities to *AP3*, indicate that this gene has a role in stamen development in the male flowers.

If *PTD* and M4 are the result of a genome duplication that has occurred in *Populus* since the genus diverged from *Arabidopsis* between 100 and 120 million years ago (Tuskan

et al., 2006), it could be that these duplicate genes have diverged in function as unisexual flowers evolved in *Populus*. There is evidence from whole genome studies that genes that are retained as duplicates in the genome frequently diversify in function, or develop subfunctionalization (Adams and Wendel, 2005). Further work will be required to test if this expression pattern is an indication of a differentiation of function of *AP3* in a unisexual flower.

5.5.6 C-class gene-expression in *P. trichocarpa* inflorescences

In the ABC model of floral development, C function in the third and fourth whorls of flowers is contributed to by the gene *AGAMOUS* (*AG*) (Theißen, 2001), and to a lesser extent *AGAMOUS*-like genes such as *SHATTERPROOF1* (*SHP1*) and *SHATTERPROOF2* (*SHP2*) which are involved in seed dispersal (Liljegren et al., 2000). In *A. thaliana*, both *SHP1* and *SHP2* are C-class genes (Table 5.2), positively regulated by *AG*, with *AG* directly regulating *SHP2* by binding to sequences in the *SHP2* promoter (Zik and Irish, 2003). *SHP1* and *SHP2* are required for fruit dehiscence in *A. thaliana*, and are functionally redundant in this species, as mutations in either of these genes independently cause no phenotypic changes, but the double mutant *shp1 shp2* result in fruits that will not break open (Liljegren et al., 2000). *SEEDSTICK* (*STK*) is another C-class gene that is necessary for the development of the funiculus, which connects the fruit to the seed, and plays a role in seed dispersal (Pinyopich et al., 2003).

The two genes that showed the most female-biased expression on the microarray were SCA19_12 and F2, and neither of these genes had been previously characterized with expression data, but both of these genes group most closely with the *A. thaliana SEEDSTICK* (*STK*) gene (Figure 5.2). While SCA19_12 and F2 show different expression levels on the

microarray, both are expressed in female flowers, and not in males. SCA12_19 may be involved in at least carpel development, if not the development of the fruit, as STK is involved in this process in *A. thaliana* (Pinyopich et al., 2003). cDNA expression patterns confirmed that SCA19_12 is expressed in female floral tissues, but not in male inflorecences, and that there was no detectible expression of this gene in leaf tissues (Figures 5.3 and 5.4). Now that SCA12_19 and F2 have been identified by this work, and the expression of SCA19_12 has been confirmed to be strongly female flower specific, future research could take a more detailed look at specific expression patterns of SCA19_12 and F2 in female flowers using an *in situ* approach or perhaps by creating transgenic *A. thaliana* mutants.

The remaining two genes that showed female-biased gender expression in this microarray study are floral MADS-box genes that have been previously studied in *Populus*. *PTAG1* and *PTAG2* have been shown to have an AG-like expression pattern in flowers. While this data shows that they have a female-biased expression pattern in floral tissues, previous work has indicated that they are also consistently expressed in vegetative tissues (Brunner et al., 2000). The phylogenetic analysis performed in this study confirms that *PTAG1* and *PTAG2* are positioned most closely with *A. thaliana AG*, based on sequence similarity of the MIK protein domains.

5.5.7 E-class gene-expression in *P. trichocarpa* inflorescences

After the inception of the ABC model of floral development it was found that a fourth class of genes was important in specifying the different whorls in floral meristems, and this class of genes was called the E-class (Theißen, 2001). This group of genes, named *SEPALLATA* (*SEP1*, *SEP2*, *SEP3* and *SEP4*), are important in the quartet model of floral organ identity in *Arabidopsis* because they code for proteins that form quartets with the other

three classes of MADS-box genes to result in the development of floral organs (Theißen, 2001). E-class genes are expressed throughout the developing floral meristems in *Arabidopsis* (Figure 5.5).

E-class genes in *Populus* have been characterized in the species *P. tremuloides* and *P. deltoides*, but not in *P. trichocarpa*. *PtMADS31*, the *P. trichocarpa* homolog of *PdMADS2*, an E-class gene first characterized in *P. deltoides* (Zhang et al., 2009), grouped with the *A. thaliana SEP* genes, as does *PTM5* (Cseke et al., 2007), an E-class gene first characterized in *P. tremuloides* (Table 5.1). These genes showed no gender-biased gene-expression in this microarray study (Table 5.1), and both genes have been previously shown to be expressed in both male and female floral tissues, as well as some vegetative tissues. In *P. deltoides*, *PdMADS2* is highly expressed in the perianth cups of male inflorescences, and the ovaries of female flowers, and also shows abundant expression in the peduncles of both genders (Zhang et al., 2009). In *P. trichocarpa*, *PtMADS31* has not been the subject of expression studies to date, so further research would be required to establish if PtMADS31 has a similar expression pattern to *PdMADS2* in *P. trichocarpa* flowers. In *P. tremuloides*, *PTM5* has been shown to be a member of the SOC1/TM3 class of MADS-box genes and is involved in protein-protein interactions with itself and other MADS-box proteins (Cseke et al., 2007), but the *P. trichocarpa* homolog has not yet been studied.

Three other MADS-box genes have been studied in *Populus*, *PTM3*, *PTM4*, and *PTM6* (Cseke et al., 2005), and while previous work has shown that these genes are expressed in both male and female tissues, none of these E-class genes showed gender-biased expression in floral tissues according to this microarray study (Table 5.1). *PTM3* and its duplicate *PTM4* are related to the *SEP1* and *SEP2* genes found in *A. thaliana*, and *PTM6* is

related to *SEP3*, and the concentration of expression of these genes in the inner whorls that contain the sexual organs of both male and female flowers suggests that in *P. deltoides* these genes are important in ensuring reproductive viability (Cseke et al., 2005). Given that *SEPALLATA* homologs in *Populus* appear to be expressed in all floral tissues studied to date in this genus, it is plausible that they play a similar role to the *SEPALLATA* genes that form protein quartets in *A. thaliana* to specify floral organ development.

5.6 Conclusions

This research updates the known floral MADS-box genes to the most current genomic annotations and locations, and places 88 floral MADS-box genes on version 2.0 of the poplar genome. My microarray experiment identified 14 putative MADS-box genes that showed gender-biased expression patterns in male and female inflorescences, and 13 of these genes were confirmed as correctly annotated as MADS-box genes. Novel expression patterns for nine floral MADS-box genes were identified with this microarray data, and the expression patterns of three of these genes was investigated in further detail using reverse-transcription PCR. I was able to show that SCA19_12 was expressed in female but not male floral tissues, and based on protein sequence homology is the *P. trichocarpa* homolog of the *A. thaliana STK* gene. I also discovered that MADS_1 and MADS_2 were expressed in male but not female floral tissues, and are homologs to the *A. thaliana* gene *PI*. None of these three genes were expressed in leaf tissues of male or female individuals at a detectible level.

Chapter 6: Conclusions

The research presented in this thesis adds to the current knowledge available on the model organism *Populus trichocarpa* with respect to the use of the genetic resources available for this species, and investigated the development of sex-linked markers in *P. trichocarpa* that could be used to identify the gender of trees prior to flowering. This thesis is comprised of four projects that took various approaches to discover how gender in *P. trichocarpa* might be regulated at a genetic level. There were two principal ideas underlying the four projects. The first idea was to develop a genetic marker that could be used to sex *P. trichocarpa* individuals of unknown gender using vegetative tissue. Having a marker like this would be useful to *Populus* breeding programs as it would allow the selection of either all male or all female progeny for clonal propagation, and it would allow the gender of experimental crosses to be identified prior to these individuals reaching the age of first flowering, which can take as long as 15 years in some *Populus* species. The second idea was to investigate the genes involved in floral development in *P. trichocarpa* and compare gene-expression patterns in this dioecious, unisexual flower species with those of the homologous genes in the model organism *A. thaliana*, which produces hermaphroditic flowers. In this thesis, chapters two, three and four report the results of projects that investigated the development of a genetic marker for gender in vegetative tissues, and chapter five details the results of a study of MADS-box genes involved in floral development in *P. trichocarpa*.

6.1 Developing a genetic marker to identify the gender of an individual using vegetative tissues

SCAR-markers have been discovered that segregate with gender in pedigreed families of *Salix viminalis*, a finding that led to the development of the hypothesis that gender in this species may be determined genetically via several loci, perhaps functioning in an epistatic way (Alström-Rapaport et al., 1998). Because *Salix* is the sister genus to *Populus* in the Salicaceae, and it is probable that both these genera evolved dioecy, and therefore genetic sex-determination mechanisms via the monoecious pathway, I wanted to see if the sex-linked markers discovered in *S. viminalis* could be used to develop a genetic maker for gender in vegetative tissues in other *Salix* species and in *P. trichocarpa*.

I identified the homologous sequence in *P. trichocarpa* for the *Salix viminalis* sex-linked marker SCAR-354 on chromosome 15 on the poplar genome. I was also able to amplify the *Salix* SCAR 354 marker sequence as well as the adjacent gene sequence for a Ssu72-like protein. I characterized the SCAR marker sequence, and identified distinct regions of the marker that varied in the degree to which they were conserved between the species sampled. When I investigated the gene-anchored sequence obtained in *S. arctica* and *S. reticulata* male and female individuals I found some evidence that gender-biased SNPs do exist in this sequence, which may explain why the SCAR 354 marker segregates with gender in pedigreed families of *S. viminalis* (Alström-Rapaport et al., 1998). However, the gender bias observed was not statistically significant. I was unable to confirm that the SCAR 354 marker that segregated with sex in *S. viminalis* was also a sex-linked marker in other *Salix* species or in *P. trichocarpa*,

Research into the genetics of gender-determination in *Populus* indicates that, as in *Salix*, one or two chromosomal regions or sex loci are involved in gender-determination. The development of the sequenced *P. trichocarpa* genome has allowed other researchers to associate the putative sex loci in this species to chromosome 19. For the project reported on in chapter three of this thesis, I investigated 24 genes in the telomeric region of chromosome 19 looking for SNP variation in the genomic sequence, or reduced recombination rates in SNPs between males and females in these genes, which could be associated with genetic markers for gender and a sex locus. The reasoning behind this was that it has been shown that areas of reduced recombination between the genders in a genome can indicate the initiation of a sex chromosome (Liu et al., 2004). I observed large variability in the number of SNPs detected in the gene sequences studied, but discovered no genetic marker that could be used to sex *P. trichocarpa* individuals of unknown gender. There was no overall trend of reduced recombination between adjacent SNPs in the gene sequences investigated, and the data collected shows that low recombination rates between SNPs are not maintained across the telomeric region on chromosome 19. If there is reduced recombination in the region of a sex locus on the telomeric end of chromosome 19, it would appear that it is very localized, as I looked at genes located less than 100kb from each other. Alternatively, it is possible that the sex locus is not in the ~1Mbp region I investigated. Gender-determination mechanisms have been investigated in at least six species of *Populus* (*P. alba*, *P. detoides*, *P. nigra*, *P. tremula*, *P. tremuloides* and *P. trichocarpa*), using pedigreed families and mapping maternal and paternal genetic markers in progeny of known gender (Paolucci et al., 2010). These studies, along with maturing genetic resources such as the re-annotated version 2.0 of the *P.*

trichocarpa genome makes it likely that developing a full understanding the genetics of gender-determination in this species will only be a matter of time.

Shortly after the release of version 1.1 of the poplar genome I conducted two microarray experiments to look for gene-expression patterns that differed between male and female leaf tissues. I was able to identify a number of genes that showed small differences in expression levels between males and females, however the majority of these differences were not statistically significant, and my subsequent investigation using reverse transcription PCR of the genes was unable to confirm the microarray results. As a result I was unable to identify potential genomic sex-linked markers in *P. trichocarpa* using this method. Sample size and replication are both critical when conducting this kind of gene-expression experiment, as illustrated by the observation that there was a greater gene-expression difference between the two microarray experiments performed than between genders. Gender-biased gene-expression may be present in vegetative tissues of *P. trichocarpa*, but if so, these effects are subtle, and therefore a much larger sample size and better environmental control would be needed for a microarray experiment to be able to detect it with any statistical confidence.

Another conclusion I came to during the course of this project was that revisiting data and updating it frequently to reflect the latest and most accurate gene annotations available is essential when working with a rapidly developing model system like *P. trichocarpa*. The release of version 2.0 of the poplar genome provided much additional information on the genes I was working with, which greatly improved the quality of my analysis, as I was able to more accurately place genes showing gender-biased expression on chromosome 19, the putative sex chromosome in *P. trichocarpa*.

6.2 Investigation of the genetics of floral development in *P. trichocarpa*

P. trichocarpa produces flowers that appear to be unisexual throughout their development, as they only initiate either male or female floral organs. There is no evidence that organs of the opposite gender are initiated and then later aborted during floral development. Studying the genetics of floral development in *P. trichocarpa* provides an opportunity to investigate how the genes involved in this process have differentiated to have either male- or female-specific expression. In chapter five of this thesis a microarray experiment using male and female inflorescence tissues showed that a number of floral MADS-box genes were differentially expressed between males and females. Since this gene family has never been systematically investigated in *Populus*, a phylogenetic analysis of *P. trichocarpa* MADS-box genes was conducted to compare the MIK protein domains of these genes to floral MADS-box genes that have had their functions thoroughly characterized in *A. thaliana.* I discovered that among the MADS-box genes that had a gender-biased expression pattern in *Populus* floral tissues there were often two *P. trichocarpa* homologs for each of the *A. thaliana* genes I looked at. There appears to have been some divergence in function of each paralogue, as they do not always share the same gender-bias in their expression, or the same levels of expression in a given tissue.

My microarray experiment identified 14 putative MADS-box genes that showed putative gender-biased expression patterns in male and female floral tissue, and 13 of these genes were confirmed as MADS-box genes. I identified which of these genes had been previously characterized in *Populus* as having functions in flower development in this species, and which of these genes were identified as having floral expression patterns for the first time by my research. Novel expression patterns for nine floral MADS-box genes were

identified with this microarray data, and the expression patterns of three of these genes were investigated in further detail using reverse transcription PCR. I was able to show that SCA19_12 was expressed in female but not male floral tissues, and determined that SCA19_12 is the *P. trichocarpa* homolog of the *A. thaliana STK* gene based on protein sequence homology. I also discovered that *Populus* MADS_1 and MADS_2 were expressed in male but not female floral tissues, and are homologs to the *A. thaliana* gene *PI*. None of these three genes were expressed in leaf tissues of male or female individuals at a detectable level.

With the release of the second version of the poplar genome I was able to update the known floral MADS-box genes to the most current genomic annotations and locations, and the total number of floral MADS-box genes according to my survey is now 88, down from the 105 first reported in Leseberg et al. (2006). Interestingly, I confirmed that there are two floral MADS-box genes located on chromosome 19, whereas it was previously reported that MADS-box genes were distributed across all chromosomes in *P. trichocarpa* except chromosome 19. One of these genes, SCA19_12, is expressed only in female floral tissues, and not in male floral tissues, or leaf tissues of either gender. Chromosome 19 is the location of the putative sex locus in *P. trichocarpa*, and the female of this species is thought to be the heterogametic gender (Yin et al., 2008), so it is interesting that a gene that is only expressed in female flowers is found on this chromosome.

6.3 Future directions and applications of this research

Given that the conserved Ssu72-like gene associated region consistently amplified in the *Salix* species I sampled, this marker may be useful for phylogenetic and population studies of willows. Current molecular phylogenetic studies in *Salix* have used chloroplast

genetic markers to resolve the phylogenetic relationships in this taxonomically difficult genus (Chen et al., 2010). Using the gene-anchored primers developed in this work it may be possible to discover species-specific single feature polymorphisms (SFPs), and these SFPs could prove to be useful phylogenetic markers for identifying *Salix* species, which can be difficult to classify based on morphological traits alone.

Advances in cost effective, high throughput next-generation sequencing technologies that are currently available from Illumina IG, Applied Biosystems SoLiD and Roche 454 Life Sciences Systems (Wang et al., 2009) will make it much easier to generate and manage larger genetic data sets, and to keep the gene annotations up to date. Currently, 20 *P. trichocarpa* accessions are included in a population transcriptome resequencing project which promises to provide much more detailed information about gene annotation of the *P. trichocarpa* genome (Geraldes et al., 2011). Data sets of this kind would make it possible to detect with greater accuracy the subtle gender-biased gene-expression differences in vegetative tissues in *P. trichocarpa* that my experiments indicated exist.

Future research into the genetics of gender-determination in poplar needs to address why genetic markers for some species of *Populus* indicate that the male is the heterogametic gender, and in other species they indicate that the female is heterogametic. Genetic data from resequencing the transcriptomes of 20 *P. trichocarpa* individuals could be used to identify genetic markers linked to gender once these trees have reached sexual maturity and can be sexed by observing floral gender. This is because this resource allows genotypic variability, which includes genetic difference between the genders, to be linked to gene-expression (Geraldes et al., 2011). Also, investigating chromosome 19 in the transcriptomes

of multiple *P. trichocarpa* individuals could explain why at least two regions on chromosome 19 in *Populus* species have been identified as being involved in sex-determination.

This research constitutes the first survey of *Populus* floral MADS-box genes conducted using the updated genome annotations available for version 2.0 of the poplar genome and therefore provides a jumping off point for further research into the genetics of floral development in *P. trichocarpa*. A logical next step for continuing work on MADS-box genes involved in floral development in *P. trichocarpa* would be to confirm the gender bias I observed in the six other genes that were identified as having novel gender-biased expression with the microarray experiment using the cDNA synthesis technique used to investigate SCA19_12, MADS_1 and MADS_2 in a larger sample size of floral tissues of male and female *P. trichocarpa* individuals. It would be important to establish if these genes are expressed only in floral tissues, or if they are expressed in other tissues as well. Eventually, conducting *in situ* experiments to determine detailed information about the expression patterns of these genes in the floral whorls of male and female *P. trichocarpa* flowers would be useful in order to establish the specific functions of these genes in floral development.

References

Acosta, I.F., Laparra, H., Romero, S.P., Schmelz, E., Hamberg, M., Mottinger, J.P., Moreno, M.A., and Dellaporta, S., L. (2009). tasselseed1 is a lipoxygenase affecting jasmonic acid signalling in sex determination on maize. Science *323*, 262-265.

Adams, K.L., and Wendel, J.F. (2005). Polyploidy and genome evolution in plants. Current Opinion in Plant Biology *8*, 135-141.

Ainsworth, C. (2000). Boys and girls come out to play: the molecular biology of dioecious plants. Annals of Botany *86*, 211-221.

Ainsworth, C., Crossley, S., Buchanan-Wollaston, V., Thangavelu, M., and Parker, J. (1995). Male and female flowers of the dioecious plant sorrel show different patterns of MADS box gene expression. The Plant Cell *7*, 1583-1598.

Ainsworth, C., Rahman, A., Parker, J., and Edwards, G. (2005). Intersex inflorescences of *Rumex acetosa* demonstrate that sex determination is unique to each flower. New Phytologist *165*, 711-720.

Allen, C.E. (1917). A chromosome difference correlated with sex in *Sphaerocarpos*. Science *46*, 466-467.

Alström-Rapaport, C., Lascoux, M., Wang, Y.C., Roberts, G., and Tuskan, G.A. (1998). Identification of a RAPD marker linked to sex determination in the basket willow (*Salix viminalis* L.). Journal of Heredity *89*, 44-49.

Alvarez-Buylla, E.R., Pelaz, S., Liljegren, S.J., Gold, S.E., Burgeff, C., Ditta, G.S., Ribas de Pouplana, L., Martinez-Castilla, L., and Yanofsky, M.F. (2000). An ancestral MADS-box gene duplication occurred before the divergence of plants and animals. Proceedings of the National Academy of Sciences *97*, 5328-5333.

Anderson, L.E. (2000). Charles E. Allen and sex chromosomes. The Bryologist *103*, 442-448.

Azuma, T., Kajita, T., Yokoyama, J., and Ohashi, H. (2000). Phylogenetic relationships of *Salix* (Salicaceae) based on rbcL sequence data. American Journal of Botany *87*, 67-75.
Banks, J.A. (2008). MicroRNA, sex determination and floral meristem determinacy in maize. Genome Biology *9*, 1-3.

Barrett, S.C.H. (2002). The evolution of plant sexual diversity. Nature Reviews *3*, 274-284.

Batzoglou, S., Jaffe, D.B., Stanley, K., Butler, J., Gnerre, S., Mauceli, E., Berger, B., Mesirov, J.P., and Lander, E.S. (2002). ARACHNE: A whole-genome shotgun assembler. Genome Research *12*, 177-189.

Bawa, K.S. (1980). Evolution of dioecy in flowering plants. Annual Review of Ecology and Systematics *11*, 15-39.

Beach, J.H., and Bawa, K.S. (1980). Role of pollinators in the evolution of dioecy from distyly. Evolution *34*, 1138-1142.

Bergero, R., and Charlesworth, D. (2009). The evolution of restricted recombination in sex chromosomes. Trends in Ecology & Evolution *24*, 94-102.

Bergero, R., Forrest, A., Kamau, E., and Charlesworht, D. (2007). Evolution strata on the X chromosomes of the dioecious plant *Silene latifolia*: Evidence from new sex-linked genes. Genetics *175*, 1945-1954.

Blackburn, K.B. (1923). Sex chromosomes in plants. Nature *112*, 687-688.

Boes, T.K., and Strauss, S.H. (1994). Floral phenology and morphology of black cottonwood, *Populus trichocarpa* (Salicaceae). American Journal of Botany *81*, 562-567.

Böhlenius, H., Huang, T., Charbonnel-Campaa, L., Brunner, A.M., Jansson, S., Strauss, S.H., and Nilsson, O. (2006). *CO/FT* regulatory module controls timing of flowering and seasonal growth cessation in trees. Science *312*, 1040-1043.

Bolstad, B.M., Irizarry, R.A., Astrand, M., and Speed, T.P. (2003). A comparison of normalization methods for high density oligonucleotide array data based on variance and bias. Bioinformatics *19*, 185-193.

Boualem, A., Fergany, M., Fernandez, R., Troadec, C., Martin, A., Morin, H., Sari, M.-A., Collin, F., Flowers, J.M., Pitrat, M.*, et al.* (2008). A conserved mutation in an ethylene biosynthesis enzyme leads to andromonoecy in melons. Science *321*, 836-838.

Boualem, A., Troadec, C., Kovalski, I., Sari, M.-A., Perl-Treves, R., and Bendahmane, A. (2009). A conserved ethylene biosynthesis enzyme leads to andromonoecy in two *Cucumis* species. PLoS ONE *4*, e6144.

Braatne, J.H., Rood, R.B., and Heilman, P.E. (1996). Life history, ecology, and conservation of riparian cottonwoods in North America. In Biology of *Populus* and its implications for management and conservation, R.F. Stettler, B.H.D. Jr., P.E. Heilman, and T.M. Hinckley, eds. (Ottawa, NRC Research Press), pp. 57-89.

Bracale, M., Caporali, E., Galli, M.G., Longo, C., Marziani-Longo, G., Rossi, G., Spada, A., Soave, C., Falavigna, A., Raffaldi, F.*, et al.* (1991). Sex determination and differentiation in *Asparagus officinalis* L. Plant Science *80*, 67-77.

Breen, A.L., Glenn, E., Yeager, A., and Olson, M.S. (2009). Nucleotide diversity among natural populations of a North American poplar (*Populus balsamifera*, Salicaceae). New Phytologist *182*, 763-773.

Browse, J. (2009). Jasmonate: Preventing the Maize tassel from getting in touch with his feminine side. Science Signalling *2*, 1-9.

Brunner, A.M. (2010). Reproductive development in *Populus*. In Plant Genetics and Genomics: Crops and Models, pp. 155-170.

Brunner, A.M., Busov, V., B., and Strauss, S.H. (2004a). Poplar genome sequence: functional genomics in an ecologically dominant plant species. Trends in Plant Science *9*, 49-56.

Brunner, A.M., and Nilsson, O. (2004b). Revisiting tree maturation and floral initiation in the poplar functional genomics era. New Phytologist *164*, 43-51.

Brunner, A.M., Rottmann, W.H., Sheppard, L.A., Krutovskii, K., DiFazio, S.P., Leonardi, S., and Strauss, S.H. (2000). Structure and expression of duplicate *AGAMOUS* orthologues in poplar. Plant Molecular Biology *44*, 619-634.

Brush, S.G. (1978). Nettie M. Stevens and the discovery of sex determination of chromosomes. Isis *69*, 162-172.

Calderon-Urrea, A., and Dellaporta, S.L. (1999). Cell death and cell protection genes determine the fate of pistils in maize. Development *126*, 435-441.

Charlesworth, B., and Charlesworth, D. (1978). A model for the evolution of dioecy and gynodioecy. American Naturalist *112*, 975-997.

Charlesworth, D., and Charlesworth, B. (1979). The evolutionary genetics of sexual systems in flowering plants. Proceedings of the Royal Society of London Series B, Biological Sciences *205*, 513-530.

Charlesworth, D., and Guttman, D.S. (1999). The evolution of dioecy and plant sex chromosome systems. In Sex determination in plants, C.C. Ainsworth, ed. (Oxford, Bios Scientific Publishers).

Charlesworth, D., and Mank, J.E. (2010). The Birds and the Bees and the Flowers and the Trees: Lessons from Genetic Mapping of Sex Determination in Plants and Animals. Genetics *186*, 9-31.

Chen, J.-H., Sun, H., Wen, J., and Yang, Y.-P. (2010). Molecular phylogeny of *Salix* L. (Salicaceae) inferred from three chloroplast datasets and its systematic implications. Taxon *59*, 29-37.

Coen, E., S., and Meyerowitz, E., M. (1991). The war of the whorls: genetic interactions controlling flower development. Nature *353*, 31-37.

Colombo, L., Franken, J., Koetje, E., van Went, J., Dons, H., Angenent, G.C., and van Tunen, A.J. (1995). The Petunia MADS box gene FBP11 determines ovule identity. The Plant Cell *7*, 1859-1868.

Comtois, P., Simon, J.P., and Payette, S. (1986). Clonal constitution and sex ratio in northern populations of balsam poplar *Populus balsamifera.* Holarctic Ecology *9*, 251-260.

Cooper, S.J., Trinklein, N.D., Anton, E.D., Nguyen, L., and Myers, R.M. (2006). Comprehensive analysis of transcriptional promoter structure and function in 1% of the human genome. Genome Research *16*, 1-10.

Cox, P.A. (1981). Niche Partitioning between Sexes of Dioecious Plants. The American Naturalist *117*, 295-307.

Crawford, R.M.M., and Balfour, J. (1983). Female Predominant Sex Ratios and Physiological Differentiation in Arctic Willows. Journal of Ecology *71*, 149-160.

Cronk, Q., C.B. (2005). Plant eco-devo: the potential of poplar as a model organism. New Phytologist *166*, 39-48.

Cseke, L.J., Cseke, S.B., Ravinder, N., Taylor, L.C., Shankar, A., Sen, B., Thakur, R., Karnosky, D.F., and Podila, G.K. (2005). *SEP*-class genes in *Populus tremuloides* and their likely role in reproductive survival of poplar trees. Gene *358*, 1-16.

Cseke, L.J., Ravinder, N., Pandey, A.K., and Podila, G.K. (2007). Identification of *PTM5* protein interaction partners, a MADS-box gene involved in aspen tree vegetative development. Gene *391*, 209-222.

Dawson, T.E., and Bliss, L.C. (1989). Patterns of water use and the tissue water relations in the dioecious shrub, *Salix arctica*: the physiological basis for habitat partitioning between the sexes. Oecologia *79*, 332-343.

Dellaporta, S.L., and Calderon-Urrea, A. (1994). The sex determination process in maize. Science *266*, 1501-1505.

Di Stilio, V.S., Kramer, E.M., and Baum, D., A (2005). Floral MADS box genes and homeotic gender dimorphism in *Thalictrum dioicum* (Ranunculaceae) - a new model for the study of dioecy. The Plant Journal *41*, 755-766.

Dichtl, B., Blank, D., Ohnacker, M., Friedlein, A., Roeder, D., Langen, H., and Keller, W. (2002). A Role for SSU72 in Balancing RNA Polymerase II Transcription Elongation and Termination. Molecular cell *10*, 1139-1150.

Dorken, M.E., and Pannell, J.R. (2009). Hermaphroditic sex allocation evolves when mating opportunities change. Current Biology *19*, 514-517.

Doyle, J.J., and Doyle, J.L. (1987). A rapid DNA isolation procedure for small quantities of fresh leaf tissue. Phytochemical Bullentin *19*, 11-15.

Eckenwalder, J.E. (1996). Systematics and evolution of *Populus*. In Biology of *Populus* and its Implications for Management and Conservation, R.F. Stettler, H.D.J. Bradshaw, P.E. Heilman, and T.M. Hinckley, eds. (Ottawa, ON, Canada, NRC Research Press, National Research Council of Canada), pp. 7-32.

Einspahr, D.W. (1960). Sex ratio in quaking aspen and possible sex-related characteristics. Proceedings of the Fifth World Forestry Congress *2*, 747-750.

Ewing, B., and Green, P. (1998). Base-calling of automated sequencer traces using phred. II. Error probabilities. Genome Research *8*, 186-194.

Ewing, B., Hillier, L., Wendl, M.C., and Green, P. (1998). Base-calling of automated sequencer traces using phred. I. Accuracy assessment. Genome Research *8*, 175-185.

Farmer, R.E.J. (1964). Sex ratio and sex-related characteristic in eastern cottonwood. Silvae Genetica *13*, 116-118.

Feng, S., Cokus, S.J., Zhang, X., Chen, P.-Y., Bostick, M., Goll, M.G., Hetzel, J., Jain, J., Strauss, S.H., Halpern, M.E*., et al.* (2010). Conservation and divergence of methylation patterning in plants and animals. Proceedings of the National Academy of Sciences *107*, 8689-8694.

Filatov, D.A., Monéger, F., Negrutiu, I., and Charlesworth, D. (2000). Low variability in a y-linked plant gene and its implications for Y-chromosome evolution. Nature *404*, 388-390.
Fisher, M.J. (1928). The morphology and anatomy of the flowers of the Salicaceae II. American Journal of Botany *15*, 372-394.

Frewen, B.E., Chen, T.H.H., Howe, G.T., Davis, J., Rohde, A., Boerjan, W., and Bradshaw, H.D. (2000). Quantitative trait loci and candidate gene mapping of bud set and bud flush in *Populus*. Genetics *154*, 837-845.

Ganem, C., Devaux, F., Torchet, C., Jacq, C., Quevillon-Cheruel, S., Labesse, G., Facca, C., and Faye, G. (2003). Ssu72 is a phosphatase essential for transcription termination of snoRNAs and specific mRNAs in yeast. EMBO J *22*, 1588-1598.

Gaudet, M., Jorge, V., Paolucci, I., Beritognolo, I., Scarascia Mugnozza, G., and Sabatti, M. (2008). Genetic linkage maps of *Populus nigra* L. including AFLPs, SSRs, SNPs, and sex trait. Tree Genetics and Genomes *4*, 25-36.

Geraldes, A., Pang, J., Thiessen, N., Cezard, T., Moore, R., Zhao, Y., Tam, A., Wang, S., Friedmann, M., Birol, I*., et al.* (2011). SNP discovery in black cottonwood (*Populus trichocarpa*) by population transcriptome resequencing. Molecular Ecology Resources *11*, 81-92.

Gilchrist, E.J., Haughn, G.W., Ying, C.C., Otto, S.P., Zhuang, J., Cheung, D., Hamberger, B., Aboutorabi, F., Kalynyak, T., Johnson, L.*, et al.* (2006). Use of ecotilling as an efficient SNP discovery tool ot survey genetic variation in wild populations of *Populus trichocarpa*. Molecular Ecology *15*, 1367-1378.

Goldberg, M.T., Spigler, R.B., and Ashman, T.-L. (2010). Comparative genetic mapping points to different sex chromosomes in sibling species of wild strawberry (*Fragaria*). Genetics, genetics.110.122911.

Goto, K., and Meyerowitz, E.M. (1994). Function and regulation of the *Arabidopsis* floral homeotic gene *PISTILLATA*. Genes & Development *8*, 1548-1560.

Graham, S.W., and Olmstead, R.G. (2000). Utility of 17 chloroplast genes for inferring the phylogeny of the basal angiosperms. American Journal of Botany *87*, 1712-1730.

Gramzow, L., and Theißen, G. (2010). A hitchhiker's guide to the MADS world of plants. Genome Biology *11*, 214-223.

Grant, M.C., and Mitton, J.B. (1979). Elevational gradients in adult sex ratios and sexual differentiation in vegetative growth rates of *Populus tremuloides* Michx. Evolution *33*, 914-918.

Grant, S., Hunkirchen, B., and Saedler, H. (1994). Developmental differences between male and female flowers in the dioecious plant *Silene latifolia*. The Plant Journal *6*, 471-480.

Gunter, L.E., Kopp, R.F., McCord, R.P., and Tuskan, G.A. (2003a). Analysis of sex-linked, sequence-characterized amplified region markers in *Salix eriocephala*. Canadian Journal of Forest Research *33*, 1785-1790.

Gunter, L.E., Roberts, G.T., Lee, K., Larimer, F.W., and Tuskan, G.A. (2003b). The development of two flanking SCAR markers linked to a sex determination locus in *Salix viminalis* L. Journal of Heredity *94*, 185-189.

Hake, L., and O'Conner, C. (2008). Genetic mechanisms for sex determination. Nature Education *1*, 1-5.

Hall, T., A. (1999). BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. Nucleic Acids Symposium Series *41*, 95-98.

Hamrick, J.L., Godt, M.J.W., and Sherman-Broyles, S.L. (1992). Factors influencing levels of genetic diversity in woody plant species. New Forests *6*, 95-124.

Hamzeh, M., and Dayanandan, S. (2004). Phylogeny of Populus (Salicaceae) based on nucleotide sequences of chloroplast TRNT-TRNF region and nuclear rDNA. American Journal of Botany *91*, 1398-1408.

Hamzeh, M., Périnet, P., and Dayanandan, S. (2006). Genetic Relationships among species of *Populus* (Salicaceae) based on nuclear genomic data. The Journal of the Torrey Botanical Society *133*, 519-527.

Hardenack, S., De, Y., Saedler, H., and Grant, S. (1994). Comparison of MADS box gene expression in developing male and female flowers of the dioecious plant white campion. The Plant Cell *6*, 1775-1787.

Henschel, K., Kofuji, R., Hasebe, M., Saedler, H., Münster, T., and Theißen, G. (2002). Two ancient classes of MIKC-type MADS-box genes are present in the moss *Physcomitrella patens*. Molecular Biology and Evolution *19*, 801-814.

Hewitt, G.M. (2004). Genetic consequences of climatic oscillations in the Quaternary. Philosophical Transactions of the Royal Society of London Series B: Biological Sciences *359*, 183-195.

Hou, W., Fan, J., Zhou, F., and Zhao, S. (2009). RAPD markers related to sex locus in *Populus tementosa*. Frontiers of Forestry in China *4*, 223-226.

Hsu, C.-Y., Adams, J.P., Kim, H., No, K., Ma, C., Strauss, S.H., Drnevich, J., Vandervelde, L., Ellis, J.D., Rice, B.M.*, et al.* (2011). *FLOWERING LOCUS T* duplication coordinates reproductive and vegetative growth in perennial poplar. Proceedings of the National Academy of Sciences *108*, 10756-10761.

Hsu, C.-Y., Liu, Y., Luthe, D.S., and Yuceer, C. (2006). Poplar *FT2* shortens the juvenile phase and promotes seasonal flowering. The Plant Cell *18*, 1846-1861.

Huber, W., von Heydebreck, A., Sültmann, H., Poustka, A., and Vingron, M. (2002). Variance stabilization applied to microarray data calibration and to the quantification of differential expression. Bioinformatics *18*, S96-S104.

Ingvarsson, P.r.K. (2005). Nucleotide polymorphism and linkage disequilibrium within and among natural populations of European aspen (*Populus tremula* L., Salicaceae). Genetics *169*, 945-953.

Ingvarsson, P.r.K. (2008). Multilocus patterns of nucleotide polymorphism and the demographic history of *Populus tremula*. Genetics *180*, 329-340.

Irish, E.E., and Nelson, T. (1989). Sex determination in monoecious and dioecious plants. The Plant Cell *1*, 737-744.

Jamilena, M., Mariotti, B., and Manzano, S. (2008). Plant sex chromosomes: molecular structure and function. Cytogenetic and Genome Research *120*, 255-264.

Jamsari, A., NItz, I., Reamon-Büttner, S.M., and Jung, C. (2004). BAC-derived diagnostic markers for sex determination in asparagus. Theoretical and Applied Genetics *108*, 1140-1146.

Janousek, B., and Mrackova, M. (2010). Sex chromosomes and sex determination pathway dynamics in plant and animal models. Biological Journal of the Linnean Society *100*, 737-752.

Jansson, S., and Douglas, C.J. (2007). *Populus*: A model system for plant biology. Annual Review of Plant Biology *58*, 435-458.

Kafadar, K., and Phang, T. (2003). Transformations, background estimation, and process effects in the statistical analysis of microarrays. Computational Statistics and Data Analysis *44*, 313-338.

Kaul, R.B. (1995). Reproductive structure and organogenesis in a cottonwood, *Populus deltoides* (Salicaceae). International Journal of Plant Sciences *156*, 172-180.

Kejnovsky, E., Kubat, Z., Hobza, R., Lengerova, M., Sato, S., Tabata, S., Fukui, K., Matsunaga, S., and Vyskot, B. (2006). Accumulation of chloroplast DNA sequences on the Y chromosome of *Silene latifolia*. Genetica *128*, 167-175.

Keller, S.R., Olson, M.S., Silim, S., Schroeder, W., and Tiffin, P. (2010). Genomic diversity, population structure, and migration following rapid range expansion in the Balsam Poplar, *Populus balsamifera*. Molecualt Ecology *19*, 1212-1226.

Kellogg, E., A., and Bennetzen, J.L. (2004). The evolution of nuclear genome structure in seed plants. American Journal of Botany *91*, 1709-1725.

Khadka, D., Nejidat, A., Tal, M., and Golan-Goldhirsh, A. (2005). Molecular characterization of a gender-linked DNA marker and a related gene in *Mercurialis annua* L. Planta *222*, 1063-1070.

Khattak, J.Z.K., Torp, A.M., and Anderson, S.B. (2006). A genetic linkage map of *Spinacia oleracea* and localization of a sex determination locus. Euphytica *148*, 311-318.

Kim, J.C., Laparra, H., Calderón-Urrea, A., Mottinger, J.P., Moreno, M.A., and Dellaporta, S.L. (2007). Cell cycle arrest of stamen initials in maize sex determination. Genetics *177*, 2547-2551.

Knopf, R.R., and Trebitsh, T. (2006). The female-specific *Cs-ACS1G* gene of cucumber. A case of gene duplication and recombination between the non-sex-specific 1-aminocyclopropane-1-carboxylate synthase gene and a branched-chain amino acid transaminase gene. Plant and Cell Physiology *47*, 1217-1228.

Kolosova, N., Miller, B., Ralph, S., Ellis, B.E., Douglas, C., Ritland, K., and Bohlmann, J. (2004). Isolation of high quality RNA from gymnosperm and angiosperm trees. BioTechniques *35*, 821-824.

Kramer, E.M., Jaramillo, M.A., and Di Stilio, V.n.S. (2004). Patterns of gene duplication and functional evolution during the diversification of the *AGAMOUS* subfamily of MADS box genes in angiosperms. Genetics *166*, 1011-1023.

Krizek, B.A., and Meyerowitz, E.M. (1996). The *Arabidopsis* homeotic genes *APETALA3* and *PISTILLATA* are sufficient to provide the B class organ identity function. Development *122*, 11-22.

Larkin, M.A., Blackshields, G., Brown, N.P., Chenna, R., McGettigan, P.A., McWilliam, H., Valentin, F., Wallace, I.M., Wilm, A., Lopez, R.*, et al.* (2007). Clustal W and Clustal X version 2.0. Bioinformatics *23*, 2947-2948.

Lebel-Hardenack, S., and Grant, S.R. (1997). Genetics of sex determination in flowering plants. Trends in Plant Science *2*, 130-136.

Leseberg, C.H., Li, A., Kang, H., Duvall, M., and Mao, L. (2006). Genome-wide analysis of the MADS-box gene family in *Populus trichocarpa*. Gene *378*, 84-94.

Librado, P., and Rozas, J. (2009). DnaSP v5: a software for comprehensive analysis of DNA polymorphism data. Bioinformatics *25*, 1451-1452.

Liljegren, S.J., Ditta, G.S., Eshed, Y., Savidge, B., Bowman, J.L., and Yanofsky, M.F. (2000). *SHATTERPROOF* MADS-box genes control seed dispersal in *Arabidopsis*. Nature *404*, 766-770.

Liu, J., Huang, Y., Ding, B., and Tauer, C.G. (1999). cDNA cloning and expression of a sweetgum gene that shows homology with *Arabidopsis AGAMOUS*. Plant Science *142*, 73-82.

Liu, S., Xu, L., Jia, Z., Xu, Y., Yang, Q., Fei, Z., Lu, X., Chen, H., and Huang, S. (2008). Genetic association of *ETHYLENE-INSENSITIVE3*-like sequence with the sex-determining *M*; locus in cucumber (*Cucumis sativus* L.). Theoretical and Applied Genetics *117*, 927-933.

Liu, Z., Moore, P.H., Ma, H., Ackerman, C.M., Ragiba, M., Yu, Q., Pearl, H.M., Kim, M.S., Charlton, J.W., Stiles, J.I.*, et al.* (2004). A primitive Y chromosome in papaya marks incipient sex chromosome evolution. Nature *427*, 348-352.

Lloyd, D.G., and Webb, C.J. (1977). Secondary sex characters in plants. Botanical Review *43*, 177-216.

Maddison, D.R., and Maddison, W.P. (2000). MacClads 4: analysis of phylogeny and character evolution, version 4.0., 4th edn (Sinauer).

Mariotti, B., Navajas-Pérez, R., Lozano, R., Parker, J.S., de la Herrán, R., Ruiz Rejón, C., Ruiz Rejón, M., Garrido-Ramos, M., and Jamilena, M. (2006). Cloning and characterization of dispersed repetitive DNA derived from microdissected sex chromosomes of *Rumex acetosa*. Genome *49*, 114-121.

Meagher, T.R. (1984). Sexual dimorphism and ecological differentiation of male and female plants. Annals of the Missouri Botanical Gardens *71*, 254-264.

Meagher, T.R. (2007). Linking the evolution of gender variation to floral development. Annals of Botany *100*, 165-176.

Mibus, H., and Tatlioglu, T. (2004). Molecular characterization and isolation of the *F/f* gene for femaleness in cucumber (*Cucumis sativus* L.). TAG Theoretical and Applied Genetics *109*, 1669-1676.

Ming, R., and Moore, P.H. (2007). Genomics of sex chromosomes. Current Opinion in Plant Biology *10*, 123-130.

Ming, R., Wnag, J., Moore, P.H., and Paterson, A.H. (2007). Sex chromosomes in flowering plants. American Journal of Botany *94*, 141-150.

Moore, R.C. (2009). Sex chromosome evolution: A 'missing link' in the evolution of sex chromosomes. Heredity *102*, 211-212.

Muenchow, G.E., and Grebus, M. (1989). The evolution of dioecy from distyly: Reevaluation of the hypothesis of the loss of long-tongued pollinators. The American Naturalist *133*, 149-156.

Münster, T., Pahnke, J., Di Rosa, A., Kim, J.T., Martin, W., Saedler, H., and Theißen, G. (1997). Floral homeotic genes were recruited from homologous MADS-box genes pre-existing in the common ancestor of ferns and seed plants. Proceedings of the National Academy of Sciences *94*, 2415-2420.

Nam, J., dePamphilis, C.W., Ma, H., and Nei, M. (2003). Antiquity and evolution of the MADS-box gene family controlling flower development in plants. Molecular Biology and Evolution *20*, 1435-1447.

Nei, M., and Miller, J.C. (1990). A simple method for estimating average number of nucleotide substitutions within and between populations from restriction data. . Genetics *125*, 873-879.

Nickerson, D.A., Tobe, V.O., and Taylor, S.L. (1997). PolyPhred: automating the detection and genotyping of single nucleotide substitutions using fluorescence-based resequencing. Nucleic Acids Research *25*, 2745-2751.

Nicolas, M., Marais, G., Hykelova, V., Janousek, B., Laporte, V., Vyskot, B., Mouchiroud, D., Negrutiu, I., Charlesworth, D., and Monéger, F. (2005). A gradual process of recombination restriction in the evolution history of the sex chromosome in dioecious plants. PLoS Biology *3*, 47-56.

Nielsen, J.L., Rood, S.B., Pearce, D.W., Letts, M.G., and Jiskoot, H. (2010). Streamside trees: responses of male, female and hybrid cottonwoods to flooding. Tree Physiology *30*, 1479-1488.

Notredame, C., Higgins, D.G., and Heringa, J. (2000). T-coffee: a novel method for fast and accurate multiple sequence alignment. Journal of Molecular Biology *302*, 205-217.

Ogata, M., Hasegawa, Y., Ohtani, H., Mineyama, M., and Miura, I. (2007). The ZZ/ZW sex-determining mechanism originated twice and independently during evolution of the frog, *Rana rugosa*. Heredity *100*, 92-99.

Ogata, M., Ohtani, H., Igarashi, T., Hasegawa, Y., Ichikawa, Y., and Miura, I. (2003). Change of the Heterogametic Sex From Male to Female in the Frog. Genetics *164*, 613-620.

Okada, S., Sone, T., Fujisawa, M., Nakayama, S., Takenaka, M., Ishizaki, K., Kono, K., Shimizu-Ueda, Y., Hanajiri, T., Yamato, K.T.*, et al.* (2001). The Y chromosome in the liverwort *Marchantia polymorpha* has accumulated unique repeat sequences harboring a male-specific gene. Proceedings of the National Academy of Sciences of the United States of America *98*, 9454-9459.

Okamuro, J.K., Caster, B., Villarroel, R., Van Montagu, M., and Jofuku, K.D. (1997). The *AP2* domain of *APETALA2* defines a large new family of DNA binding proteins in *Arabidopsis*. Proceedings of the National Academy of Sciences *94*, 7076-7081.

Olson, M.S., Robertson, A.L., Takebayashi, N., Silim, S., Schroeder, W.R., and Tiffin, P. (2010). Nucleotide diversity and linkage disequilibrium in balsam poplar (*Populus balsamifera*). New Phytologist *186*, 526-536.

Pakull, B., Groppe, K., Mecucci, F., Gaudet, M., Sabatti, M., and Fladung, M. (2011). Genetic mapping of linkage group XIX and identification of sex-linked SSR markers in a *Populus tremula* x *Populus tremuloides* cross. Canadian Journal of Forest Research *41*, 245-253.

Pakull, B., Groppe, K., Meyer, M., Markussen, T., and Fladung, M. (2009). Genetic linkage mapping in aspen (*Populus tremula* L. and *Populus tremuloides* Michx.). Tree Genetics and Genomes *5*, 505-515.

Paolucci, I., Gaudet, M., Jorge, V., Beritognolo, I., Terzoli, S., Kuzminsky, E., Muleo, R., Scarascia Mugnozza, G., and Sabatti, M. (2010). Genetic linkage maps of *Populus alba* L. and comparitive mapping analysis of sex determination across *Populus* species. Tree Genetics and Genomes *6*, 1-13.

Pelaz, S., Ditta, G.S., Baumann, E., Wisman, E., and Yanofsky, M.F. (2000). B and C floral organ identity functions require *SEPALLATA* MADS-box genes. Nature *405*, 200-203.

Perl-Treves, R., Kahana, A., Rosenman, N., Xiang, Y., and Silberstein, L. (1998). Expression of Multiple *AGAMOUS*-like genes in male and female flowers of cucumber (*Cucumis sativus* L.). Plant and Cell Physiology *39*, 701-712.

Pinyopich, A., Ditta, G.S., Savidge, B., Liljegren, S., J., Baumann, E., Wisman, E., and Yanofsky, M., F. (2003). Assessing the redundancy of MADS-box genes during carpel and ovule development. Nature *424*, 85-88.

Polzin, T., and Daneschmand, S.V. (2003). On Steiner trees and minimum spanning trees in hypergraphs. Operations Research Letters *31*, 12.

Rambaut, A. (1996). Se-Al: Sequence Alignment Editor
Reamon-Büttner, S.M., Schondelmaier, J., and Jung, C. (1998). AFLP markers tightly linked to the sex locus in *Asparagus officinalis* L. Molecular Breeding *4*, 91-98.

Renner, S., S., and Ricklefs, R., E. (1995). Dioecy and its correlates in the flowering plants. American Journal of Botany *82*, 596-606.

Renner, S.S., and Won, H. (2001). Repeated evolution of dioecy from monoecy in Siparunaceae (Laurales). Systematic Biology *50*, 700-712.

Rice, W.R. (1987). Genetic hitchhiking and the evolution of reduced genetic activity of the Y sex chromosome. Genetics *116*, 161-167.

Rood, S., B., and Polzin, M.L. (2003). Big old cottonwoods. Canadian Journal of Botany *81*, 764-767.

Sakamoto, K., Abe, T., Matsuyama, T., Yoshida, S., Ohmido, N., Fukui, K., and Satoh, S. (2005). RAPD markers encoding retrotransposable elements are linked to the male sex in *Cannabis sativa* L. Genome *48*, 931-936.

Sakamoto, K., Ohmido, N., Fukui, K., Kamada, H., and Satoh, S. (2000). Site-specific accumulation of a LINE-like retrotransposon in a sex chromosome of the dioecious plant *Cannabis sativa*. Plant Molecular Biology *44*, 723-732.

Scotti, I., and Delph, L.F. (2006). Selective trade-offs and sex-chromosome evolution in *Silene latifolia* Evolution *60*, 1793-1800.

Seefelder, S., Ehrmaier, H., Schweizer, G., and Seigner, E. (2000). Male and female genetic linkage map of hops, *Humulus lupulus*. Plant Breeding *119*, 249-255.

Semerikov, V., Lagercrantz, U., Tsarouhas, V., Rönnberg-Wästljung, A., Alström-Rapaport, C., and Lascoux, M. (2003). Genetic mapping of sex-linked markers in *Salix viminalis* L. Heredity *91*, 293-299.

Shapiro, H. (2005). Outline of the Assembly Process: JAZZ, the JGI In-House Assembler (Lawrence Berkley National Laboratory).

Sheppard, L.A., Brunner, A.M., Krutovskii, K.V., Rottmann, W.H., Skinner, J.S., Vollmer, S.S., and Strauss, S.H. (2000). A *DEFICIENS* homolog from the dioecious tree black cottonwood is expressed in female and male floral meristems of the two-whorled, unisexual flowers. Plant Physiology *124*, 627-640.

Shibata, F., Hizume, M., and Kuroki, Y. (1999). Chromosome painting and Y chromosomes and isolation of a Y chromosome-specific repetitive sequence in the dioecious plant *Rumex acetosa*. Chromosoma *108*, 266-270.

Skinner, J., S., Meilan, R., Ma, C., and Strauss, S.H. (2003). The *Populus* PTD promoter imparts floral-predominant expression and enables high levels of floral-organ ablation in *Populus*, *Nicotiana* and *Arabidopsis*. Molecular Breeding *12*, 119-132.

Slavov, G.T., Leonardi, S., Adams, W.T., Strauss, S.H., and DiFazio, S.P. (2010). Population substructure in continuous and fragmented stands of *Populus trichocarpa*. Heredity *105*, 348-357.

Smyth, G.K. (2005). Limma: linear models for microarray data. In Bioinformatics and Computational Biology Solutions using R and Bioconductor, R. Gentleman, V. Carey, S. Dudoit, R. Irizarry, and W. Huber, eds. (New York, Springer), pp. 397-420.

Spigler, R.B., Lewers, K.S., Main, D.S., and Ashman, T.-L. (2008). Genetic mapping of sex determination in a wild strawberry, *Fragaria virginiana*, reveals earliest form of sex chromosome. Heredity *101*, 507-527.

Stamatakis, A. (2006). RAxML-VI-HPC: maximum likelihood-based phylogenetic analyses with thousands of taxa and mixed models. Bioinformatics *22*, 2688-2690.

Stellari, G.M., Jaramillo, M.A., and Kramer, E.M. (2004). Evolution of the *APETALA3* and *PISTILLATA* lineages of MADS-box-containing genes in the basal angiosperms. Molecular Biology and Evolution *21*, 506-519.

Sterky, F., Bhalerao, R.R., Unneberg, P., Segerman, B., Nilsson, P., Brunner, A.M., Charbonnel-Campaa, L., Lindvall, J.J., Tandre, K., Strauss, S.H.*, et al.* (2004). A *Populus* EST resource for plant functional genomics. Proceedings of the National Academy of Sciences of the United States of America *101*, 13951-13956.

Stevens, P.F. (2001). Angiosperm Phylogeny Website (Missouri Botanical Gardens, University of Missouri, St. Louis, USA).

Swofford, D.L. (2001). PAUP* 4.0: Phylogenetic Analysis Using Parsimony (* and other methods). (Sinauer Associates).

Tanurdzic, M., and Banks, J.A. (2004). Sex-determining mechanisms in land plants. Plant Cell *16*, S61-71.

Taylor, M., and Semple, C. (2002). Sushi gets serious: the draft genome sequence of the pufferfish *Fugu rubripes*. Genome Biology *3*, reviews1025.1021 - reviews1025.1026.

Telgmann-Rauber, A., Jamsari, A., Kinney, M.S., Pires, C.J., and Jung, C. (2007). Genetic and physical maps around the sex-determining *M*-locus of the dioecious plant asparagus. Molecular Genetics and Genomics *278*, 221-234.

Theißen, G. (2001). Development of floral organ identity: stories from the MADS house. Current Opinion in Plant Biology *4*, 75-85.

Trebitsh, T., Staub, J.E., and O'Neill, S.D. (1997). Identification of a 1-aminocyclopropane-1-carboxylic acid synthase gene linked to the female (F) locus that enhances female sex expression in cucumber. Plant Physiol *113*, 987-995.

Tschaplinski, T.J., Tuskan, G.A., and Gunderson, C.A. (1994). Water-stress tolerance of black and eastern cottonwood clones and four hybrid progeny. I. Growth, water relations, and gas exchange. Canadian Journal of Forest Research *24*, 364-371.

Tuskan, G.A., DiFazio, S., Jansson, S., Bohlmann, J., Grigoriev, I., Hellsten, U., Putnam, N., Ralph, S., Rombauts, S., Salamov, A.*, et al.* (2006). The genome of black cottonwood, *Populus trichocarpa* (Torr. & Gray). Science *313*, 1596-1604.

van den Berg, B.H.J., McCarthy, F.M., Lamont, S.J., and Burgess, S.C. (2010). Re-annotation Is an essential step in systems biology modeling of functional genomics data. PLoS ONE *5*, e10642.

van Doorn, G.S., and Kirkpatrick, M. (2007). Turnover of sex chromosomes induced by sexual conflict. Nature *449*, 909-912.

Vandenbussche, M., Zethof, J., Souer, E., Koes, R., Tornielli, G.B., Pezzotti, M., Ferrario, S., Angenent, G.C., and Gerats, T. (2003). Toward the analysis of the petunia MADS box gene family by reverse and forward transposon insertion mutagenesis approaches: B, C, and D floral organ identity functions require *SEPALLATA*-like MADS box genes in petunia. Plant Cell *15*, 2680-2693.

Veyrunes, F., Waters, P.D., Miethke, P., Rens, W., McMillan, D., Alsop, A.E., Grützner, F., Deakin, J.E., Whittington, C.M., Schatzkamer, K.*, et al.* (2008). Bird-like sex chromosomes of platypus imply recent origin of mammal sex chromosomes. Genome Research *18*, 965-973.

Vuylsteke, M., van Eeuwijk, F., Van Hummelen, P., Kuiper, M., and Zabeau, M. (2005). Genetic analysis of variation in gene expression in *Arabidopsis thaliana*. Genetics *171*, 1267-1275.

Vyskot, B., and Hobza, R. (2004). Gender in plants: sex chromosomes are emerging from the fog. Trends in Genetics *20*, 432-438.

Wang, Z., Gerstein, M., and Snyder, M. (2009). RNA-Seq: a revolutionary tool for transcriptomics. Nat Rev Genet *10*, 57-63.

Wullschleger, S., D., Jansson, S., and Taylor, G. (2002b). Genomics and forest biology: *Populus* emerges as the perennial favorite. The Plant Cell *14*, 2651-2655.

Wullschleger, S., D., Tuskan, G.A., and DiFazio, S.P. (2002a). Genomics and the tree physiologist. Tree Physiology *22*, 1273-1276.

Yamato, K.T., Ishizaki, K., Fujisawa, M., Okada, S., Nakayama, S., Fujishita, M., Bando, H., Yodoya, K., Hayashi, K., Bando, T.*, et al.* (2007). Gene organization of the liverwort Y chromosome reveals distinct sex chromosome evolution in a haploid system. Proceedings of the National Academy of Sciences *104*, 6472-6477.

Yin, T., DiFazio, S.P., Gunter, L.E., Zhang, X., Sewell, M.M., Woolbright, S.A., Allan, G.J., Kelleher, C.T., Douglas, C.J., Wang, M.*, et al.* (2008). Genome structure and emerging evidence of an incipient sex chromosome in *Populus*. Genome Research *18*, 422-430.
Yu, Q., Navajas-Pérez, R., Tong, E., Robertson, J., Moore, P., Paterson, A., and Ming, R. (2008). Recent origin of dioecious and gynodioecious Y chromosomes in papaya. Tropical Plant Biology *1*, 49-57.

Yuceer, C., Land, S.B., Kubiske, M.E., and Harkess, R.L. (2003). Shoot morphogenesis associated with flowering in *Populus deltoides* (Salicaceae). American Journal of Botany *90*, 196-206.

Zhang, B., Zhang, X., Li, H., Zhou, X., and Su, X. (2009). Cloning and expression analysis of an E-class MADS-box gene from *Populus deltoides*. African Journal of Biotechnology *8*, 4789-4796.

Zhang, L.-B., Simmons, M.P., Kocyan, A., and Renner, S.S. (2006). Phylogeny of the Cucurbitales based on DNA sequences of nine loci from three genomes: Implications for morphological and sexual system evolution. Molecular Phylogenetics and Evolution *39*, 305-322.

Zhang, W., Wang, X., Yu, Q., Ming, R., and Jiang, J. (2008). DNA methylation and heterochromatinization in the male-specific region of the primitive Y chromosome of papaya. Genome Research *18*, 1938-1943.

Zik, M., and Irish, V.F. (2003). Flower development: Initiation, differentiation, and diversification. Annual Review of Cell Developmental Biology *19*, 119-140.

Zluvova, J., Nicolas, M., Berger, A., Negrutiu, I., and Monéger, F. (2006). Premature arrest of the male flower meristem precedes sexual dimorphism in the dioecious plant *Silene latifolia*. Proceedings of the National Academy of Sciences *103*, 18854-18859.

Zwickl, D.J. (2006). Genetic algorithm approaches for the phylogenetic analysis of large biological sequence datasets under the maximum likelihood criterion. In The University of Texas at Austin (Austin, The University of Texas), pp. 125.

Appendices

Appendix A List of PCR primers designed to amplify genes studied in chapters three, four and five

Gene name	Organism and genome	Primer name	Length	Tm	Primer sequence	Product size
eugene3.01170047	*P.trichocarpa* V 1.1 Genome	SCA117_1_1F	20	60.14	TGAAGAGCTCGAGACCCAGT	211
eugene3.01170047	*P.trichocarpa* V 1.1 Genome	SCA117_1_1R	20	60.04	TGCCTTGTCGATCCCTATTC	211
fgenesh4_pg.C_scaffold_117000045	*P.trichocarpa* V 1.1 Genome	SCA117_2_1F	20	59.84	TGTGCCGAAGAAAATGACAG	243
fgenesh4_pg.C_scaffold_117000045	*P.trichocarpa* V 1.1 Genome	SCA117_2_1R	20	59.73	GAGCGTTGCATGTGAATGTT	243
grail3.0117003001	*P.trichocarpa* V 1.1 Genome	SCA117_3_1F	20	60.22	AAGCATCAAATGGCGAAGAC	238
grail3.0117003001	*P.trichocarpa* V 1.1 Genome	SCA117_3_1R	20	59.94	ACCCTTGCCTCAATTCCTTT	238
gw1.117.122.1	*P.trichocarpa* V 1.1 Genome	SCA117_4_1F	20	60.09	TGGAGGCAACTTGTTTTTCC	155
gw1.117.122.1	*P.trichocarpa* V 1.1 Genome	SCA117_4_1R	20	59.86	TCCTCAAATCCCATCCAGAC	189

Gene name	Organism and genome	Primer name	Length	Tm	Primer sequence	Product size
gw1.117.122.1	*P.trichocarpa* V 1.1 Genome	SCA117_4_2R	20	60.33	CTTACCCATCCGCAACATTC	208
fgenesh4_pg.C_scaffold_117000051	*P.trichocarpa* V 1.1 Genome	SCA117_5_1F	20	59.62	ACAGTGGGCATCACAGTACG	150
fgenesh4_pg.C_scaffold_117000051	*P.trichocarpa* V 1.1 Genome	SCA117_5_1R	20	59.98	GCTATGGTGGCGAGAGAAAG	150
fgenesh4_pg.C_scaffold_117000051	*P.trichocarpa* V 1.1 Genome	SCA117_5_2R	20	59.98	AGCTATGGTGGCGAGAGAAA	151
fgenesh4_pg.C_scaffold_117000053	*P.trichocarpa* V 1.1 Genome	SCA117_6_1F	20	60.01	TTTGATGGCCTGAATGATGA	153
fgenesh4_pg.C_scaffold_117000053	*P.trichocarpa* V 1.1 Genome	SCA117_6_1R	20	60.17	CTCCCCATATCTCGAAGCAA	212
fgenesh4_pg.C_scaffold_117000053	*P.trichocarpa* V 1.1 Genome	SCA117_6_2F	20	59.62	GAAGATTGCCGCTAGCACTT	151
fgenesh4_pg.C_scaffold_117000053	*P.trichocarpa* V 1.1 Genome	SCA117_6_2R	20	60.01	TCATCATTCAGGCCATCAAA	151
fgenesh4_pg.C_scaffold_117000054	*P.trichocarpa* V 1.1 Genome	SCA117_7_1F	20	60.07	ATCTCGGACTCCCGACTTTT	211
fgenesh4_pg.C_scaffold_117000054	*P.trichocarpa* V 1.1 Genome	SCA117_7_1R	20	60.13	TTTTGAGCAGCCTGAGAGGT	211

Gene name	Organism and genome	Primer name	Length	Tm	Primer sequence	Product size
e_gw1.117.150.1	*P.trichocarpa* V 1.1 Genome	SCA117_8_1F	20	60.14	AACATCAGGCGTGTGCATTA	177
e_gw1.117.150.1	*P.trichocarpa* V 1.1 Genome	SCA117_8_1R	20	60.12	ACCTTTGCCGAGTCAACATC	177
e_gw1.117.150.1	*P.trichocarpa* V 1.1 Genome	SCA117_8_2R	20	59.81	TTTGAAGGCATGAGATGTGC	247
gw1.117.169.1	*P.trichocarpa* V 1.1 Genome	SCA117_9_1F	20	59.99	TGAGGCTTTTGTGGAGCTTT	150
gw1.117.169.1	*P.trichocarpa* V 1.1 Genome	SCA117_9_1R	20	59.67	CAAAACAGGGCATCTCTTCC	150
gw1.117.169.1	*P.trichocarpa* V 1.1 Genome	SCA117_9_2F	20	60.30	TGGAGCGGAAACACTACACA	214
gw1.117.169.1	*P.trichocarpa* V 1.1 Genome	SCA117_9_2R	20	60.07	CAGGGCATCTCTTCCACAAT	214
eugene3.01170072	*P.trichocarpa* V 1.1 Genome	SCA117_10_1F	20	60.30	AAAGAATGGAGGGGGACTTG	242
eugene3.01170072	*P.trichocarpa* V 1.1 Genome	SCA117_10_1R	20	60.49	AGCCTTGGGGACTAAGGTTG	242
fgenesh4_pg.C_scaffold_117000045	*P.trichocarpa* V 1.1 Genome	SCA117_2WG_1F	20	60.07	AGGCCTTCCCTCATGAAAGT	853

Gene name	Organism and genome	Primer name	Length	Tm	Primer sequence	Product size
fgenesh4_pg.C_scaffold_117000045	*P.trichocarpa* V 1.1 Genome	SCA117_2WG_1R	20	59.94	CGAGATTTTGCCAACCATTT	853
fgenesh4_pg.C_scaffold_117000045	*P.trichocarpa* V 1.1 Genome	SCA117_2WG_2F	20	59.85	CATGGCTGCATGATAGCACT	922
fgenesh4_pg.C_scaffold_117000045	*P.trichocarpa* V 1.1 Genome	SCA117_2WG_2R	20	59.94	CGAGATTTTGCCAACCATTT	922
gw1.117.122.1	*P.trichocarpa* V 1.1 Genome	SCA117_4WG_1F	20	60.07	CATGGAGGATGGCTTTGACT	980
gw1.117.122.1	*P.trichocarpa* V 1.1 Genome	SCA117_4WG_1R	20	60.17	CTCCCCATATCTCGAAGCAA	980
fgenesh4_pg.C_scaffold_117000051	*P.trichocarpa* V 1.1 Genome	SCA117_5WG_1F	21	59.92	TGCTATCAACAAGCACAGTGG	852
fgenesh4_pg.C_scaffold_117000051	*P.trichocarpa* V 1.1 Genome	SCA117_5WG_1R	22	57.81	AATAGCTTTGTCGTACGCTTTC	852
gw1.117.169.1	*P.trichocarpa* V 1.1 Genome	SCA117_9WG_1F	20	60.15	GGGTAGGGATCACGGAGATT	559
gw1.117.169.1	*P.trichocarpa* V 1.1 Genome	SCA117_9WG_1R	20	59.67	AAAACAGGGCATCTCTTCCA	559
eugene3.01170072	*P.trichocarpa* V 1.1 Genome	SCA117_10WG_1F	20	59.80	TTCTCATCAAGGGAGGCACT	853

Gene name	Organism and genome	Primer name	Length	Tm	Primer sequence	Product size
eugene3.01170072	*P.trichocarpa* V 1.1 Genome	SCA117_10WG_1 R	20	60.49	AGCCTTGGGGACTAAGGTTG	853
eugene3.01170072	*P.trichocarpa* V 1.1 Genome	SCA117_10WG_2 F	20	59.80	TTCTCATCAAGGGAGGCACT	897
eugene3.01170072	*P.trichocarpa* V 1.1 Genome	SCA117_10WG_2 R	21	60.26	TCAGTCTTGGGCAACTTGTGC	897
gw1.117.255.1	*P.trichocarpa* V 1.1 Genome	SCA117_A_1F	20	60.07	ATGGGCTCCTCCTTCTTGTT	935
gw1.117.255.1	*P.trichocarpa* V 1.1 Genome	SCA117_A_1R	20	59.99	GGATCACCATGGAAATTTGG	935
eugene3.01170002	*P.trichocarpa* V 1.1 Genome	SCA117_B_1F	27	57.93	TCAGGATATGATGTATTATATG GACAA	705
eugene3.01170002	*P.trichocarpa* V 1.1 Genome	SCA117_B_1R	22	60.12	TTCATTGGTGACATATGGTCGT	705
fgenesh4_pg.C_scaffold_117000010	*P.trichocarpa* V 1.1 Genome	SCA117_C_1F	21	59.64	TTATACAGGCAATGTGCAACG	619
fgenesh4_pg.C_scaffold_117000010	*P.trichocarpa* V 1.1 Genome	SCA117_C_1R	24	59.67	TTCTCACCATCAAGTCTTAAAT CG	619
fgenesh4_pg.C_scaffold_117000025	*P.trichocarpa* V 1.1 Genome	SCA117_D_1F	20	59.52	CTAAAGGTCGCTGCAAGGTT	634

Gene name	Organism and genome	Primer name	Length	Tm	Primer sequence	Product size
fgenesh4_pg.C_scaffold_117000025	*P.trichocarpa* V 1.1 Genome	SCA117_D_1R	20	59.59	AAAAGGTGTTTGGGGGTTTT	634
eugene3.01170041	*P.trichocarpa* V 1.1 Genome	SCA117_E_1F	24	59.43	CTTTTGGAAACTATGGATCAAACA	617
eugene3.01170041	*P.trichocarpa* V 1.1 Genome	SCA117_E_1R	25	60.25	GAATGAAAATACATGCTTCTCCAAC	617
eugene3.01170064	*P.trichocarpa* V 1.1 Genome	SCA117_F_1F	20	60.03	TTCTTCCGGAGGCCTTAAAT	925
eugene3.01170064	*P.trichocarpa* V 1.1 Genome	SCA117_F_1R	20	60.03	TTGCCTCCTCCCTATTCCTT	925
gw1.117.220.1	*P.trichocarpa* V 1.1 Genome	SCA117_G_1F	20	59.56	CCTTGGAGAAGGTGGTTTTG	867
gw1.117.220.1	*P.trichocarpa* V 1.1 Genome	SCA117_G_1R	20	60.12	ATTGGTCCACAGGAGCAGTC	867
gw1.117.235.1	*P.trichocarpa* V 1.1 Genome	SCA117_H_1F	20	59.85	GACAACACCCACATCGACAC	853
gw1.117.235.1	*P.trichocarpa* V 1.1 Genome	SCA117_H_1R	20	59.99	TGCTGGAGACTGATTTGTGC	853
e_gw1.117.150.1	*P.trichocarpa* V 1.1 Genome	SCA117_8WG_1F	20	59.88	AACCCAGTACACCCAAGTCG	957

Gene name	Organism and genome	Primer name	Length	Tm	Primer sequence	Product size
e_gw1.117.150.1	*P.trichocarpa* V 1.1 Genome	SCA117_8WG_1R	20	59.83	CATGAGCGAGCCATTGTAAA	957
POPTR_0019s01700	*P.trichocarpa* V 2.0 Genome	SCA19_1_1F	20	59.97	AAATATTGGTGCCAGCGTTC	633
POPTR_0019s01700	*P.trichocarpa* V 2.0 Genome	SCA19_1_1R	20	60.30	GCTGGCTGGTGTTGTTTCTT	633
POPTR_0019s01740	*P.trichocarpa* V 2.0 Genome	SCA19_2_1F	20	60.05	CACCGCAAAGCACTGTTAGA	970
POPTR_0019s01740	*P.trichocarpa* V 2.0 Genome	SCA19_2_1R	20	60.05	GTCCAGAAATCCTGCCAAAA	970
POPTR_0019s01780	*P.trichocarpa* V 2.0 Genome	SCA19_3_1F	20	59.96	AGGGGTTCGAGCTTTGGTAT	863
POPTR_0019s01780	*P.trichocarpa* V 2.0 Genome	SCA19_3_1R	20	60.04	CCTGACATCTTTTCCCCTGA	863
POPTR_0019s01830	*P.trichocarpa* V 2.0 Genome	SCA19_4_1F	20	59.98	CAGATCAGAGGCTGGTGTCA	957

Gene name	Organism and genome	Primer name	Length	Tm	Primer sequence	Product size
POPTR_0019s01830	*P.trichocarpa* V 2.0 Genome	SCA19_4_1R	20	59.91	AACCTGGCTGTGCAGAAAGT	957
POPTR_0019s01850	*P.trichocarpa* V 2.0 Genome	SCA19_5_1F	20	59.99	AGTCAAAGCTCCGAAGACCA	942
POPTR_0019s01850	*P.trichocarpa* V 2.0 Genome	SCA19_5_1R	20	59.85	CTGGACTCATAGCACGACCA	942
POPTR_0019s01870	*P.trichocarpa* V 2.0 Genome	SCA19_6_1F	20	59.55	GTTCAGTTGGATGTGCCAAA	979
POPTR_0019s01870	*P.trichocarpa* V 2.0 Genome	SCA19_6_1R	20	59.78	CCCAGATAACCTTTCCACCA	979
POPTR_0019s13010	*P.trichocarpa* V 2.0 Genome	SCA19_7_1F	20	60.42	AAGCCAAACGAAGCAACAAG	939
POPTR_0019s13010	*P.trichocarpa* V 2.0 Genome	SCA19_7_1R	20	59.43	ATGACGGAACCTCGGTTAAA	939

Gene name	Organism and genome	Primer name	Length	Tm	Primer sequence	Product size
POPTR_0019s03090	*P.trichocarpa* V 2.0 Genome	SCA_19_9_1F	20	60.05	GGAATGGAAGTTCGAGTGGA	912
POPTR_0019s03090	*P.trichocarpa* V 2.0 Genome	SCA_19_9_1R	20	59.70	CATACCTGCCACTCTGGTCA	912
POPTR_0019s03100	*P.trichocarpa* V 2.0 Genome	SCA_19_10_1F	20	59.97	CAAGGAATGGGTGTTCGAGT	899
POPTR_0019s03100	*P.trichocarpa* V 2.0 Genome	SCA_19_10_1R	20	60.04	TCTTAAGATGCCCCATTTGC	899
POPTR_0019s12360	*P.trichocarpa* V 2.0 Genome	SCA_19_11_1F	20	60.05	GGCCCAAACCAAACCTTATT	992
POPTR_0019s12360	*P.trichocarpa* V 2.0 Genome	SCA_19_11_1R	20	60.49	ACCACTGGGTATCCCACCTT	992
POPTR_0019s10580	*P.trichocarpa* V 2.0 Genome	SCA_19_12_1F	20	60.20	GCAGACATGGGAAGAGGAAA	722

Gene name	Organism and genome	Primer name	Length	Tm	Primer sequence	Product size
POPTR_0019s10580	*P.trichocarpa* V 2.0 Genome	SCA_19_12_1R	20	59.50	CAAAATACCACTTGCACACCA	722
POPTR_0002s07920	*P.trichocarpa* V 2.0 Genome	MADS_1_1F	23	60.03	CCTCACACCTTCACTTTGAAATC	879
POPTR_0002s07920	*P.trichocarpa* V 2.0 Genome	MADS_1_1R	20	60.11	TTTCACGGGCTAAGTGGTTC	879
POPTR_0005s20480	*P.trichocarpa* V 2.0 Genome	MADS_2_1F	21	58.63	CCTCCTCACACCTTCAATTTC	874
POPTR_0005s20480	*P.trichocarpa* V 2.0 Genome	MADS_2_1R	20	59.58	TAGGGTTTTGAATGCACACG	874

Appendix B List of gene names, genomic locations, and functional annotations for genes studied in chapters two, three and four

Working name for gene	Poplar genome version 1.1 gene name	Poplar genome version 2.0 gene name	Location on version 2.0 poplar genome	Function annotation for gene
Ssu72-like	fgenesh1_pg.C_LG_XV 000380	POPTR_0015s04700	scaffold_15: 4923219 - 4926558	DEAD/DEAH box helicase
N/A	eugene3.00660277	Overlaps with 2 genes on scaffold_3:POPTR_0003s03450 and POPTR_0003s03440	POPTR_0003s03450: scaffold_3: 3897472 - 3898473 and POPTR_0003s03440:scaffold_3: 3882517 - 3883355	There are no functional annotations for these two loci
N/A	fgenesh4_pg.C_scaffold_277 000004	Overlaps with 4 genes on scaffold_17: POPTR_0017s13140, POPTR_0017s13180, POPTR_0017s13220, POPTR_0017s13260	POPTR_0017s13140:scaffold_17: 13300328 - 13301186 POPTR_0017s13180: scaffold_17: 13358880 - 13359781 POPTR_0017s13220: scaffold_17: 13422659 - 13423406 POPTR_0017s13260: scaffold_17: 13439842 - 13440015	POPTR_0017s13140: Rapid ALkalinization Factor (RALF) POPTR_0017s13180: There are no functional annotations for this locus POPTR_0017s13220: Rapid ALkalinization Factor (RALF) POPTR_0017s13260: There are no functional annotations for this locus
SCA 117 A	gw1.117.255.1	POPTR_0019s01120	scaffold_19: 849164 - 851761	Protein tyrosine kinase
SCA 117 B	eugene3.01170002	Sequence not included in V 2.0	N/A	N/A

Working name for gene	Poplar genome version 1.1 gene name	Poplar genome version 2.0 gene name	Location on version 2.0 poplar genome	Function annotation for gene
SCA 117 C	fgenesh4_pg.C_scaffold_117000010	POPTR_0019s01190 POPTR_0019s02960	POPTR_0019s01190 sca_19:914104-917075 POPTR_0019s2960 scaffold_19: 2875642 - 2876509	POPTR_0019s01190 and POPTR_0019s02960: There are no functional annotations for these gene
SCA 117 D	fgenesh4_pg.C_scaffold_117000025	POPTR_0019s01270	scaffold_19: 1075484 - 1078988	Terpene synthase family, metal binding domain
SCA 117 E	eugene3.01170041	POPTR_0019s01340	scaffold_19: 1177994 - 1181330	terpene synthase/cyclase family protein; similar to myrcene/ocimene synthase
SCA 117 1	eugene3.01170047	POPTR_0019s01520	POPTR_0019s01520 = scaffold_19: 1339019 - 1340779	Transferase family
SCA 117 2	fgenesh4_pg.C_scaffold_117000045	POPTR_0019s01530	POPTR_0019s01530 = scaffold_19: 1347076 - 1348225	there are no functional annotations for this gene
SCA 117 3	grail3.0117003001	POPTR_0019s01540	POPTR_0019s01540 = scaffold_19: 1348343 - 1349893	Transferase family
SCA 117 4	gw1.117.122.1	POPTR_0019s01560	scaffold_19: 1379318 - 1382368	Leucine Rich Repeat
SCA 117 5	fgenesh4_pg.C_scaffold_117000051	POPTR_0019s01570	POPTR_0019s01570 = scaffold_19: 1391060 - 1395729	Apoptotic ATPase, protein binding
SCA 117 6	fgenesh4_pg.C_scaffold_117000053	Overlaps with 2 genes on scaffold_19: POPTR_0019s01560 POPTR_0019s01570	POPTR_0019s01560 = scaffold_19: 1379318 - 1382368 POPTR_0019s01570 = scaffold_19: 1391060 - 1395729	Apoptotic ATPase
SCA 117 7	fgenesh4_pg.C_scaffold_117000054	Overlaps with 2 genes on scaffold_19: POPTR_0019s01560 POPTR_0019s01570	POPTR_0019s01560 = scaffold_19: 1379318 - 1382368 POPTR_0019s01570 = scaffold_19: 1391060 - 1395729	Apoptotic ATPase

Working name for gene	Poplar genome version 1.1 gene name	Poplar genome version 2.0 gene name	Location on version 2.0 poplar genome	Function annotation for gene
SCA 117 8	e_gw1.117.150.1	POPTR_0019s01630	POPTR_0019s01630 = scaffold_19: 1438061 - 1439700	Alginate lyase
SCA 117 F	eugene3.01170064	Overlaps with 2 genes on scaffold_19: POPTR_0019s01660 and POPTR_0019s01670	POPTR_0019s01660 = scaffold_19: 1460825 - 1462203 POPTR_0019s01670 = scaffold_19:scaffold_19: 1463832 - 1466771	Apoptotic ATPase
SCA 117 9	gw1.117.169.1	Between 2 genes on scaffold_19:POPTR_0019s01660 and POPTR_0019s01670	POPTR_0019s01660 = scaffold_19: 1460825 - 1462203 POPTR_0019s01670 = scaffold_19:scaffold_19: 1463832 - 1466771	Apoptotic ATPase
SCA 117 10	eugene3.01170072	POPTR_0019s01790	scaffold_19: 1546894 - 1550419	Transcription elongation factor
SCA 117 G	gw1.117.220.1	POPTR_0019s01880	scaffold_19: 1630220 - 1633020	Serine/threonine protein kinase
SCA 117 H	gw1.117.235.1	Oversaps with 2 genes on scaffold_19: POPTR_0019s01940 and POPTR_0019s01950	scaffold_19: 1728118 - 1733133	Serine/threonine protein kinase
SCA 19 11	eugene3.00190854	POPTR_0019s12360	scaffold_19: 13401257 - 13403527	Chitin recognition protein
SCA 19 12	eugene3.00190689	POPTR_0019s10580	scaffold_19: 12132258 - 12133799	K-box region, regulation of transcription, DNA-dependent

Appendix C List of gene names, genomic locations, and functional annotations for genes studied in chapter five

Working name for gene	Poplar genome version 2.0 gene name	Location on version 2.0 poplar genome	Function annotation for gene
SCA19_12	POPTR_0019s10580	scaffold_19: 12132258 - 12133799	K-box region, regulation of transcription, DNA-dependent
F2	POPTR_0013s09980	scaffold_13: 9955912 - 9966253	MADS box transcription factor
F3	POPTR_0011s03140	scaffold_11: 3200188 - 3207704	MADS-box transcription factor
F4	POPTR_0008s09800	scaffold_8: 6076113 - 6078596	MADS-box protein regulation of transcription, DNA-dependent
PTAP2 and PTM1	POPTR_0010s16380	scaffold_10: 15345650 - 15351172	MADS-box transcription factor
PTAP2 and PTM1	POPTR_0010s16380	scaffold_10: 15345650 - 15351172	MADS-box transcription factor
PTD	POPTR_0007s13660	scaffold_7: 13686094 - 13687851	SRF-type transcription factor (DNA-binding and dimerisation domain)
PTAG1	POPTR_0004s06300	scaffold_4: 4997208 - 5005680	SRF-type transcription factor (DNA-binding and dimerisation domain)
M1	POPTR_0004s11420	scaffold_4: 10038841 - 10040605	MADS-box protein
M2	POPTR_0012s05960	scaffold_12: 5936482 - 5942087	MADS box transcription factor
M4	POPTR_0005s12000	scaffold_5: 8770246 - 8772014	MADS-box protein involved in regulation of transcription, DNA-dependent
M5	POPTR_0002s02910	scaffold_2: 1789707 - 1795019	Kinesin-like protein involved in microtubule motor activity
M6	POPTR_0017s13410	scaffold_17: 13578276 - 13585496	MADS box transcription factor

Working name for gene	Poplar genome version 2.0 gene name	Location on version 2.0 poplar genome	Function annotation for gene
MADS_1	POPTR_0002s07920	scaffold_2: 5450756 - 5453302	MADS box transcription factor
MADS_2	POPTR_0005s20480	scaffold_5: 19532615 - 19535235	MADS box transcription factor
PtMADS31	POPTR_0008s09790	scaffold_8: 6063858 - 6070604	MADS box transcription factor
PTM3 and PTM4	POPTR_0004s11440	scaffold_4: 10100959 - 10108181	MADS box transcription factor
PTM5	POPTR_0014s07010	scaffold_14: 5281611 - 5290965	MADS box transcription factor
PTM6	POPTR_0001s13650	scaffold_1: 10729860 - 10732842	Regulation of transcription, DNA-dependent

Appendix D List of all MADS-box genes reported in Leseberg et al. (2006) updated to version 2.0 of the poplar genome

Leseberg name	Locus Name	Location	Research on this gene	Functional annotations for this locus
PtMADS1	POPTR_0001s08500 and POPTR_0001s08510	POPTR_0001s08500=scaffold_1: 6475276 - 6476293 and POPTR_0001s08510=scaffold_1: 6478981 - 6480482	BLAST hit to 2 genes on version 2.0 of the poplar genome.	POPTR_0001s08500=Pfam:01486 K-box region Panther:11945 MADS BOX PROTEIN POPTR_0001s08510=Pfam:00319 SRF-type transcription factor (DNA-binding and dimerisation domain) Panther:11945 MADS BOX PROTEIN
PtMADS2	POPTR_0001s33600	scaffold_1: 31831353 - 31831577		Pfam:00319 SRF-type transcription factor (DNA-binding and dimerisation domain) Panther:11945 MADS BOX PROTEIN
PtMADS3 + 4	POPTR_0001s29100	scaffold_1: 27887506 - 27887724	These two genes are annotated as a single gene on version 2.0 of the genome.	Pfam:00319 SRF-type transcription factor (DNA-binding and dimerisation domain) Panther:11945 MADS BOX PROTEIN
PtMADS5	POPTR_0001s13660	scaffold_1: 10747910 - 10748155		Pfam:00319 SRF-type transcription factor (DNA-binding and dimerisation domain) Panther:11945 MADS BOX PROTEIN
PtMADS6	POPTR_0001s13650	scaffold_1: 10729860 - 10732842	*P. trichocarpa* homolog for PTM6 SEP-class gene in *Populus tremuloides,* Cseke et al 2005	Pfam:01486 K-box region Panther:11945 MADS BOX PROTEIN
PtMADS7	POPTR_0002s10580	scaffold_2: 7672322 - 7674041		Pfam:01486 K-box region Panther:11945 MADS BOX PROTEIN
PtMADS8	POPTR_0002s11030	scaffold_2: 8048500 - 8050136		Pfam:01486 K-box region Panther:11945 MADS BOX PROTEIN
PtMADS9	POPTR_0002s15300	scaffold_2: 11431129 - 11433605		Pfam:01486 K-box region Panther:11945 MADS BOX PROTEIN

Leseberg name	Locus Name	Location	Research on this gene	Functional annotations for this locus
PtMADS10	POPTR_0002s02990	scaffold_2: 1842523 - 1843856		Pfam:01486 K-box regionPfam:00319 SRF-type transcription factor (DNA-binding and dimerisation domain)Panther:11945 MADS BOX PROTEINKEGGORTH:09264 MADS-box transcription factor, plant
PtMADS11	POPTR_0002s07920	scaffold_2: 5450756 - 5453302	MADS_1 MADS-box gene that this research has shown to be only expressed in male floral tissue, not in female flowers tissue, or leaf tissue from either gender.	Pfam:01486 K-box region Pfam:00319 SRF-type transcription factor (DNA-binding and dimerisation domain) Panther:11945 MADS BOX PROTEIN KOG:0014 MADS box transcription factor KEGGORTH:09264 MADS-box transcription factor, plant
PtMADS12	POPTR_0003s11960	scaffold_3: 12666516 - 12672466		Pfam:01486 K-box region Pfam:00319 SRF-type transcription factor (DNA-binding and dimerisation domain) Panther:11945 MADS BOX PROTEIN KOG:0014 MADS box transcription factor
PtMADS13	POPTR_0003s16800 and POPTR_0003s16810	POPTR_0003s16800 = scaffold_3: 16329000 - 16330057 POPTR_0003s16810 = scaffold_3: 16335820 - 16337536	BLAST hit to 2 genes on version 2.0 of the poplar genome.	POPTR_0003s16800 = Pfam:00319 SRF-type transcription factor (DNA-binding and dimerisation domain) Panther:11945 MADS BOX PROTEIN POPTR_0003s16810 = Pfam:01486 K-box region Panther:11945 MADS BOX PROTEIN
PtMADS14 and 15	POPTR_0003s16850	scaffold_3: 16367370 - 16368102	These two genes are annotated as a single gene on version 2.0 of the genome.	Pfam:00319 SRF-type transcription factor (DNA-binding and dimerisation domain)Panther:11945 MADS BOX PROTEIN
PtMADS16	POPTR_0004s11430	scaffold_4: 10049087 - 10049726		Pfam:00319 SRF-type transcription factor (DNA-binding and dimerisation domain) Panther:11945 MADS BOX PROTEIN

Leseberg name	Locus Name	Location	Research on this gene	Functional annotations for this locus
PtMADS17	POPTR_0004s11440	scaffold_4: 10100959 - 10108181		Pfam:01486 K-box region Pfam:00319 SRF-type transcription factor (DNA-binding and dimerisation domain) Panther:11945 MADS BOX PROTEIN KOG:0014 MADS box transcription factor
PtMADS18	POPTR_0004s06300	scaffold_4: 4997208 - 5005680	PTAG1 (Brunner et al., 2000)	Pfam:01486 K-box region Pfam:00319 SRF-type transcription factor (DNA-binding and dimerisation domain) Panther:11945 MADS BOX PROTEIN KOG:0014 MADS box transcription factor KEGGORTH:09264 MADS-box transcription factor, plant
PtMADS19	POPTR_0009s08270	scaffold_9: 7613164 - 7616576		Pfam:01486 K-box region Panther:11945 MADS BOX PROTEIN KEGGORTH:09264 MADS-box transcription factor, plant
PtMADS20	POPTR_0009s06060	scaffold_9: 5973865 - 5977200		Pfam:01486 K-box region Pfam:00319 SRF-type transcription factor (DNA-binding and dimerisation domain) Panther:11945 MADS BOX PROTEIN KOG:0014 MADS box transcription factor
PtMADS21	POPTR_0005s17950	scaffold_5: 15803268 - 15804499		Pfam:01486 K-box region Panther:11945 MADS BOX PROTEIN
PtMADS22	POPTR_0005s20480	scaffold_5: 19532615 - 19535235	MADS_2 MADS-box gene that this research has shown to be only expressed in male floral tissue, not in female flowers tissue, or leaf tissue from either gender.	Pfam:01486 K-box regionPfam:00319 SRF-type transcription factor (DNA-binding and dimerisation domain)Panther:11945 MADS BOX PROTEINKOG:0014 MADS box transcription factorKEGGORTH:09264 MADS-box transcription factor, plant
PtMADS23	POPTR_0005s19420	scaffold_5: 18264598 - 18266213		Pfam:01486 K-box region Panther:11945 MADS BOX PROTEIN

Leseberg name	Locus Name	Location	Research on this gene	Functional annotations for this locus
PtMADS24	POPTR_0006s04730	scaffold_6: 3266443 - 3269027		Pfam:01486 K-box region Pfam:00319 SRF-type transcription factor (DNA-binding and dimerisation domain) Panther:11945 MADS BOX PROTEIN KOG:0014 MADS box transcription factor
PtMADS25	POPTR_0007s13660	scaffold_7: 13686094 - 13687851	F6	Pfam:00319 SRF-type transcription factor (DNA-binding and dimerisation domain) Panther:11945 MADS BOX PROTEIN
PtMADS26	POPTR_0007s14310	scaffold_7: 14130991 - 14137078		Pfam:01486 K-box region Pfam:00319 SRF-type transcription factor (DNA-binding and dimerisation domain) Panther:11945 MADS BOX PROTEIN KOG:0014 MADS box transcription factor
PtMADS27 and 28 and 29	POPTR_0007s03280	scaffold_7: 1888163 - 1889185	These three genes are annotated as a single gene on version 2.0 of the genome.	Panther:11945 MADS BOX PROTEIN
PtMADS30	POPTR_0007s07620	scaffold_7: 5972870 - 5975044		Pfam:01486 K-box regionPfam:00319 SRF-type transcription factor (DNA-binding and dimerisation domain)Panther:11945 MADS BOX PROTEINKOG:0014 MADS box transcription factor
PtMADS31	POPTR_0008s09790	scaffold_8: 6063858 - 6070604	Zhang et al, 2009., in *P.deltoides*, found that homolog to this gene is involved in foral development in males and females - SEP type expression, E-class gene.	Pfam:01486 K-box region Pfam:00319 SRF-type transcription factor (DNA-binding and dimerisation domain) Panther:11945 MADS BOX PROTEIN KOG:0014 MADS box transcription factor

Leseberg name	Locus Name	Location	Research on this gene	Functional annotations for this locus
PtMADS32	POPTR_0008s09800	scaffold_8: 6073641 - 6078620	PTM2 and PTAP1 sequence blasts to this sequence	Pfam:01486 K-box region Pfam:00319 SRF-type transcription factor (DNA-binding and dimerisation domain) Panther:11945 MADS BOX PROTEIN KOG:0014 MADS box transcription factor KEGGORTH:09264 MADS-box transcription factor, plant GO:0006355 regulation of transcription, DNA-dependent
PtMADS33	POPTR_0010s16380	scaffold_10: 15345650 - 15351172	PTAP2 (Brunner et al., 2004) and PTM1 (Cseke et al, 2005) are same gene, and BLAST to this gene so it shows both male and femlae biased expression.	Pfam:01486 K-box regionPfam:00319 SRF-type transcription factor (DNA-binding and dimerisation domain)Panther:11945 MADS BOX PROTEINKOG:0014 MADS box transcription factorKEGGORTH:09264 MADS-box transcription factor, plant
PtMADS34	POPTR_0011s03140	scaffold_11: 3200188 - 3207704	PTAG2 from Brunner et al., 2000	Pfam:01486 K-box region Pfam:00319 SRF-type transcription factor (DNA-binding and dimerisation domain) Panther:11945 MADS BOX PROTEIN KOG:0014 MADS box transcription factor KEGGORTH:09264 MADS-box transcription factor, plant
PtMADS35 and 36	POPTR_0012s10190	scaffold_12: 11018129 - 11024115	These two genes are annotated as a single gene on version 2.0 of the genome.	Pfam:01486 K-box region Pfam:00319 SRF-type transcription factor (DNA-binding and dimerisation domain) Panther:11945 MADS BOX PROTEIN KOG:0014 MADS box transcription factor
PtMADS37	POPTR_0012s14020	scaffold_12: 13998081 - 14005758		Pfam:01486 K-box region Pfam:00319 SRF-type transcription factor (DNA-binding and dimerisation domain) Panther:11945 MADS BOX PROTEIN KOG:0014 MADS box transcription factor

Leseberg name	Locus Name	Location	Research on this gene	Functional annotations for this locus
PtMADS38	POPTR_0012s14770	scaffold_12: 14522968 - 14525400		Pfam:01486 K-box regionPfam:00319 SRF-type transcription factor (DNA-binding and dimerisation domain)Panther:11945 MADS BOX PROTEINKOG:0014 MADS box transcription factor
PtMADS39	POPTR_0012s05960	scaffold_12: 5936482 - 5942087	Male biased expression on the microarray study, no expression data other than this research.	Pfam:01486 K-box region Pfam:00319 SRF-type transcription factor (DNA-binding and dimerisation domain) Panther:11945 MADS BOX PROTEIN KOG:0014 MADS box transcription factor
PtMADS40	POPTR_0014s07000	scaffold_14: 5271849 - 5273587		Pfam:01486 K-box region Panther:11945 MADS BOX PROTEIN
PtMADS41	POPTR_0014s07010	scaffold_14: 5281611 - 5290965	*P. trichocarpa* homolog for PTM5 SEP-class gene in Populus tremuloides, Cseke et al., 2005	Pfam:01486 K-box region Pfam:00319 SRF-type transcription factor (DNA-binding and dimerisation domain) Panther:11945 MADS BOX PROTEIN KOG:0014 MADS box transcription factor
PtMADS42	POPTR_0019s10540	scaffold_19: 12079388 - 12082794		Pfam:01486 K-box region Pfam:00319 SRF-type transcription factor (DNA-binding and dimerisation domain) Panther:11945 MADS BOX PROTEIN KOG:0014 MADS box transcription factor
PtMADS43	POPTR_0019s10580	scaffold_19: 12126290 - 12134031	SCA19_12 - MADS-box gene on Chromosome 19 that this research has shown to be only expressed in female floral tissue, not in male flowers tissue, or leaf tissue from either gender.	Pfam:01486 K-box regionPfam:00319 SRF-type transcription factor (DNA-binding and dimerisation domain)Panther:11945 MADS BOX PROTEINKOG:0014 MADS box transcription factor

Leseberg name	Locus Name	Location	Research on this gene	Functional annotations for this locus
PtMADS44	POPTR_0015s11040	scaffold_15: 11727470 - 11733274		Pfam:01486 K-box region Pfam:00319 SRF-type transcription factor (DNA-binding and dimerisation domain) Panther:11945 MADS BOX PROTEIN KOG:0014 MADS box transcription factor
PtMADS45	POPTR_0015s14950	scaffold_15: 14342335 - 14344327		Pfam:01486 K-box region Pfam:00319 SRF-type transcription factor (DNA-binding and dimerisation domain) Panther:11945 MADS BOX PROTEIN KOG:0014 MADS box transcription factor
PtMADS46	POPTR_0015s14010	scaffold_15: 13758181 - 13760371		Pfam:01486 K-box region Pfam:00319 SRF-type transcription factor (DNA-binding and dimerisation domain) Panther:11945 MADS BOX PROTEIN KOG:0014 MADS box transcription factor
PtMADS47 and 48	POPTR_0017s07170	scaffold_17: 5557162 - 5558254	These two genes are annotated as a single gene on version 2.0 of the genome.	Pfam:00319 SRF-type transcription factor (DNA-binding and dimerisation domain)Panther:11945 MADS BOX PROTEIN
PtMADS49	POPTR_0017s13390 and POPTR_0017s13410	POPTR_0017s13390 = scaffold_17: 13572863 - 13573949 POPTR_0017s13410 =	Used POPTR_0017s13410 in the phylogeny= M6 Male	POPTR_0017s13390 = there are no functional annotations for this locus POPTR_0017s13410 = Pfam:01486 K-box region Pfam:00319 SRF-type transcription factor (DNA-binding and dimerisation domain) Panther:11945 MADS BOX PROTEIN KOG:0014 MADS box transcription factor
PtMADS50	POPTR_0017s13390	POPTR_0017s13390 = scaffold_17: 13572863 - 13573949	BLAST hit to 2 genes on version 2.0 of the poplar genome.	POPTR_0017s13390 = there are no functional annotations for this locus
PtMADS51	POPTR_0013s09980	scaffold_13: 9955912 - 9966253	Female biased expression on the microarray study, no expression data other than this research.	Pfam:01486 K-box region Pfam:00319 SRF-type transcription factor (DNA-binding and dimerisation domain) Panther:11945 MADS BOX PROTEIN KOG:0014 MADS box transcription factor

Leseberg name	Locus Name	Location	Research on this gene	Functional annotations for this locus
PtMADS52	POPTR_0013s10200	scaffold_13: 10214970 - 10216137		Panther:11945 MADS BOX PROTEIN
PtMADS53	POPTR_0003s16840	scaffold_3: 16351961 - 16352345		Pfam:00319 SRF-type transcription factor (DNA-binding and dimerisation domain) Panther:11945 MADS BOX PROTEIN
PtMADS54 and 55	POPTR_0003s16830	scaffold_3: 16347023 - 16350147	These two genes are annotated as a single gene on version 2.0 of the genome.	Pfam:01486 K-box region Panther:11945 MADS BOX PROTEIN
PtMADS56	POPTR_0001s25850	scaffold_1: 24929361 - 24929969		Pfam:00319 SRF-type transcription factor (DNA-binding and dimerisation domain)Panther:11945 MADS BOX PROTEINKOG:0014 MADS box transcription factor
PtMADS57 and 75	POPTR_0005s14660	scaffold_5: 11665785 - 11666699	These two genes are annotated as a single gene on version 2.0 of the genome.	Pfam:00319 SRF-type transcription factor (DNA-binding and dimerisation domain) Panther:11945 MADS BOX PROTEIN KOG:0014 MADS box transcription factor
PtMADS58	Hits between POPTR_0002s06270 and POPTR_0002s06280		Does not align with reading frame of a gene.	
PtMADS59	POPTR_0004s13590	scaffold_4: 13919101 - 13919616		Pfam:00319 SRF-type transcription factor (DNA-binding and dimerisation domain) Panther:11945 MADS BOX PROTEIN KOG:0014 MADS box transcription factor
PtMADS60	POPTR_0004s13600	scaffold_4: 13937678 - 13938214		Pfam:00319 SRF-type transcription factor (DNA-binding and dimerisation domain) Panther:11945 MADS BOX PROTEIN KOG:0014 MADS box transcription factor
PtMADS61	POPTR_0009s05070	scaffold_9: 5245134 - 5245484		Panther:11945 MADS BOX PROTEIN

Leseberg name	Locus Name	Location	Research on this gene	Functional annotations for this locus
PtMADS62 and 63	POPTR_0009s08740 and POPTR_0009s08750	POPTR_0009s08740=scaffold_9: 7896599 - 7897252 POPTR_0009s08750=scaffold_9: 7900899 - 7901868	These two genes are annotated as to the same two genes on version 2.0 of the genome.	POPTR_0009s08740=Pfam:00319 SRF-type transcription factor (DNA-binding and dimerisation domain) Panther:11945 MADS BOX PROTEIN KOG:0014 MADS box transcription factor POPTR_0009s08750=Pfam:00319 SRF-type transcription factor (DNA-binding and dimerisation domain) Panther:11945 MADS BOX PROTEIN KOG:0014 MADS box transcription factor
PtMADS64	POPTR_0006s27460 and POPTR_0006s27470	POPTR_0006s27460=scaffold_6: 25357649 - 25358757 POPTR_0006s27470=scaffold_6: 25361513 - 25362412	BLAST hit to 2 genes on version 2.0 of the poplar genome.	POPTR_0006s27460= Panther:11945:SF19 MADS BOX PROTEIN POPTR_0006s27470=Pfam:00319 SRF-type transcription factor (DNA-binding and dimerisation domain)Panther:11945 MADS BOX PROTEINKOG:0014 MADS box transcription factor
PtMADS65	POPTR_0007s03410	scaffold_7: 1958989 - 1959761		Pfam:00319 SRF-type transcription factor (DNA-binding and dimerisation domain) Panther:11945 MADS BOX PROTEIN KOG:0014 MADS box transcription factor
PtMADS66	POPTR_0008s04110	scaffold_8: 2294059 - 2294718		Pfam:00319 SRF-type transcription factor (DNA-binding and dimerisation domain) Panther:11945 MADS BOX PROTEIN KOG:0014 MADS box transcription factor
PtMADS67 and PtMADS73	POPTR_0012s11190	scaffold_12: 11727312 - 11728355		Pfam:00319 SRF-type transcription factor (DNA-binding and dimerisation domain) Panther:11945 MADS BOX PROTEIN KOG:0014 MADS box transcription factor
PtMADS68 and PtMADS70	POPTR_0016s12640	scaffold_16: 12022576 - 12023214		Pfam:00319 SRF-type transcription factor (DNA-binding and dimerisation domain) Panther:11945 MADS BOX PROTEIN KOG:0014 MADS box transcription factor

Leseberg name	Locus Name	Location	Research on this gene	Functional annotations for this locus
PtMADS69	POPTR_0017s07460	scaffold_17: 5801036 - 5801934		Pfam:00319 SRF-type transcription factor (DNA-binding and dimerisation domain) Panther:11945 MADS BOX PROTEIN KOG:0014 MADS box transcription factor
PtMADS71	POPTR_0004s00400	scaffold_4: 124160 - 124946		Pfam:00319 SRF-type transcription factor (DNA-binding and dimerisation domain)Panther:11945 MADS BOX PROTEINKOG:0014 MADS box transcription factor
PtMADS72	POPTR_0004s01080	scaffold_4: 542625 - 543137		Pfam:00319 SRF-type transcription factor (DNA-binding and dimerisation domain) Panther:11945 MADS BOX PROTEIN
PtMADS74 and PtMADS77	POPTR_0017s06710	scaffold_17: 5128975 - 5129586 ::	These two genes are annotated as a single gene on version 2.0 of the genome.	Pfam:00319 SRF-type transcription factor (DNA-binding and dimerisation domain) Panther:11945 MADS BOX PROTEIN KOG:0014 MADS box transcription factor
PtMADS76	POPTR_0017s06800 and POPTR_0017s06810	POPTR_0017s06800=scaffold_17: 5188482 - 5189081 POPTR_0017s06810=scaffold_17: 5192102 - 5192713	BLAST hit to 2 genes on version 2.0 of the poplar genome.	POPTR_0017s06800=Pfam:00319 SRF-type transcription factor (DNA-binding and dimerisation domain) Panther:11945 MADS BOX PROTEIN KOG:0014 MADS box transcription factor POPTR_0017s06810=Pfam:00319 SRF-type transcription factor (DNA-binding and dimerisation domain) Panther:11945 MADS BOX PROTEIN KOG:0014 MADS box transcription factor
PtMADS78	POPTR_0016s12650	scaffold_16: 12031661 - 12032191		Pfam:00319 SRF-type transcription factor (DNA-binding and dimerisation domain) Panther:11945 MADS BOX PROTEIN GO:0003700 transcription factor activity GO:0043565 sequence-specific DNA binding GO:0006355 regulation of transcription, DNA-dependent GO:0005634 nucleus

Leseberg name	Locus Name	Location	Research on this gene	Functional annotations for this locus
PtMADS79	POPTR_0001s22220	scaffold_1: 20864473 - 20865483		Pfam:00319 SRF-type transcription factor (DNA-binding and dimerisation domain)Panther:11945 MADS BOX PROTEIN
PtMADS80	POPTR_0001s33740	scaffold_1: 31907047 - 31908123		Pfam:00319 SRF-type transcription factor (DNA-binding and dimerisation domain) Panther:11945 MADS BOX PROTEIN KOG:0014 MADS box transcription factor
PtMADS81	POPTR_0002s25700	scaffold_2: 22772493 - 22773710		Pfam:00319 SRF-type transcription factor (DNA-binding and dimerisation domain) Panther:11945 MADS BOX PROTEIN
PtMADS82	POPTR_0007s03260	scaffold_7: 1869051 - 1877508		Pfam:01486 K-box region Pfam:00319 SRF-type transcription factor (DNA-binding and dimerisation domain) Panther:11945 MADS BOX PROTEIN KOG:0014 MADS box transcription factor
PtMADS83	POPTR_0008s02120	scaffold_8: 1011737 - 1012369		Pfam:00319 SRF-type transcription factor (DNA-binding and dimerisation domain) Panther:11945 MADS BOX PROTEIN KOG:0014 MADS box transcription factor
PtMADS84	POPTR_0010s18860	scaffold_10: 16975875 - 16976760		Pfam:00319 SRF-type transcription factor (DNA-binding and dimerisation domain) Panther:11945 MADS BOX PROTEIN
PtMADS85	POPTR_0010s24540	scaffold_10: 20709023 - 20709655		Pfam:00319 SRF-type transcription factor (DNA-binding and dimerisation domain) Panther:11945 MADS BOX PROTEIN KOG:0014 MADS box transcription factor
PtMADS86	POPTR_0012s11170	scaffold_12: 11722595 - 11723767		Pfam:00319 SRF-type transcription factor (DNA-binding and dimerisation domain)Panther:11945 MADS BOX PROTEINKOG:0014 MADS box transcription factor

Leseberg name	Locus Name	Location	Research on this gene	Functional annotations for this locus
PtMADS87	POPTR_0015s08890	scaffold_15: 10253929 - 10254633		Pfam:00319 SRF-type transcription factor (DNA-binding and dimerisation domain) Panther:11945 MADS BOX PROTEIN
PtMADS88 and PtMADS89	POPTR_0004s23090	scaffold_4: 21882354 - 21883063	These two genes are annotated as a single gene on version 2.0 of the genome.	Pfam:00319 SRF-type transcription factor (DNA-binding and dimerisation domain) Panther:11945 MADS BOX PROTEIN KOG:0014 MADS box transcription factor
PtMADS90		No BLAST hit on version 2.0 of the poplar genome for this gene.		
PtMADS91	POPTR_0002s09290	scaffold_2: 6583977 - 6587230		Pfam:00319 SRF-type transcription factor (DNA-binding and dimerisation domain) Panther:11945 MADS BOX PROTEIN KOG:0014 MADS box transcription factor
PtMADS92	POPTR_0005s16500	scaffold_5: 13345268 - 13345565		Pfam:00319 SRF-type transcription factor (DNA-binding and dimerisation domain) Panther:11945 MADS BOX PROTEIN
PtMADS93	POPTR_0007s02770	scaffold_7: 1612601 - 1612946		Pfam:00319 SRF-type transcription factor (DNA-binding and dimerisation domain) Panther:11945 MADS BOX PROTEIN
PtMADS94	POPTR_0007s05980	scaffold_7: 4126423 - 4128527		Pfam:00319 SRF-type transcription factor (DNA-binding and dimerisation domain) Panther:11945 MADS BOX PROTEIN KOG:0014 MADS box transcription factor
PtMADS95	POPTR_0008s08780	scaffold_8: 5431595 - 5435799		Pfam:00319 SRF-type transcription factor (DNA-binding and dimerisation domain)Panther:11945 MADS BOX PROTEINKOG:0014 MADS box transcription factor
PtMADS96	POPTR_0010s17450	scaffold_10: 16038303 - 16041326		Pfam:00319 SRF-type transcription factor (DNA-binding and dimerisation domain) Panther:11945 MADS BOX PROTEIN KOG:0014 MADS box transcription factor

Leseberg name	Locus Name	Location	Research on this gene	Functional annotations for this locus
PtMADS97	POPTR_0018s08400	scaffold_18: 9348208 - 9348865		Pfam:00319 SRF-type transcription factor (DNA-binding and dimerisation domain) Panther:11945 MADS BOX PROTEIN
PtMADS98 and PtMADS99	POPTR_0007s10345	scaffold_7: 10218463 - 10218950	These two genes are annotated as a single gene on version 2.0 of the genome.	Pfam:00319 SRF-type transcription factor (DNA-binding and dimerisation domain) Panther:11945 MADS BOX PROTEIN
PtMADS100	POPTR_0006s21630	scaffold_6: 20654652 - 20656061		Pfam:00319 SRF-type transcription factor (DNA-binding and dimerisation domain) Panther:11945 MADS BOX PROTEIN KOG:0014 MADS box transcription factor
PtMADS101	POPTR_0013s01880	scaffold_13: 1089954 - 1090806		Pfam:00319 SRF-type transcription factor (DNA-binding and dimerisation domain) Panther:11945 MADS BOX PROTEIN KOG:0014 MADS box transcription factor
PtMADS102	POPTR_0013s00350	scaffold_13: 120118 - 120792		Pfam:00319 SRF-type transcription factor (DNA-binding and dimerisation domain) Panther:11945 MADS BOX PROTEIN KOG:0014 MADS box transcription factor
PtMADS103	POPTR_0016s06850	scaffold_16: 4631275 - 4632726		Pfam:00319 SRF-type transcription factor (DNA-binding and dimerisation domain)Panther:11945 MADS BOX PROTEINKOG:0014 MADS box transcription factor
PtMADS104	POPTR_0005s00420	scaffold_5: 134895 - 135674		Pfam:00319 SRF-type transcription factor (DNA-binding and dimerisation domain) Panther:11945 MADS BOX PROTEIN KOG:0014 MADS box transcription factor
PtMADS105	POPTR_0013s11210	scaffold_13: 11884123 - 11884934		Pfam:00319 SRF-type transcription factor (DNA-binding and dimerisation domain) Panther:11945 MADS BOX PROTEIN

Appendix E Protein alignment of MADS-box protein domain for *Populus trichocarpa* MADS-box genes used to construct the phylogenetic tree in chapter five

```
Gene name          10         20         30         40         50         60         70         80
>AP1       MSRGRVQLKR IENKINRQVT FSKRRAGLLK KAHEISVLCD AEVALVVFSH KGKLFEYSTD -S-CM----- ----------
>PTAP1_1   MGRGRVQLKR IENKINRQVT FSKRRTGLLK KAHEISVLCD AEVALIVFSH KGKLFEYSTN -A-CM----- ----------
>PTAP1_2   MGRGRVQLKR IENKINRQVT FSKRRTGLLK KANEISVLCD AEVALIVFSH KGKLFEYSTD -D-SM----- ----------
>M1        ---------- ---------- ---------- ---------- -----NFVEH SSEIFHFF-G ---SM----- ----------
>M2        MGRGRIQLKR IENKINRQVT FSKRRSGLLK KAHEISVLCD AEVALIVFST KGKLFEYATD -S-CM----- ----------
>M6        MGRGRVQLKR IENNISRQVT FSKRRTGLLK KAHEISVLCD ADVAVIVFST KGKLFEYSTD -S-SM----- ----------
>PI        MGRGKIEIKR IENANNRVVT FSKRRNGLVK KAKEITVLCD AKVALIIFAS NGKMIDYCCP -SMDL----- ----------
>MADS_1    MGRGKIEIKR IENSSNRQVT YSKRRSGIIK KAKEITVLCD AQVSLVIFAS SGRMHEYCSP -STTV----- ----------
>MADS_2    MGRGKIEIKR IENASNRQVT YSKRKNGIIK KAKEITVLCD AQVSLVIFAS SGRMHEYCSP -STTV----- ----------
>AP3       MARGKIQIKR IENQTNRQVT YSKRRNGLFK KAHELTVLCD ARVSIIMFSS SNKLHEYISP -NTTT----- ----------
>PTD       MGRGKIEIKK IENPTNRQVT YSKRRNGIFK KAQELTVLCD AKVSLIMFSN TNKLNEYISP -STST----- ----------
>PMADS30   MGRGKIEIKK IENPTNRQVT YSKRRNGIFK KAQELTVLCD AKVSLIMFSN TNKFHEYISP -STTT----- ----------
>M4        MGRGKIEIKK IENPTNRQVT YSKRRNGIFK KAQELTVLCD AKVSLIMFSN TNKFHEYISP -STTT----- ----------
>AG        SGRGKIEIKR IENTTNRQVT FCKRRNGLLK KAYELSVLCD AEVALIVFSS RGRLYEYSNN -S--V----- ----------
>PTAG1     LGRGKVEIKR IENTTNRQVT FCKRRSGLLK KAYELSVLCD AEVALIVFSS RGRLYEYSND -S--V----- ----------
>PTAG2     LGRGKVEIKR IENTTNRQVT FCKRRNGLLK KAYELSVLCD AEVALIVFSS RGRLYEYSNN -S--V----- ----------
>SHP1      LGRGKIEIKR IENTTNRQVT FCKRRNGLLK KAYELSVLCD AEVALVIFST RGRLYEYANN —S--FIYLLL EKKKKKKKKK
>SHP2      IGRGKIEIKR IENTTNRQVT FCKRRNGLLK KAYELSVLCD AEVALVIFST RGRLYEYANN -S--V----- ----------
>STK       MGRGKIEIKR IENSTNRQVT FCKRRNGLLK KAYELSVLCD AEVALIVFST RGRLYEYANN -N--I----- ----------
>SCA19_12  MGRGKIEIKR IENTTNRQVT FCKRRNGLLK KAYELSVLCD AEVALIVFSS RGRLYEYANN -N--I----- ----------
>F2        MGRGKIEIKR IENTTNRQVT FCKRRNGLLK KAYELSVLCD AEVSLIVFSS RGRLYEYANN -N--I----- ----------
>PTM3_PTM4 MGRGRVELKR IENKINRQVT FAKRRNGLLK KAYELSVLCD AEVALIIFSN RGKLYEFCST -S-NM----- ----------
>SEP1      MGRGRVELKR IENKINRQVT FAKRRNGLLK KAYELSVLCD AEVALIIFSN RGKLYEFCSS -S-NM----- ----------
>SEP2      MGRGRVELKR IENKINRQVT FAKRRNGLLK KAYELSVLCD AEVSLIVFSN RGKLYEFCST -S-NM----- ----------
>SEP3      MGRGRVELKR IENKINRQVT FAKRRNGLLK KAYELSVLCD AEVALIIFSN RGKLYEFCSS -S-SM----- ----------
>SEP4      MGRGKVELKR IENKINRQVT FAKRRNGLLK KAYELSVLCD AEIALLIFSN RGKLYEFCSS PS-GM----- ----------
>PTM5      MVRGKTQMRR IENATSRQVT FSKRRNGLLK KAFELSVLCD AEVALIVFSP RGKLYEFAST -S--M----- ----------
>PtMADS31  MGRGRVELKR IENKINRQVT FAKRRNGLLK KAYELSVLCD AEVALIIFSN SGKLFEFCSS -S-NM----- ----------
>PtMADS7   MAREKIKIKK IDNVTARQVT FSKRRRGLFK KAEELSVLCD AEVAVIIFSA TGKLFEYSSS -S--M----- ----------
>PtMADS12  MVRGKTQMKR IENATSRQVT FSKRRNGLLK KAFELSVLCD AEVALIVFSS RGKLYEFSSS -S--I----- ----------
>PtMADS17  MGRGRVELKR IENKINRQVT FAKRRNGLLK KAYELSVLCD AEVALIIFSN RGKLYEFCST -S-NM----- ----------
>PtMADS21  MAREKIKIKK IDNVAARQVT FSKRRRGLLK KAEELSVLCD VEVAVIIFSA TGKLFEYSSS -S--M----- ----------
>PMADS27   MTRKKIQIKK IDNTAARQVT FSKRRRGLFK KAYELSTLCD AEIALTVFSA TGKLFEYSNT -R--T----- ----------
>PtMADS82  MTRKKIQIKK IDNTAARQVT FSKRRRGLFK KAYELSTLCD AEIALMVFSA TGKLFEYSNS -S--M----- ----------
```

```
Gene name          90        100       110        120        130        140        150        160
>AP1        ---------- EKILERYERY SYAERQLIAP —ESDVNTNWS MEYNRLKAKI ELLERNQRHY LGEDLQAMSP KELQNLEQQL
>PTAP1_1    ---------- EKILERHERY SYAERQLVAT -DLDSQG-NW TLEYNRLKAK VELLQRNHRH YLGEDLDSVS LKELQNLEQQ
>PTAP1_2    ---------- EKILERYERY SYAERQLVAT -DLDSQG-NW TLEYNRLKAK VELLQRNHRN YLGEDLDSMS LKELQNLEQQ
>M1         ---------- ESILERYERC SYLEQQLVPN -GSEHQE-SW SLEHPKLMAR VEILQRNLRN YAGQELDPLS LKELQYLEQQ
>M2         ---------- ERILERYERY SYAERQLLAN DDPENHG-SW TLEYAKLKAR VDVLQRNQRH FMGEDLDSLN IKELQNLEHQ
>M6         ---------- ESILERYERC SYAEQQFVPH -GPEHQG-SW FLEHPKLRAR VELLQRNLRN YTGQDLDPLS YKELQHLEQK
>PI         ---------- GAMLDQYQKL SG--KK-WD- AK---HE-NL SNEIDRIKKE NDSLQLELRH LKGEDIQSLN LKNLMAVEHA
>MADS_1     ---------- VDLLDKYHKQ SG--KRLWD- AK---HE-NL SNEIDRIKKE NESMQIELRH LKGQDISSLP HKELMAIEEA
>MADS_2     ---------- VDLLDKYHKQ SG--KRLWD- AK---HE-NL SKEIDRIKKE NDSMQIELRH LKGEDISSLH HTELMAIEEA
>AP3        ---------- KEIVDLYQTI SDVDVWA--- TQ---YE-RM QETKRKLLET NRNLRTQIKQ RLGECLDELD IQELRRLEDE
>PTD        ---------- KKIYDQYQNA LGIDLWG--- TQ---YE-KM QEHLRKLNDI NHKLRQEIRK YHVIKTQNET YRKKVS-D--
>PMADS30    ---------- KKIYDQYQKA LGIDLWS--- AQ---YE--- ------IS-I NHKLKKEIRQ RIGEDLNELS IDHLRVLEQN
>M4         ---------- KKIYDQYQKA LGIDLWS--- AQ---YE--- ------IS-I NHKLKKEIRQ RIGEDLNELS IDHLRVLEQN
>AG         ---------- KGTIERYKK AISDNSNTGSV AEIN-AQ-YYQ QESAKLRQQ IISIQNSNRQ LMGETIGSMS PKELRNLEGR
>PTAG1      ---------- KSTIERYKK ASADSSNTGSV SEAN-AQ-YYQ QEAAKLRSQ IGNLQNSNRH MLGEALSSLS VKELKSLEIR
>PTAG2      ---------- KSTIERYKK ACADSSNNGSV SEAN-AQ-FYQ QEAAKLRSQ IGNLQNSNRN MLGESLSALS VKELKSLEIK
>SHP1       NLWIYSSHVV RGTIERYKK ACSDAVNPPSV TEAN-TQ-YYQ QEASKLRRQ IRDIQNSNRH IVGESLGSLN FKELKNLEGR
>SHP2       ---------- RGTIERYKK ACSDAVNPPTI TEAN-TQ-YYQ QEASKLRRQ IRDIQNLNRH ILGESLGSLN FKELKNLESR
>STK        ---------- RSTIERYKK ACSDSTNTSTV QEIN-AA-YYQ QESAKLRQQ IQTIQNSNRN LMGDSLSSLS VKELKQVENR
>SCA19_12   ---------- RSTIDRYKK ASSDSSNASSI TEIN-AQ-YYQ QESAKLRQQ IQMLQNSNRH LMGDAVSNLS VKELKQLENR
>F2         ---------- RSTIDRYKK VSSDSSNTASI TEIN-AQ-YYQ QESAKMRQQ IQLLQNSNRH LMGEAVSNLS VKELKQLENR
>PTM3_PTM4  ---------- LKTLERYQK CSYGAEEVN-K PAKE-LE-SSY REYLKVKAR FEALQRTQRN LLGEDLGPLN TKELEQLERQ
>SEP1       ---------- LKTLDRYQK CSYGSIEVNNK PAKE-LE-NSY REYLKLKGR YENLQRQQRN LLGEDLGPLN SKELEQLERQ
>SEP2       ---------- LKTLERYQK CSYGSIEVNNK PAKE-LE-NSY REYLKLKGR YENLQRQQRN LLGEDLGPLN SKELEQLERQ
>SEP3       ---------- LRTLERYQK CNYGAPEPNVP SRELAVELSSQ QEYLKLKER YDALQRTQRN LLGEDLGPLS TKELESLERQ
>SEP4       ---------- ARTVDKYRK HSYATMDPN-Q SAKD-LQ-DKY QDYLKLKSR VEILQHSQRH LLGEELSEMD VNELEHLERQ
>PTM5       ---------- QETIERYRR HVKENNTNKQP VEQN-ML-QLK EEAASMIKK IEHLEVSKRK LLGECLGSCT VEELQQIEQQ
>PtMADS31   ---------- ATTIEKYQR FSYGALEGG-Q SEKETQQ-NNY QEYLKLKTR VDVLQRSQRN LLGEDLGNLG TMELDQLENQ
>PtMADS7    ---------- KDVLARYNL HSNNLDKINPP SLE--LQL-EN SNHMRLSKE VSEKSHQLRR MRGEDLHGLN IEELQQLEKA
>PtMADS12   ---------- NRTIERYQK RAKDVGISSKM VQDN-IQ-PVK EDTFTLAKK IELLEVSKRK LLGEGLETCS TDDLQQLENQ
>PtMADS17   ---------- LKTLERYQK CSYGAEEVN-K PAKE-LE-SSY REYLKVKAR FEALQRTQRN LLGEDLGPLN TKELEQLERQ
>PtMADS21   ---------- KDVLARYNL HSNNLDKLNQP SLE--LQL-EN SNHMRLRKE VSEKSHQLRR MRGEELQGLN IEELQQLEKV
>PMADS27    ---------- ID------- ---------- ---------- --------- ---------- ---------- ----------
>PtMADS82   ---------- GQVIERRNL HPKNINTLDQP SLE--KQL-DG GVHAMLIKE IAKKNRELRH MRGEDLQGLD LEELQKLEKI
```

```
Gene name          170        180        190
>AP1         DTALKHIRTR KNQLMYESIN ELQKKE----
>PTAP1_1     IDTALKLIRE RKNHLMYQSI SELQIKE---
>PTAP1_2     IDTALKHIRA RKNHLMSQSI SELQRKE---
>M1          IDTALKRIRS RKNQLIHESL NELRKKE---
>M2          IDSALKHVRS RKNQL----- ----------
>M6          IDTALKSVRS RKNQLVHESL AEMQKKE---
>PI          IEHGLDKVRD HQMEILISK- R---RNE---
>MADS_1      LDTGLAAVRK KQMEFHSML- E---QNE---
>MADS_2      LDAGLAAVRK KQMEYHSML- E---QNE---
>AP3         MENTFKLVRE RKFKSLGNQI ETTKKKN---
>PTD         ---------- KKN-ILPLQF HASK------
>PMADS30     MTEALNGVRG RKYHVIKTQT ETYKKKV---
>M4          MTEALNGVRG RKYHVIKTQT ETYKKKV---
>AG          LERSITRIRS KKNELLFSEI DYMQKRE---
>PTAG1       LEKGISRIRS KKNELLFAEI EYMQKRE---
>PTAG2       LEKGIGRIRS KKNELLFAEI EYMQKRE---
>SHP1        LEKGISRVRS KKNELLVAEI EYMQKRE---
>SHP2        LEKGISRVRS KKHEMLVAEI EYMQKRV---
>STK         LEKAISRIRS KKHELLLVEI ENAQKRE---
>SCA19_12    LERGITRIRS KKHELLLAEI EYLQKRE---
>F2          LERGMTRIRS KKHELLLAEI EYMQKRE---
>PTM3_PTM4   LESSLNQVRS TKTQYMLDQL ADLQNKE---
>SEP1        LDGSLKQVRS IKTQYMLDQL SDLQNKE---
>SEP2        LDGSLKQVRC IKTQYMLDQL SDLQGKE---
>SEP3        LDSSLKQIRA LRTQFMLDQL NDLQSKE---
>SEP4        VDASLRQIRS TKARSMLDQL SDLKTKE---
>PTM5        LERSVSTIRA RKNQVFKEQI ELLRQKE---
>PtMADS31    LDSSLKQIRS RKGQFVLDEL SELQRKE---
>PtMADS7     LEVGLSRVLE TKGERIMNEI STLERKG---
>PtMADS12    LGRSLTRIRA RKNQLFRERI EKLKGEE---
>PtMADS17    LESSLNQVRS TKTQYMLDQL ADLQNKE---
>PtMADS21    LEVGLCCVLE TKGERIMNEI STLERKG---
>PMADS27     --TSICEIL- ---------- LMLPMYE---
>PtMADS82    MEGSLRRLVE EKGGKIINEI DALKTKG---
```