Elucidating the genetic basis and molecular mechanisms of dioecy in *Populus*

Dissertation submitted to the University of Hamburg,
Faculty of Mathematics, Informatics and Natural Sciences,
Department of Biology
by
Ana Paula Leite Montalvão

Doctoral Thesis
Universität Hamburg
2022

Thesis Committee:

Supervisor:
PD Dr. Birgit Kersten
Thünen Institute of Forest Genetics
Genome Research Group, Research Area Bioinformatics

Co-supervisor:
Prof. Dr. Julia Kehr
University of Hamburg
Faculty of Mathematics, Informatics and Natural Sciences, Department of Biology
Institute of Plant Sciences and Microbiology, Molecular plant genetics

Other members:

Prof. Dr. Sigrun Reumann
University of Hamburg
Faculty of Mathematics, Informatics and Natural Sciences, Department of Biology
Institute of Plant Sciences and Microbiology, Plant biochemistry and infection biology

PD Dr. Cornelia Heinze
University of Hamburg
Faculty of Mathematics, Informatics and Natural Sciences, Department of Biology
Institute of Plant Sciences and Microbiology, Molecular phytopathology

Disputation date: 25.03.2022

This work was carried out from October 2018 to December 2021 at the Thünen Institute of Forest Genetics, Großhansdorf, Germany, in the Genome Research Group of PD Dr. Matthias Fladung. The project was financed by a grant to Dr. Niels A. Müller by the Deutsche Forschungsgemeinschaft (DFG grant n° MU 4357/1-1).

A short-term scholarship for international students was granted to Ana Paula Leite Montalvão as part of the funding program STIBET III Matching Funds offered by the German Academic Exchange Service (DAAD) to finance the completion of the doctoral thesis for a period of 6 months (October 2021 to March 2022).

This dissertation is submitted as a cumulative thesis based on the work contained in four (4) peer-reviewed articles and one (1) in preparation for submission. The articles are organized in a thematic order in the following 5 main chapters:

Chapter 1

Leite Montalvão AP, Kersten B, Fladung M, Müller NA (2021). The diversity and dynamics of sex determination in dioecious plants. *Frontiers in Plant Science* 11, 580488. https://doi.org/10.3389/fpls.2020.580488

Chapter 2

Müller NA*, Kersten B*, **Leite Montalvão AP**, Mähler N, Bernhardsson C, Bräutigam K, Carracedo Lorenzo Z, Hönicka H, Kumar V, Mader M, Pakull B, Robinson KM, Sabatti M, Vettori C, Ingvarsson PK, Cronk Q, Street NR, Fladung M (2020). A single gene underlies the dynamic evolution of poplar sex determination. *Nature Plants* 6, 630-637. https://doi.org/10.1038/s41477-020-0672-9

Chapter 3

Leite Montalvão AP, Eikhof A, Müller NA. Recombination and segregation analysis of the white poplar SDR (In preparation).

Chapter 4

Kim G*, **Leite Montalvão AP***, Kersten B, Fladung M, Müller NA (2021). The genetic basis of sex determination in *Populus* provides molecular markers across the genus and reveals convergent evolution. *Silvae Genetica* 70, 145-155.
https://doi.org/10.2478/sg-2021-0012

Chapter 5

Leite Montalvão AP, Kersten B, Kim G, Fladung M, Müller NA. *ARR17* controls dioecy in *Populus* by repressing B-class MADS-box gene expression. *Philosophical Transactions of the Royal Society B* (in press).

* equal contribution.

Table of contents

Table of Figures

Abstract

More than 100 years ago, botanists and biologists were already interested in the fascinating diversity of plant mating systems and sex determination of monoecious and dioecious plants. Today, despite the remarkable progress that has been made, many exciting questions about the evolutionary dynamics and the genetic and molecular mechanisms underlying sexual differentiation remain open. With the advance of genomic technologies, including long-read sequencing, the genetic basis of dioecy in an expanding number of species is finally being tackled. In line with that, **Chapter 1** presents a review of the current literature concerning the diversity of sex determination, the contrasting evolutionary pathways to dioecy and the sex-determining genes identified in various dioecious plants. The following chapters combine existing and newly generated data and present the groundwork to elucidate the genetic basis and molecular mechanisms of dioecy in *Populus*. Upon a CRISPR-Cas9 induced mutation, the poplar *ARR17* gene was experimentally validated as a single factor of sex determination in poplars in **Chapter 2**. This single-gene system in *Populus* is conserved across the species despite millions of years of independent evolution, and repeated sex chromosome turnovers. Sex determination can be achieved by a dominant negative switch – the Y-specific partial *ARR17* duplicates – producing small RNAs that likely act via RNA-directed DNA methylation (RdDM) to silence the *ARR17* gene in poplar species with a XY system (i.e. aspen, balsam poplar) or by a presence/absence mutation of *ARR17* in the sex-determining region (SDR) in a ZW system (i.e. white poplar). There are multiple possible scenarios in which sex-determining genes and SDRs can arise. Accumulating data indicate that the SDRs within the genus *Populus* evolved independently. Accordingly, **Chapter 3** assessed the potential suppression of recombination along the white poplar SDR. Hemizygosity of the W-specific poplar sex-determining gene *ARR17* may cause the recombination suppression between Z and W. Female-specific hemizygous regions around the *ARR17* gene were also identified in closely related species to white poplar in **Chapter 4**. Additionally, the Y-specific partial *ARR17* duplicates as well as the W-specific *ARR17* gene were further explored by utilizing them as molecular markers. This chapter also demonstrates the convergent evolution of sex determination in *Populus*. The investigation of the genetic basis clarified the biological

function of the *ARR17* gene in *Populus*, but the molecular function of *ARR17* remained unclear. Consequently, **Chapter 5** explored the genetic networks regulated by *ARR17*. The poplar *ARR17* gene is only expressed in a certain tissue during a specific time window. Although *ARR17* is annotated as a type-A response regulator supposedly involved in cytokinin hormone signaling, CRISPR-Cas9-mediated *arr17* knockout only affected the expression of a strikingly small number of genes indicating a specific role in the regulation of floral development rather than a generic function in hormone signaling. When *ARR17* is active, female development is initiated, while, if *ARR17* is inactive, B class MADS-box genes are expressed, and male development occurs. Interestingly, recent data suggest that a shared basis of sex determination mediated by *ARR17* between species in the dioecious sister genera *Populus* and *Salix* and the closely related monoecious species *Poliothyrsis sinensis* might be possible. Future studies on more systems, including non-dioecious ones, are crucial for allowing a broader perspective and to elucidate the biological reasons for the different evolutionary dynamics between hermaphroditic, monoecious and dioecious species.

Kurzfassung

Schon vor mehr als 100 Jahren interessierten sich Botaniker und Biologen für die faszinierende Vielfalt der pflanzlichen Paarungssysteme und die Geschlechtsbestimmung bei ein- und zweihäusigen Pflanzen. Heute sind trotz der enormen Fortschritte viele spannende Fragen über die evolutionäre Dynamik und die genetischen und molekularen Mechanismen, die der sexuellen Differenzierung zugrunde liegen, noch offen. Mit dem Fortschritt der Genomik, einschließlich der Long-Read-Sequenzierung, kann die genetische Grundlage der Zweihäusigkeit bei einer wachsenden Zahl von Arten aufgeklärt werden. In **Kapitel 1** wird ein Überblick über die aktuelle Literatur zur Vielfalt der Geschlechtsbestimmung, zu den unterschiedlichen evolutionären Wegen zur Zweihäusigkeit und zu den in verschiedenen zweihäusigen Pflanzen identifizierten geschlechtsbestimmenden Genen gegeben. In den folgenden Kapiteln werden vorhandene und neu gewonnene Daten kombiniert und die Grundlagen für die Aufklärung der genetischen Grundlagen und molekularen Mechanismen der Zweihäusigkeit in der Gattung *Populus* vorgestellt. Mithilfe von CRISPR-Cas9-induzierten Mutationen wurde in **Kapitel 2** das Pappel-Gen *ARR17* experimentell als einzelner Faktor der Geschlechtsbestimmung bei Pappeln validiert. Dieses Einzelgensystem in *Populus* ist trotz Millionen von Jahren unabhängiger Evolution und trotz wiederholter Umstrukturierungen der Geschlechtschromosomen über die gesamte Gattung hinweg konserviert. Die Geschlechtsbestimmung kann durch einen dominanten negativen Schalter - die Y-spezifischen partiellen *ARR17*-Kopien - erreicht werden. Hier werden kleine RNAs produziert, die wahrscheinlich über RNA-gesteuerte DNA-Methylierung (RdDM) wirken, um das *ARR17*-Gen in Pappelarten mit einem XY-System (z.B. Espe, Balsam-Pappel) zu Silencen. Alternativ kann die Geschlechtsbestimmung durch eine An-/Abwesenheitsmutation von *ARR17* in der geschlechtsbestimmenden Region (SDR) in einem ZW-System (z.B. Weiß-Pappel) reguliert werden. Es gibt mehrere mögliche Szenarien, in denen geschlechtsbestimmende Gene und SDRs entstehen können. Neue Daten zeigen, dass sich die SDRs innerhalb der Gattung *Populus* unabhängig voneinander entwickelt haben. Daher wurde in **Kapitel 3** die mögliche Suppression der Rekombination entlang der SDR der Weißpappel untersucht. Hemizygotie des W-spezifischen

geschlechtsbestimmenden Gens *ARR17* der Pappel könnte die Rekombinationsunterdrückung zwischen Z und W im geschlechtsbestimmenden System mit nur einem Gen verursachen. W-Chromosom-spezifische hemizygote Regionen um das *ARR17*-Gen wurden auch bei eng mit Weiß-Pappel verwandten Arten in **Kapitel 4** identifiziert. Darüber hinaus wurden die Y-spezifischen partiellen *ARR17*-Kopien sowie das W-spezifische *ARR17* weiter untersucht, indem sie als molekulare Marker verwendet wurden. Dieses Kapitel zeigt auch die konvergente Evolution der Geschlechtsbestimmung bei *Populus*. Die Untersuchung der genetischen Grundlage erklärt die biologische Funktion von *ARR17* in Pappeln, nicht aber die molekulare Funktion. Daher wurden in **Kapitel 5** die genetischen Netzwerke untersucht, die durch *ARR17* reguliert werden. Das *ARR17*-Gen wird nur in einem bestimmten Gewebe und in einem bestimmten Zeitfenster exprimiert. Obwohl *ARR17* als Typ-A-Response Regulator annotiert ist, der an der Cytokinin-Hormonsignalisierung beteiligt ist, wirkte sich der CRISPR-Cas9-vermittelte Knockout von *ARR17* nur auf die Expression einer geringen Anzahl von Genen aus, was auf eine spezifische Rolle bei der Regulierung der Blütenentwicklung und nicht auf eine generelle Funktion bei der Hormonsignalübertragung hindeutet. Wenn *ARR17* aktiv ist, wird die weibliche Entwicklung eingeleitet, wenn *ARR17* inaktiv ist, werden MADS-Box-Gene der Klasse B exprimiert und die männliche Entwicklung wird initiiert. Interessanterweise deuten neuere Daten darauf hin, dass die Arten der zweihäusigen Schwestergattungen *Populus* und *Salix* und die eng verwandte einhäusige Art *Poliothyrsis sinensis* eine gemeinsame Grundlage für die durch *ARR17* vermittelte Geschlechtsbestimmung haben könnten. Zukünftige Studien an weiteren Systemen, einschließlich nicht zweihäusiger Arten, sind von entscheidender Bedeutung, um eine breitere Perspektive zu ermöglichen und die biologischen Gründe für die unterschiedliche evolutionäre Dynamik zwischen hermaphroditen, einhäusigen und zweihäusigen Arten zu ergründen.

General introduction

Sex determination is an important developmental event in the life cycle of sexually reproducing organisms. For over a century, it has intrigued both animal and plant researchers. In flowering plants, the observation of floral structures and the diversity of floral forms has shaped the basis of our current knowledge on plant sexual diversity (Darwin, 1877; Linnaenus, 1758). While pronounced differences between males and females are the norm for many groups of animals, most plant species are hermaphrodites (Bachtrog et al., 2014; Käfer et al., 2017). Only 10% of all flowering plant species separate unisexual male and female flowers, which may occur either in different locations of the same individual (monoecy) or in different individuals (dioecy) (Renner, 2014; Renner and Ricklefs, 1995; Westergaard, 1958). A breakthrough discovery through cytological analysis on the inheritance patterns of chromosomes in insects led to the discovery of sex chromosomes (Stevens, 1905; Wilson, 1905). In plants, sex chromosomes were first identified in the liverwort *Sphaerocarpos donnelli* Austin (Allen, 1917). Following the discovery of sex chromosomes and the fact that sex can be genetically determined, interesting evolutionary theories on how sex chromosomes originate and evolve were formulated motivating several studies to better understand the evolution and the diversity of sex determination systems (**Figure 1**).

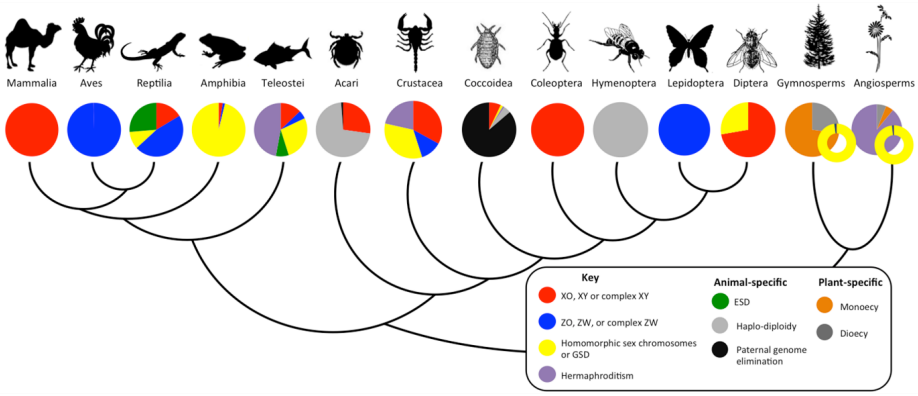


FIGURE 1. THE DIVERSITY OF SEX DETERMINATION SYSTEMS IN ANIMAL AND PLANT SPECIES. THERE CAN BE GENETIC SEX DETERMINATION (GSD) OR ENVIRONMENTAL SEX DETERMINATION (ESD). THE SEX CHROMOSOMES CAN BE CLASSIFIED ACCORDING TO THE HETEROGAMETIC SEX (I.E. XY, ZW AND VARIATIONS). FIGURE FROM (BACHTROG ET AL., 2014)[1].

[1] Copyright: © 2014 Bachtrog et al. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

The separation of the sexes in flowering plants

Flowering plants have gone through numerous transitions between sexual systems across the phylogeny (Ashman et al., 2014; Bachtrog et al., 2014). But why did the so-called "perfect" flowers, that is hermaphrodites, evolve independently into "imperfect" forms (e.g. dioecious) in diverse lineages, possibly a thousand times (Renner, 2014)? And most importantly, what kind of molecular mechanisms cause differential sex manifestation – are there many ways or are constraints on the underlying molecular networks forcing parallel evolution?

Despite being rare, dioecious plants are spread across the entire phylogeny (Renner, 2014). The consensus is that dioecy evolved from either hermaphroditism or monoecy (Charlesworth, 2013; Ming et al., 2011). The most recognized evolutionary model (in the literature referred as the "two-gene model") states that dioecy likely evolved from a hermaphrodite ancestor through at least two successive mutations (a male-sterility mutation creates females, and a female-sterility mutation creates males) linked in a non-recombining sex-determining region (SDR) in the chromosome (Charlesworth and Charlesworth, 1978). Hence, identifying the genes located within the SDR has been considered a starting point to propose candidate genes responsible for sex determination (see **Chapter 1**) (**Leite Montalvão et al., 2021**).

Genetic sex determination (GSD) in dioecious plants is often associated with sex chromosomes. The classical model for the evolution of sex chromosomes in flowering plants assumes a linear process, with the linkage between the two genes as the first step towards the evolution of suppressed recombination, which is expected to spread, and progress towards chromosome differentiation and, finally, degeneration (Bull, 1983; Charlesworth et al., 2000, 2005; Charlesworth, 2016; Furman et al., 2020; Ming et al., 2011). Sex chromosome pairs, therefore, may present different levels of genetic divergence, and are classified either in homomorphic or heteromorphic, the latter differing in size and identified through cytological analysis (Filatov, 2015; Furman et al., 2020). Indeed, genetic degeneration has been experimentally demonstrated in a few model plants e.g. *Carica papaya* L. and *Silene latifolia* Poir. (Bergero et al., 2015; Gschwend et al., 2012; Papadopulos et al., 2015) and sex chromosome differentiation and degeneration have been thought to be inevitable and irreversible consequences of

sex chromosome evolution (Westergaard, 1958). Recent comparative studies, however, indicate that this may not be always the case, as recombination suppression may or may not occur (Charlesworth, 2021), and the degree of divergence is not necessarily associated with SDR size or age (Renner and Müller, 2021; Wright et al., 2016).

In addition to the different levels of chromosome differentiation, different types of sex chromosome systems exist. The originally discovered XX/XY system consists of males carrying two different chromosomes (male heterogametic) with a XY karyotype, while females carry two X chromosomes (XX). Conversely, in the ZW/ZZ system, the females are the heterogametic sex represented by ZW, while males are ZZ (Charlesworth, 2013; Wilson, 1909; Wright et al., 2016). In dioecious plants, the XY system is apparently predominant when comparing the sex-determining systems of 85 species with known sex chromosomes (**Chapter 1**: **Figure 1**). While 84.7% of these species present male heterogamety (XY system), only 15.3% present female heterogamety (ZW system) (**Leite Montalvão et al., 2021**).

Due to the myriad of independent evolutionary origins of dioecy, one cannot expect different dioecious species to have the same sex-determining genes thus, different genetic and molecular mechanisms may be underlying the development of unisexual flowers (Diggle et al., 2011). Nevertheless, parallel evolution (i.e. repeated but independent usage of the same genetic solutions) (Stern, 2013) could also play a role. The two-gene model has strongly influenced the research on the evolution of dioecy for several decades. Recent work, however, provides exciting exceptions to the classical theory highlighting single-gene sex-determining systems (or sex switches), listed as one of the most important future issues of plant sex chromosome research (Charlesworth, 2016).

Long-read genome sequencing

The development and improvement of genomic technologies is helping to unlock the genetic basis of dioecy in an expanding number of species (Carey et al., 2021; Henry et al., 2018; **Leite Montalvão et al., 2021**; Ming et al., 2011).

The introduction of next generation sequencing (NGS) allowed cost-effective high-throughput sequencing and thus great advances in genome assembly, population genomics and genome-wide association (GWA) studies (Carey et al., 2021; Jiao and Schneeberger, 2017; Vekemans et al., 2021). In sex determination research, genome assembly (i.e. alignment and merging of DNA fragments to match the original sequence), is an essential step. NGS offers affordable and highly accurate short reads (100-300 bp), nevertheless, on sex chromosomes where recombination stops over a relevant portion of the genome (e.g. large SDRs), the accumulation of repetitive sequences will make scaffolding a challenging task, and likely result in an incomplete genome assembly (Vekemans et al., 2021). Moreover, duplicated sequences that might be involved in sex determination (Akagi et al., 2014) tend to collapse into a single genomic region when analyzed with short-read sequencing.

The introduction of long-read sequencing (or third generation sequencing) bridge the gap between high-quality reference sequences and low-cost short-read assemblies (Jiao and Schneeberger, 2017). Additionally, the production of longer contigs outperforms short-read scaffolds and aids in resolving regions with suppressed recombination. Long-read sequencing can be performed nowadays by employing Pacific Biosciences (PacBio) single molecule real-time (SMRT) sequencing and Oxford Nanopore Technologies (ONT) platforms. The SMRT sequencing detects a difference in fluorescence corresponding to the addition of a specific nucleotide by a polymerase at the bottom of a well, while the flow cell utilized in the ONT devices (**Figure 2**) identifies sequences of DNA based on the changes in electrical conductivity as the strands pass through the biological nanopores (Amarasinghe et al., 2020; Nurk et al., 2020). The application of ONT allows DNA strands of indefinite length, depending on the quality and fragmentation of the input DNA, producing ultra-long reads which can be assembled in mega-base contigs that can span complex regions (Vekemans et al., 2021) and even entire chromosomes.

These technologies may rely on distinct principles, nevertheless the resulting data can be analyzed utilizing similar tools and pipelines. Moreover, even though assemblies based only on long reads often produce highly complete and contiguous genomes, there are many situations where accuracy may be improved by employing PacBio-HiFi reads or hybrid approaches, using short reads together with long reads. The additional

application of chromatin conformation data (Hi-C) and optical mapping data will further improve the quality and completeness of the assembly (Amarasinghe et al., 2020; Nurk et al., 2020).

FIGURE 2. LONG-READ SEQUENCING WITH OXFORD NANOPORE TECHNOLOGIES USING MINION. SEQUENCING CAN BE PERFORMED AFTER HIGH QUALITY HIGH MOLECULAR WEIGHT DNA EXTRACTION, LIBRARY PREPARATION AND LOADING INTO THE MINION DEVICE. SEQUENCE OUTPUTS PROVIDE RAW DATA IN FASTQ FORMAT WHICH SUBSEQUENTLY CAN BE, FOR EXAMPLE, ASSEMBLED (E.G. *DE NOVO* ASSEMBLY), POLISHED AND ANNOTATED. PHOTO CREDITS: NIELS A. MÜLLER.

The genus *Populus* as a model system to study the evolution of dioecy

Members of the family Salicaceae (Malphigiales order, Eurosid I clade) are among the main systems for which sex determination and sex chromosome evolution have been explored. Aspen, cottonwood, and poplar are the common names of ecologically and economically important trees from the genus *Populus*. The *Populus* trees are wind-pollinated dioecious species adapted to a wide range of habitats spread worldwide occurring mostly in the northern hemisphere, although occasionally reported in southern regions or, in some cases, restricted to specific habitats (**Figure 3**). Currently, 29 species are recognized and subdivided into six sections (i.e. Tacamahaca, Abaso, Turanga, Populus, Leucoides and Aigeiros) based on morphological similarity and crossing possibilities (Cronk et al., 2015; Slavov and Zhelev, 2010). However, despite an apparent low species diversity, an accurate number of species within the genus *Populus* is still under debate (WFO, 2022). Divergence signs are obscured mostly due to a wide range of differing morphological characters and the frequent hybridization across the

genus (both within and among sections), giving rise to hybrids with commercially valuable traits widely used in breeding programs.

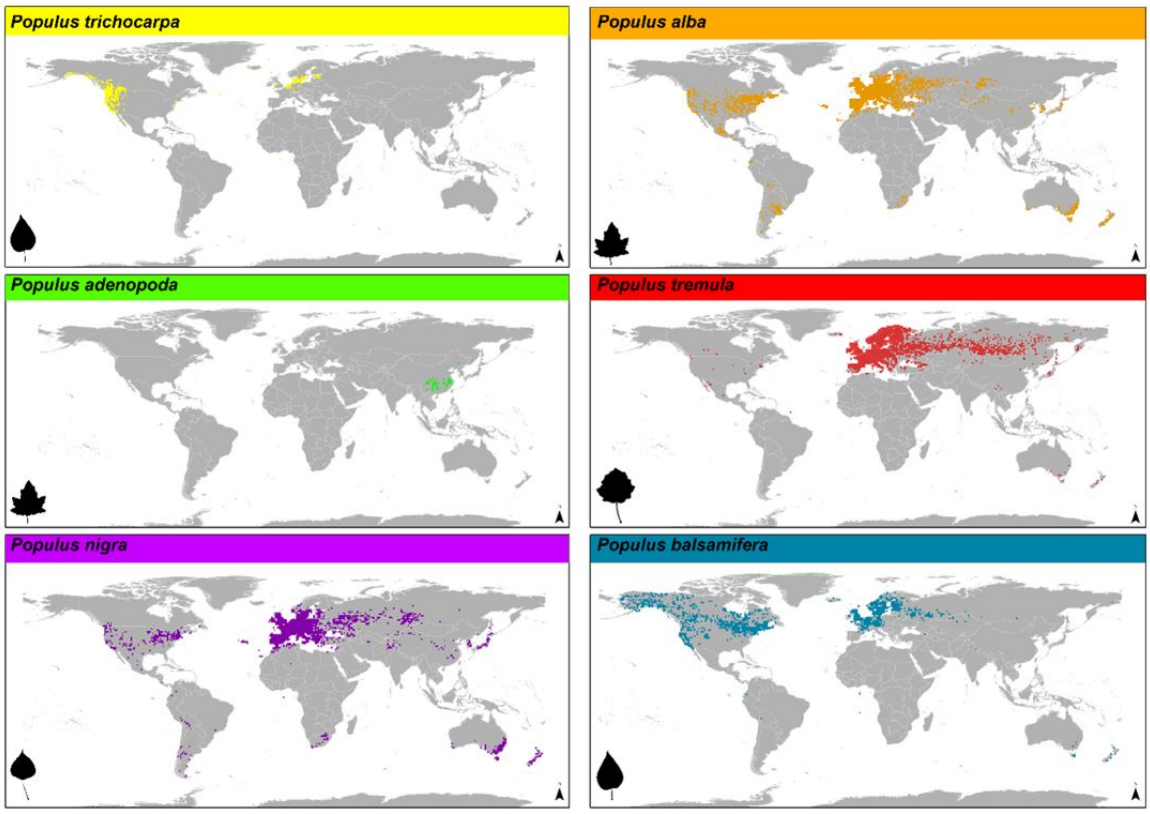


FIGURE 3. WORLDWIDE OCCURRENCES OF ECOLOGICALLY AND ECONOMICALLY IMPORTANT *POPULUS* SPECIES. *POPULUS TRICHOCARPA* (BLACK COTTONWOOD), *P. ALBA* (WHITE POPLAR), *P. ADENOPODA* (CHINESE ASPEN), *P. TREMULA* (EUROPEAN ASPEN), *P. NIGRA* (BLACK POPLAR) AND *P. BALSAMIFERA* (BALSAM POPLAR). REGISTERED OCCURRENCE COORDINATES ACQUIRED FROM THE GLOBAL BIODIVERSITY INFORMATION FACILITY – GBIF (WWW.GBIF.ORG) AND CONVERTED TO PIXELS USING ARCGIS SOFTWARE.

The species from the genus *Populus* are long-lived organisms and can have a life span of over 100 years (**Figure 4a**). Reproduction occurs for the first time after an individual reaches sexual maturity, which usually takes a decade. Subsequently, flowering occurs seasonally in springtime (Cronk et al., 2015). The shoot architecture accommodates vegetative buds, leaves and reproductive buds, which later will produce an inflorescence known as catkin (**Figure 4b**). Poplar catkins will only develop one floral sexual structure, either carpels (female) or stamens (male). There are no rudimentary or non-functional floral organs from the opposite sex (Boes and Strauss, 1994; Diggle et al., 2011).

FIGURE 4. ARBORETUM AT THE THÜNEN INSTITUTE OF FOREST GENETICS IN GROßHANSDORF, GERMANY. (A) ASPEN TREES MARKED FOR SAMPLING. (B) MALE AND FEMALE CATKINS OF *P. TREMULA*.

While several biological traits are best studied in established model systems such as *Arabidopsis thaliana* L., trees present unique physiological and genetic aspects that are better investigated in trees themselves (Bradshaw et al., 2000; Cronk, 2005). A female black cottonwood *P. trichocarpa* Torr. & A. Gray, genotype Nisqually-1, was the first woody species and the third plant after Arabidopsis and rice with its nuclear genome fully sequenced and assembled (Tuskan et al., 2006). Sequencing a *Populus* genome was possible due to its relatively compact genome size of approximately 500 mega bases (Mb) (Bradshaw et al., 2000; Cronk, 2005; Tuskan et al., 2006). Additional characteristics such as the remarkably fast growth rate, ease of vegetative propagation, and potential for genetic manipulation have been explored (Bradshaw et al., 2000; Cronk, 2005; Strauss and Martin, 2004) establishing poplars as a prominent model plant in forest genetics. The highly efficient *Agrobacterium*-mediated transformation system available for poplar (Fladung et al., 1997; Kumar and Fladung, 2001) is unmatched by any other forest tree offering a unique advantage to forest geneticists. When paired with the exceptional early flowering system established for different aspen lines at the Thünen Institute of Forest Genetics (Hoenicka et al., 2016), it allows the experimental confirmation of genetic and genomic hypothesis regarding floral development and enables the elucidation of the genetic basis of poplar dioecy.

Early flowering aspen lines

Species from the genus *Populus* present a long vegetative phase and regularly take a decade to reach sexual maturity. Due to minimal sexual dimorphism (McKown et al.,

2017; Renner and Müller, 2021; Robinson et al., 2014) it is only possible to determine the sex of an individual at the time of flowering. This long generation time is an obstacle to practical breeding and selection. A transgenic approach allied to the discovery of the genetic factors controlling the flowering network can advance breeding programs or the development of strategies for conservation, for example.

Studies of flowering time genes added an extra layer to the potential of poplars in plant biology. The genetic and molecular characterization of different flowering phenotypes in *A. thaliana* mutants allowed the identification of several flowering time control genes, including the *FLOWERING LOCUS T* (FT) (Kardailsky et al., 1999; Koornneef et al., 1991; Pin and Nilsson, 2012). The manipulation of flowering time in *Populus tremula* L. has been achieved at the Thünen Institute of Forest Genetics via an *Agrobacterium*-mediated system (Hoenicka et al., 2016) utilizing the construct HSP::AtFT with the *A. thaliana FT* gene. Short- and long read sequencing of two transgenic early-flowing poplar lines confirmed vector-free single-locus integration of the T-DNA including the *FT* gene (Kersten et al., 2020). In this artificial early flowering aspen lines, the time to flowering is reduced to less than one year. The control of *FT* expression by the heat inducible HSP promoter derived from soybean (Hoenicka et al., 2012) allows flower induction upon a heat shock treatment (**Figure 5**).

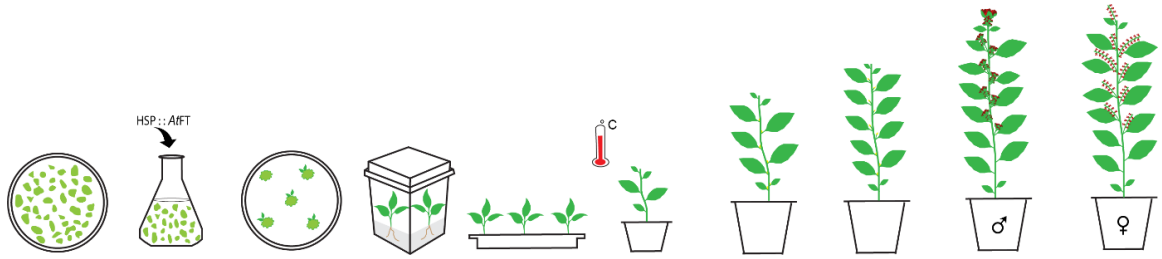


FIGURE 5. *AGROBACTERIUM*-MEDIATED TRANSFORMATION SYSTEM AND EARLY FLOWER INDUCTION IN POPLAR. AFTER *AGROBACTERIUM*-MEDIATED TRANSFORMATION AND REGENERATION, PLANTS ARE CULTIVATED *IN VITRO* UNTIL READY FOR TRANSPLANTATION INTO SOIL. UPON THE HEALTHY GROWTH OF THE PLANTS, A HEAT SHOCK TREATMENT (40 °C FOR 2 HOURS DAILY) IS APPLIED, AND FLOWERING STARTS AFTER ONE MONTH. AFTERWARDS, THE PLANTS MUST BE PLACED UNDER A COLD TREATMENT FOR PROPER FLOWER AND POLLEN DEVELOPMENT.

Genetics of sex determination in *Populus*

The mechanism of sex determination in the genus *Populus* remained unknown despite active investigation during the last one and a half decades. Genetic mapping revealed sex-linked markers consistently associated to the linkage group XIX (Gaudet et al., 2007;

Pakull et al., 2009; Paolucci et al., 2010; Yin et al., 2008) indicating the sex determination system in the genus *Populus* is genetically controlled by chromosome 19. With the employment of new genomic technologies, at least two different genomic architectures of sex determination were observed. While a relatively small (~100 kb) SDR is located at the very beginning of chromosome 19 in the balsam poplars *P. trichocarpa* and *P. balsamifera* L. (Geraldes et al., 2015; Yin et al., 2008), a much larger (~1 Mb) SDR resides within the centromeric region of chromosome 19 in the quaking aspen *P. tremuloides* Michx. and European aspen *P. tremula* (Kersten et al., 2014; Pakull et al., 2011, 2009; Müller et al., 2020). Whereas the aspens *P. tremula* and *P. tremuloides* share the same genetic architecture of sex determination in a XY system (Pakull et al., 2009), the white poplar *P. alba* L. sex-linked region was suggested to exhibit yet another genomic locus and a different heterogametic system, i.e. a ZW system (Paolucci et al., 2010).

Among the efforts to elucidate the sex-determining system in *Populus*, the sex-linked region in *P. tremuloides* was screened for potential candidate genes for sex determination (Kersten et al., 2014). A total of seven candidate genes were identified. Among those, a homologue of the *P. trichocarpa* gene (Potri.019G047300) showing similarity with *TORMOZEMBRYO DEFECTIVE* of *A. thaliana* captured attention due to its predicted role in flower development (Kersten et al., 2014; Pakull et al., 2015). Next generation sequencing identified it as a male-specific gene, located within the SDR. This sex-linked gene, named *TOZ19*, is missing completely (*P. tremuloides*) or partially (*P. tremula*) in females (Pakull et al., 2015). It should be noted that the *TOZ19* gene is a pseudogene in the aspens. Several premature stop codons and a large truncation preclude any coding function, therefore indicating that it is not directly involved in the process of sex determination. Nevertheless, the *TOZ19* gene has been validated as a marker to identify sex in aspens (Pakull et al., 2015).

Genome-wide association studies in *P. balsamifera and P. trichocarpa*, also searching for potential candidate genes underlying sex determination in *Populus,* revealed a candidate gene with a high number of sex-linked single nucleotide polymorphisms (SNPs) (Geraldes et al., 2015; McKown et al., 2017). The gene (Potri.019G133600) is homologous to the *A. thaliana* gene pair *ARABIDOPSIS RESPONSE REGULATOR 16/17* (*ARR16*/*ARR17*), further referred just as ***ARR17***, which is annotated as a type-A response

regulator supposedly involved in cytokinin signaling (Geraldes et al., 2015; Kieber and Schaller, 2018; To et al., 2004). Intriguingly, in *P. trichocarpa* the position of the SDR at the beginning of the chromosome did not match the position of the *ARR17* gene, located at the end (Geraldes et al., 2015). Preliminary analyses of short-read sequencing data from a *TOZ19*-probed bacterial artificial chromosome (BAC) clone of one male aspen individual (Kersten et al., 2017) and of existing data in balsam poplars revealed additional partial *ARR17* duplicates on the Y chromosome in these species. Additionally, initial data analysis from pool-seq data of male and female white poplars indicated that males did not exhibit any *ARR17* sequences in contrast to females, suggesting that *ARR17* might be a feminizing factor. But then, how do the additional partial *ARR17*duplicates in male aspen and balsam poplars elicit male flower development? Is *ARR17* a master sex determinant shared across the genus *Populus*? And what is the function of *ARR17*?

Objectives and hypotheses

In addition to representing an exciting model system for the evolution of sex determination, as presented in the previous sections, the collective genetic, genomic, and biotechnological amenability of *Populus* is absolutely unique among dioecious plant species. Therefore, it offers exceptional possibilities for studying the evolution of sex determination, not only because dioecy occurs throughout the entire genus but also because it allows to unambiguously test the hypothesis of a single gene-based mechanism. The dissection of the genetic mechanisms underlying sex determination in *Populus* will allow several evolutionary, developmental, and economic questions to be resolved. Therefore, the aim of this thesis was to explore the genes underlying dioecy in *Populus* and to link the high-level regulators to the underlying sex-differentiating gene network(s). The specific aims were to i) resolve the genetics underlying sex determination in aspen and white poplar, ii) clarify the role of *ARR17* in dioecy across the genus *Populus,* iii) identify sex-specific differential expression potentially pinpointing the sex-differentiating gene network(s) and iv) develop genetic markers for different *Populus* species.

In particular, the following hypotheses were tested:

(i) Y-specific partial *ARR17* duplicates are responsible for sex determination in aspen and balsam poplars.

(ii) Sex determination in white poplar is based on a ZW system with a Z-specific deletion of *ARR17*.

(iii) *ARR17* is a feminizing factor in aspen.

(iv) *ARR17* is a master sex switch shared across the genus *Populus*.

(v) Sex-specific expression networks will provide first clues on the molecular connection of *ARR17* to floral phenotypes.

(vi) The SDR of white poplar shows suppressed recombination and it is linked to a segregation distorter.

(vii) Genetic markers for sex determination in *Populus* can be developed based on W-specific *ARR17* in ZW systems and on Y-specific partial *ARR17* duplicates in XY systems.

In short, **Chapter 1** offers a broad context review of the current literature concerning the sex-determining genes and mechanisms in various dioecious plants. **Chapter 2** demonstrates the experimental validation of *ARR17* as a master regulator of sex determination in poplar. **Chapter 3** explores possible recombination suppression along the white poplar SDR. **Chapter 4** reports an example of convergent evolution of sex determination in *Populus* and demonstrates the application of the W-specific *ARR17* as molecular sex marker in ZW systems while the Y-specific partial *ARR17* duplicates can be employed in a XY system. Finally, **Chapter 5** investigates the molecular function of *ARR17*, and the genetic networks connected to the floral phenotype.

In summary, dioecy doesn't always require two segregating alleles linked together in the SDR. A segregating mutation in a single gene suffices for dioecy to be expressed. In poplars, a single gene, annotated as an A-type response regulator named *ARR17*, controls sex determination in a switch-like manner. Additionally, *ARR17* has a highly targeted function on specifying floral organ identity. When *ARR17* is active, female development is initiated, while, if *ARR17* is inactive, B class MADS-box genes are expressed, and male development occurs.

Chapter 1

The diversity and dynamics of sex determination in dioecious plants

Ana P. Leite Montalvão, Birgit Kersten*, Matthias Fladung & Niels A. Müller

Thünen Institute of Forest Genetics, Großhansdorf, Germany
* Corresponding author

Published *in Frontiers in Plant Science* (2021)
doi: 10.3389/fpls.2020.580488

Abstract

The diversity of inflorescences among flowering plants is captivating. Such charm is not only due to the variety of sizes, shapes, colors, and flowers displayed, but also to the range of reproductive systems. For instance, hermaphrodites occur abundantly throughout the plant kingdom with both stamens and carpels within the same flower. Nevertheless, 10% of flowering plants have separate unisexual flowers, either in different locations of the same individual (monoecy) or on different individuals (dioecy). Despite their rarity, dioecious plants provide an excellent opportunity to investigate the mechanisms involved in sex expression and the evolution of sex-determining regions (SDRs) and sex chromosomes. The SDRs and the evolution of dioecy have been studied in many species ranging from Ginkgo to important fruit crops. Some of these studies, for example in asparagus or kiwifruit, identified two sex-determining genes within the non-recombining SDR and may thus be consistent with the classical model for the evolution of dioecy from hermaphroditism via gynodioecy, that predicts two successive mutations, the first one affecting male and the second one female function, becoming linked in a region of suppressed recombination. On the other hand, aided by genome sequencing and gene editing, single factor sex determination has emerged in other species, such as persimmon or poplar. Despite the diversity of sex-determining mechanisms, a tentative comparative analysis of the known sex-determining genes and candidates in different species suggests that similar genes and pathways may be employed repeatedly for the evolution of dioecy. The cytokinin signaling pathway appears important for sex determination in several species regardless of the underlying genetic system. Additionally, tapetum-related genes often seem to act as male-promoting factors when sex is determined via two genes. We present a unified model that synthesizes the genetic networks of sex determination in monoecious and dioecious plants and will support the generation of hypothesis regarding candidate sex determinants in future studies.

Keywords: dioecy, monoecy, hermaphroditism, sex determination via one gene, sex determination via two genes, sex chromosomes, sex-determining region

Introduction

Contrary to most animals, hermaphroditism occurs widely in plants (Westergaard 1958; Bawa 1980; Renner 2014). However, the separation of the sexes, either in different locations of the same individual (monoecy), or in different individuals (dioecy) has been recognized in varying frequencies across the numerous plant species (Renner and Ricklefs 1995; Renner 2014). Unisexual gametophytes are widespread within the bryophyte lineages, with 68% of mosses, 57% of liverwort and 40% of hornwort species (Villarreal and Renner 2013). Among the seed plants there is a striking discrepancy: while in gymnosperms 65% of the species are dioecious (Walas et al. 2018), in angiosperms, dioecy is a comparatively uncommon phenomenon, comprising only 5-6% of all species (Renner 2014). However, even though dioecy is considered rare among flowering plants, its occurrence has been reported in several phylogenetic taxa (around 15,600 species spread over 175 families and 987 genera), suggesting that its evolution occurred independently hundreds if not thousands of times (Westergaard 1958; Renner 2014).

In humans, mammals, some insects and several plants, the males are heterogametic, meaning that males carry two different sex chromosomes denoted as X and Y, while females carry two X chromosomes. In many insects, the X:O system exists, where females carry two X chromosomes but the males only one. The second sex chromosome is absent in males. In birds, some reptiles and a few plants, the females are the heterogametic sex, represented by ZW, while males are ZZ. In bryophytes, with a predominant haploid phase, the male sex chromosome is referred to as U while the female is referred to as V (Bachtrog et al. 2014; Capel 2017; Renner et al. 2017). In contrast to animals, sex chromosomes in plants have been identified in very few species to date (Ming et al. 2011; Renner 2014; Muyle et al. 2017), especially because of the low number of species with heteromorphic sex chromosomes. Figure 1 presents dioecious plant species with cytogenetic and/or molecular evidence for the presence of sex chromosomes and their sex determination systems in a phylogenetic perspective. Among these plant species, male heterogamety (XY) is predominant (84.7%), while female heterogamety (ZW) only comprises 15.3%.

Interestingly, turnovers in the heterogametic systems have been reported in several plant families and genera. For instance, the genera *Dioscorea, Populus, Salix* and *Silene*

all comprise species with XY and ZW systems (Figure 1). One theory suggested that transitions between heterogametic systems may be selected when the sex chromosome is degenerated and the heterozygous sex has low fitness (Blaser et al. 2014), and another possibility is when sexually antagonistic polymorphism is maintained on an autosome, a new sex-determinant that arises in the region becomes advantageous (Doorn and Kirkpatrick 2007). Recent findings indicated that these transitions among heterogametic systems can be favored by haploid selection (Scott et al. 2018). Shifts between different sex chromosome systems (XY ↔ ZW) are also well documented across several clades of animal species and appear to be a common phenomenon, especially in reptiles, frogs and fishes (Vicoso et al. 2013; Bachtrog et al. 2014; Zhou et al. 2014; Jeffries et al. 2018; Pennell et al. 2018; Kottler et al. 2020).

This diversity of sexual systems in plants has an important impact on evolutionary biology as well as importance for seed and fruit production (Renner 2016). Despite the extensive amount of theoretical work regarding the evolution of dioecy and the possible resulting sex-determining systems, experimental data are only recently becoming available. This review aims to explore and synthesize the diversity of sex-determining mechanisms in several dioecious plant species. A unified model of sex determination is presented and possibly enriched functional categories of sex-determining genes are highlighted.

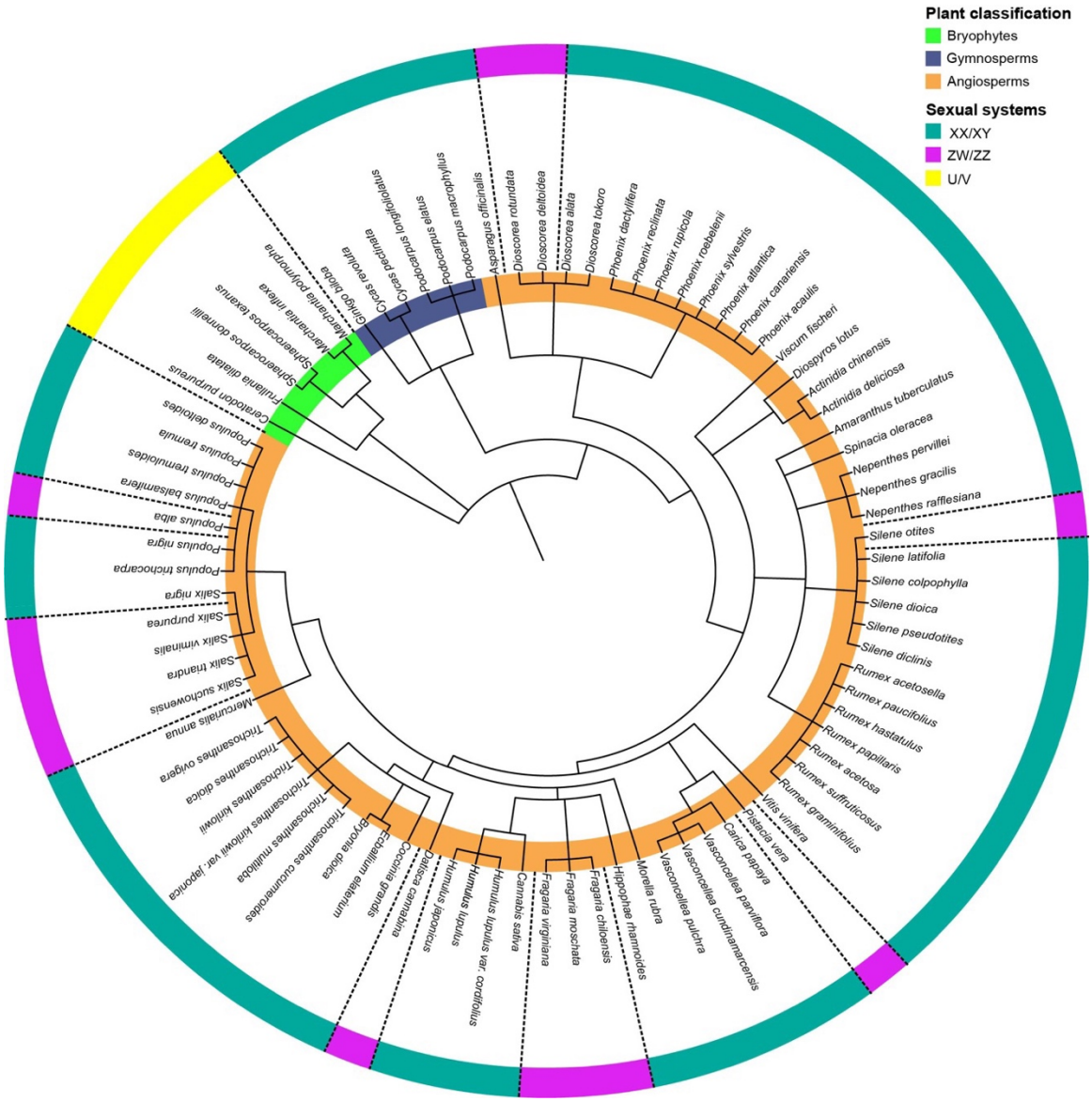


Figure 1. Phylogenetic tree of species with known sex determination systems and sex chromosomes. Plant classification (inner circle): bryophytes (green), gymnosperms (blue) and angiosperms (orange); Sexual systems (outer circle): male heterogametic system XX/XY (green), female heterogametic system ZW/ZZ (pink), haploid U/V system (yellow). The phylogenetic tree was built using phyloT v2: a tree generator (based on NCBI taxonomy)[1]. The list of the species with their respective references is presented in the Supplementary Table 1 (Table S1).

Genetic mechanisms of sex determination

The numerous independent evolutionary origins of dioecy suggest that many different genetic and molecular mechanisms are involved in the separation of the sexes in plants (Diggle et al. 2011). Studies in diverse species aimed at clarifying these mechanisms and ultimately explaining the evolution of dioecy and switches between sexual systems. The

[1] https://phylot.biobyte.de/

recent advances in molecular techniques are making this increasingly possible. By identifying and functionally characterizing the genes underlying sex determination, possible evolutionary pathways to dioecy can be inferred.

For dioecy to evolve, either from a hermaphroditic or a monoecious state, at least two changes, i.e. mutations, are necessary. The most influential work proposing a model for the evolution of dioecy, (Charlesworth and Charlesworth 1978), concludes that the most likely evolutionary pathway from hermaphroditism to dioecy involves two successive mutations with a gynodioecious intermediate. First, a recessive male-sterility mutation gives rise to a gynodioecious population (co-existence of females and hermaphrodites). Second, a dominant female-sterility mutation, which needs to be linked with the first mutation into a region of suppressed recombination, results in dioecy. While it is stated that both mutations might occur in a single gene, the most likely outcome of the "gynodioecy pathway" is two sex-determining genes in the SDR, one regulating female floral organ development and the other one male floral organ development. On the other hand, a single-gene sex-determining system can evolve if the factors regulating female and male function are connected by an epistatic genetic interaction rather than physical linkage (Golenberg and West 2013).

We will first review studies that indicate two sex-determining genes at the SDR, followed by work indicating a single sex-determining gene.

Sex determination via two genes

Strong experimental support for two sex-determining genes at the SDR has been shown for kiwifruit (*Actinidia deliciosa)* and asparagus (*Asparagus officinalis)*. Additionally, in date palm (*Phoenix dactylifera)* and grapevine (*Vitis vinifera)* sex determination via two genes appears likely and may thus also be consistent with the "gynodioecy pathway".

Kiwifruit (*A. deliciosa*) is a major fruit crop with an XY system of sex determination. Studies in kiwifruit demonstrated that two genes are responsible for sex determination, one affecting ovule production, and another one the production of pollen. A male-specific type-C cytokinin response regulator called "*SHY GIRL*" (*SyGI*) was identified as a suppressor of feminization (Su^F). The model systems *Arabidopsis thaliana* and *Nicotiana*

tabacum were used to validate the functionality of this gene. Transgenic expression of *SyGI* resulted in a stable suppression of carpel development (Akagi et al. 2018). Subsequently, the male-promoting factor (M_1), called "*FRIENDLY BOY*" (*FrBy*) was identified as the second Y-encoded gene responsible for sex determination in kiwifruit, specifically for the development of androecia. This gene is related to the *MICROSPORE AND TAPETUM REGULATOR 1* (*MTR1*) protein family, which, in rice, contributes to tapetum degradation affecting male fertility (Tan et al. 2012). The function of this second gene was validated in model plants as well as in kiwifruit. The artificial introduction of the *FrBy* gene into a female kiwifruit cultivar resulted in hermaphrodites (Akagi et al. 2019).

Similar to kiwifruit, in garden asparagus (*A. officinalis*) two genes were identified as the sex-determining genes: one of which is the Y-specific *SUPPRESSOR OF FEMALE FUNCTION* (*SOFF*) gene, acting as suppressor of femaleness. Experimental validation was achieved using a gamma irradiation knockout that resulted in the conversion of males to hermaphrodites (Harkess et al. 2017). The *DEFECTIVE IN TAPETUM DEVELOPMENT AND FUNCTION 1* (*TDF1*), encoding a MYB transcription factor and expressed only in males, was recognized by different research groups as a strong candidate for sex determination acting as a promoter of male function (Harkess et al. 2017; Murase et al. 2017; Tsugama et al. 2017). Ethyl methanesulfonate (EMS) mutagenesis of *aspTDF1* resulted in the conversion of males to asexual neuters. The knockout of both genes (*SOFF* and *aspTDF1*) converted males to females (Harkess et al. 2020).

These results show functional evidence that two sexually antagonistic genes at the SDR are necessary to determine sex in asparagus and in kiwifruit. Both species reveal distinct male-promoting factors (*FrBy* in kiwifruit and *aspTDF1* in asparagus), yet both having functions in the tapetum which is essential for male flower fertility.

The date palm (*P. dactylifera*), an important commercial fruit crop, presents a male heterogametic system (XY), and all 14 known species from the genus *Phoenix* are dioecious (Cherif et al. 2016). Recent work has identified sex-linked markers and a sex-linked region of ~6 Mb (Hazzouri et al. 2019) although candidates for sex-determining genes remained unidentified until recently. Torres et al. (2018) uncovered male-specific sequences in 13 species of *Phoenix* whereas no unique female-specific sequences were observed. Candidate genes potentially involved in sex determination in *P. dactylifera*

were revealed with similarity to *CYTOCHROME P450* (*CYP450*), ortholog of *CYP703A3* from rice (*Oryza sativa*), *GLYCEROL-3-PHOSPHATE ACYLTRANSFERASE 3-LIKE* (*GPAT3-like*), an ortholog of *GPAT3* from *A. thaliana* and the gene *LONELY-GUY* (*LOG*). The identified genes have known functions in sexual development in other monocot species. Both *CYP* and *GPAT3-like* are expressed only in *Phoenix* males and seem to be important for male flower development and fertility. In rice, both *CYP703* and *GPAT3* are expressed in tapetal cells and have functions in pollen formation and anther development. The deletion of the homologs of *CYP703* in rice, maize, and *A. thaliana* (Morant et al. 2007; Yang et al. 2014) and of *GPAT3* in rice (Men et al. 2017) led to male sterility. *LOG* genes are a family of genes with an important role in cytokinin activation and a potential role for female flower development (Kurakawa et al. 2007). In rice, *LOG* mutants presented flowers without ovules (Yamaki et al. 2011). While the functionality of these genes in date palm remains to be tested, the data are consistent with sex determination via two genes.

All grapevines (*V. vinifera*) are dioecious, however, during domestication, humans have generated a hermaphroditic grapevine subspecies (*Vitis vinifera* ssp. *vinifera*) (Fechter et al. 2012; Coito et al. 2018). Different models have been proposed to explain the genetic basis of sex determination in grapevines, but only recently evidence was put together to help clarifying these hypotheses. A genetic map demonstrated the sex-determining region contains several genes with potential involvement in flower development (Fechter et al. 2012; Picq et al. 2014). Haplotype-resolved genomes of hermaphrodite, female and male grapevines finally resolved the sex-determining region which spans approximately 260 kb on chromosome 2 (Zhou et al. 2019; Massonnet et al. 2020). The gene content and variability were characterized, and candidate genes proposed. Of ten genes with female-specific single nucleotide polymorphisms (SNPs), the *INAPERTURATE POLLEN 1* (*INP1*) gene was revealed as a likely candidate for the male-promoting factor (Massonnet et al. 2020). In *A. thaliana*, *INP1* is necessary for fertile pollen (Dobritsa and Coerper 2012). The results also showed that all individuals with female flowers were homozygous for an eight bp deletion in *VviINP1* indicating that this may be the causal polymorphism leading to male-sterility. In contrast, all individuals with male flowers carried one functional and one non-functional copy of *VviINP1*. Convincing candidate genes for the dominant female suppressor include the *ADENINEPHOSPHORIBOSYL*

TRANSFERASE (*APRT3*), a cytokinin regulator (Coito et al. 2018; Badouin et al. 2020) and the transcription factor *YABBY3* (Massonnet et al. 2020) that belongs to a gene family previously implicated in the development of carpels in *A. thaliana* (Villanueva et al. 1999). While future studies are necessary to understand the specific roles and connections of these different factors, the current data provide strong evidence for sex determination via (at least) two genes.

Artificial generation of dioecy via monoecy

Dioecy was artificially engineered in the monoecious species maize (*Zea mays*) and melon (*Cucumis melo*). Nearly a century ago, two genetically interacting genes were identified to control sex expression in monoecious maize: the *TASSEL SEED* (*Ts*) gene, which is a female suppressor, and the *SILKLESS* (*Sk*) gene, which protects female floral organ development from the action of *Ts*. In a *sk* mutant background, a single segregating *ts* mutation could be employed for the artificial production of dioecious maize (Jones 1934). In the monoecious melon, a network of three genes controls sex expression (Boualem et al. 2008, 2015). *CmACS11* controls the development of pistillate flowers, just like *Sk* in maize. *CmWIP1* suppresses female flower development, just as *Ts* in maize. Finally, *CmACS7* represses male flower development. The *1-AMINOCYCLOPROPANE-1-CARBOXYLIC ACID SYNTHASE* (*ACS*) genes are key enzymes in ethylene synthesis in plants, while *WIP* is a zinc finger transcription factor (Appelhagen et al. 2010). Just as in maize, the artificial conversion from monoecy to dioecy was achieved by producing a population with a non-functional *acs11* gene and a segregating *wip1* mutation (Boualem et al. 2015). Both examples demonstrate a potential pathway for the evolution of dioecy that results in the fixation of one mutation and a single segregating mutation controlling sex determination (Muller 1932; Westergaard 1958; Golenberg and West 2013; Renner 2016; Cronk and Müller 2020).

Sex determination via one gene

In several non-plant taxa, a single regulator gene is the primary mechanism for sex determination. For instance, the *SEX-DETERMINING REGION Y* (*SRY*) is the master switch of sex determination in most mammals (Gubbay et al. 1990; Kashimada and Koopman

2010; Li et al. 2014). Additionally, one single gene, that is *DOUBLESEX AND MAB-3 RELATED TRANSCRIPTION FACTORS* (*DMRT*), can control sex in several different groups of animals (e.g. birds, fish, frogs) (Hodgkin 2002; Smith et al. 2009).

The diversity of sexual systems in plants indicates various mechanisms, including sex determination via a single sex switch (female ↔ male). Experimental support for such switches in dioecious plant species was provided by Akagi et al. (2014) in the Caucasian persimmon and by Müller et al. (2020) in *Populus* spp. (Figure 2).

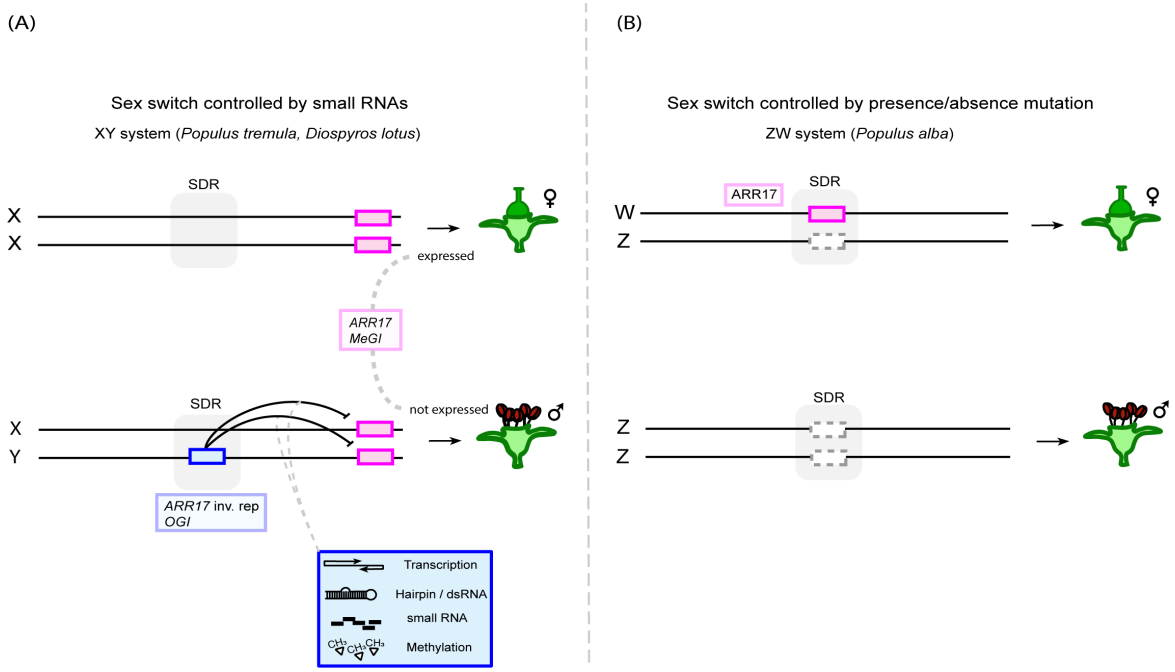


Figure 2. Single-gene sex determination enables turnovers between XY and ZW systems. (A) The feminizing sex switch (magenta box), i.e. *ARR17* and *MeGI* in poplar and persimmon, respectively, is located outside of the sex-determining region (SDR, indicated by grey shading), because a hairpin RNA-encoding Y-chromosomal sequence (blue box) controls its activity in trans via small RNAs, via a dominant repressing action. (B) Intriguingly, the same sex gene appears to be a sex-determining gene in a ZW system in white poplar (*P. alba*). Copies of the gene are found in the SDR of this species, and its absence from the Z chromosome (dashed grey box) leads to recessive loss of female and activation of male function.

The genus *Diospyros* contains approximately 475 species of which all are dioecious (Renner 2016). A male-specific sex-determining gene was described in the Caucasian persimmon (*Diospyros lotus*), that presents a male heterogametic system (XY) and a small SDR (Akagi et al. 2014). The male-specific *Oppressor of MeGI* (*OGI*) produces an RNA hairpin and, through a small RNA-based mechanism apparently causing DNA methylation, represses the autosomal *MALE GROWTH INHIBITOR* (*MeGI*) gene, allowing male development (Figure 2A; Table 1). Since no transformation protocol is available for

D. lotus, the functional characterization of *OGI* and *MeGI* was performed in *Nicotiana benthamiana, N. tabacum* and *A. thaliana*, showing that the overexpression of *OGI* suppressed the expression of *MeGI* and the overexpression of *MeGI* inhibited male flower function (Akagi et al. 2014). *MeGI* is a homolog of the *HOMEODOMAIN LEUCINE ZIPPER PROTEIN* (*Vrs1*), a gene from barley (*Hordeum vulgare*) (Komatsuda et al. 2007) that, when mutated, produces a group of three fertile flowers instead of a single central one. In persimmons, male flowers are composed by stamens, organized in groups of three, and non-developed carpels, suggesting the strong influence of this gene on sex determination and flower development (Akagi et al. 2016). Although experimental data only confirm the influence of the *OGI*/*MeGI* module on androecia and not gynoecia development, the *D. lotus* system has been suggested to function via the single *OGI*/*MeGI* sex switch. Recently, a gene network analysis identified the pathways regulating male and female sexual organ development. Furthermore, expression levels of cytokinin-related genes during gynoecium differentiation are correlated with *MeGI* expression levels (Yang et al. 2019). This suggests the cytokinin signaling pathway could play a role in the gynoecium differentiation in persimmon flowers.

In the genus *Populus*, the SDR was identified on chromosome 19 in different species (Gaudet et al. 2008; Paolucci et al. 2010; Kersten et al. 2014; Geraldes et al. 2015). Genome-wide association studies (GWAS) revealed the ortholog of the *ARABIDOPSIS RESPONSE REGULATOR 17* gene (*ARR17*) as a strong candidate for sex determination in the closely related balsam poplars *P. balsamifera* and *P. trichocarpa* (Geraldes et al. 2015; McKown et al. 2017). Further analysis identified partial duplicates of *ARR17* in the male-specific region of the Y chromosome (MSY) (Müller et al. 2020). Notably, these duplicates are present in aspens and balsam poplars, which represent two different sections of the genus, suggesting the possibility of a shared mechanism of sex determination. Despite these commonalities, phylogenetic analysis indicated that the sex-linked *ARR17* duplicates evolved independently (Müller et al. 2020; Zhou et al. 2020b). Long read sequencing showed that the partial duplicates are arranged as inverted repeats, giving rise to small RNAs and apparently causing DNA methylation and silencing of the *ARR17* gene, reminiscent of the *OGI*/*MeGI* system of persimmon (Figure 2A) (Bräutigam and Cronk 2018; Müller et al. 2020). Most importantly, the functionality of *ARR17* as a sex switch was demonstrated by CRISPR/Cas9-mediated *arr17* knockout in

early flowering aspen lines, reverting females to fully functional males (Müller et al. 2020). This complete sex reversal, rather than a reversion to hermaphrodites or neuters, demonstrates that *ARR17* functions as a single-gene sex switch.

Interestingly, white poplars (*P. alba*) present a female heterogametic system (ZW) (Paolucci et al. 2010). Long read sequencing and *de novo* assembly of a female white poplar identified a W chromosome-specific contig with three complete copies of *ARR17* (Müller et al. 2020). Male white poplars do not carry any *ARR17* sequences in their genome. Sex determination in white poplars thus appears to be based on a simple presence/absence mutation of *ARR17* (Figure 2B). Intriguingly, the single-gene-based mechanism of dioecy provides a simple and elegant means of changing the heterogametic system. The transition from a dominant suppressor, which acts in trans, to a recessive presence/absence mutation, which acts in cis, leads to the switch in the heterogametic system from XY to ZW (Figure 2) (Müller et al. 2020).

Unified model of sex determination

Interestingly, all currently known molecular mechanisms underlying sex determination can be represented in a single genetic network consisting of suppressors and promoters of femaleness and maleness connected either via genetic linkage or an epistatic genetic interaction (Figure 3). For monoecious species and dioecious species that determine sex via one gene, an additional high-level sex switch is needed. This switch (*ASC11* in melon, *ARR17* in poplar and *MeGI* in persimmon) represses a female suppressor (*WIP1* in melon) that in turn represses a male suppressor (*ACS7* in melon) by genetic interaction. In sex-determining systems via two genes, there is no genetic interaction, but the female suppressor and male promoter are linked into the sex-determining region. This unified model of plant sex determination emphasizes the differences and commonalities of the different systems and highlights the reason why one might hypothesize that single-gene sex-determining systems could be common in dioecious species that evolved via monoecy.

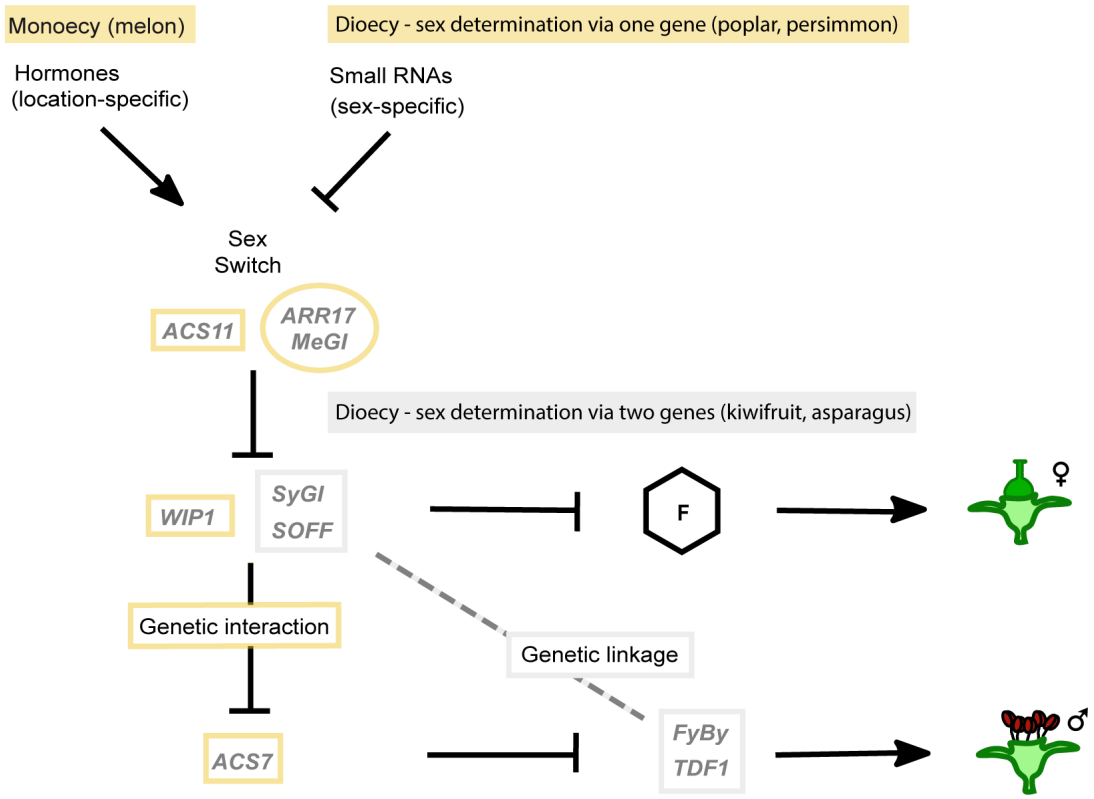


Figure 3. Unified model of sex determination in monoecious and dioecious plants. All experimentally proven systems of sex determination can be combined into a single genetic network comprising female and male suppressors, female and male promoters and high-level sex switches. In the "sex switch systems", i.e. monoecy and sex determination via one gene, indicated in yellow, an epistatic genetic interaction connects the development of female and male flowers making them in principle mutually exclusive. In sex determination via two genes ("gynodioecy model") indicated in grey, instead of a genetic interaction, the linkage between the female suppressor (*SOFF* in asparagus and *SyGI* in kiwifruit) and a male promoter (*TDF1* in asparagus and *FrBy* in kiwifruit) are essential for connecting female and male flower development. Female and male floral organs are controlled independently, therefore requiring genetic linkage to avoid hermaphrodites or asexual neuters. Female promoters, represented by "F", have not been identified for any of the studied species to date.

Experimental findings in selected other species

Sex determination has been studied in many other plant species. In the following paragraphs we review studies that provide robust data on sex-determining regions and potential candidate genes.

The early-diverged lineage of land plants represented by *Ginkgo biloba* exhibits an ancient origin of dioecy. Recent data indicate a male heterogametic system (XY) (Zhang et al. 2019). The species has a large (10.61 Gb) and repetitive genome (Guan et al. 2016), which was assembled into 12 chromosomes with approximately 9 Gb (Zhang et al. 2019).

A large region (~4.6 Mb) on chromosome 2 was identified as the SDR. Within this region, 16 protein-coding genes were found, and from those, five were suggested as potential sex-determining genes mainly due to their connection to previously reported functions in sex determination in other species: homologs of *RESPONSE REGULATOR 12* (*RR12*) and *RESPONSE REGULATOR 2* (*RR2*) (Gb_15883 and Gb_15884, respectively), both previously reported to be related to cytokinin and sex determination in kiwifruit (Akagi et al. 2018); a homolog of *EARLY FLOWERING 6* (*ELF6*) (Gb_15885), which encodes a H3K4 demethylase involved in the regulation of flowering time; a homolog of *BRASSINOSTEROID-RELATED ACYLTRANSFERASE 1* (*BAT1*) (Gb_15886), which regulates sex determination in maize (Hartwig et al. 2011); and a homolog of *AGAMOUS-like 6* (*AGL6*) (Gb_28587), reported to specify floral organ identity in rice (Li et al. 2010). To further explore sex determination in *G. biloba*, data of gene expression in different developmental stages should reveal clues to advance the investigations on this non-model gymnosperm species.

Fragaria spp. may display various sexual systems (e.g. hermaphrodite, gynodioecy, subdioecy and dioecy) across the species providing a great opportunity for new insights into the evolution of unisexuality in plants. The dioecious octoploid wild strawberry (*F. virginiana*) was characterized as having homomorphic sex chromosomes with a female heterogametic system ZW (Spigler et al. 2008; Goldberg et al. 2010; Tennessen et al. 2016). Strikingly, a small female-specific sequence (13 kb) was recognized as the SDR "cassette", which is located at different genomic positions in three related *Fragaria* species (Wei et al. 2017; Tennessen et al. 2018). The translocation of the SDR cassette demonstrates a possible way of sex chromosome turnover (Wei et al. 2017; Tennessen et al. 2018). Interestingly, only two protein-coding genes, *GMEW* (*GDP-mannose 3,5-epimerase 2, GME)* and *RPP0W (60S acidic ribosomal protein P0, RPP0*), were found in this "cassette". Nevertheless, it remains unclear how these candidate genes act in sex determination (Tennessen et al. 2018). Moreover, the SDR "cassette" might only control male function, while female function is controlled by a second locus (Spigler et al. 2008).

In willow (*Salix* spp.), the SDR was identified on chromosome 15 with female heterogamety (ZW) in *S. viminalis* (Pucholt et al. 2015), *S. suchowensis* (Hou et al. 2015; Chen et al. 2016), *S. purpurea* (Zhou et al. 2018) and *S. triandra* (Li et al. 2020). A recent

study revealed large palindromic structures on the W chromosome of *S. purpurea* and an ortholog of *ARR17* (*Salix purpurea RESPONSE REGULATOR* 9, *SpRR9*) was suggested as a strong candidate gene for sex determination (Zhou et al. 2020a). In contrast, in another species, *S. nigra,* a relatively small SDR (~2 Mb) was identified on chromosome 7 presenting a male heterogametic system (XY) (Sanderson et al. 2020). The underlying mechanisms for sex determination in *Salix* remain unclear; however, there is a possibility of a shared mechanism of sex determination despite the dynamic turnover of sex chromosomes in Salicaceae species.

Sex determination has also been investigated in *Nepenthes* pitcher plants (Scharmann et al. 2019). The species of this genus are all dioecious and carnivorous. Based on wild populations of males and females of three different species (*N. pervillei, N. gracilis, N. rafflesiana*), data supporting a male heterogametic system (XY) were presented. Two expressed sex-linked genes were identified: the homologs of the *A. thaliana* genes *DYSFUNCTIONAL TAPETUM 1* (*DYT1*) and *SEPALLATA 1* (*SEP1*); The first with important role in tapetum development and pollen fertility and the second as a regulator of floral organ identity. The *DYT1* gene functions in the tapetum, similar to the male-promoting genes in kiwifruit and asparagus. This opens the possibility of sex determination via two genes, where *DYT1* could function as the male-promoting factor.

Silene latifolia, (white campion), is a widely studied species and a model for studying sex chromosome evolution. It presents heteromorphic sex chromosomes and a male heterogametic system (XY) (Blackburn 1923; Bernasconi et al. 2009; Kejnovsky and Vyskot 2010; Muyle et al. 2012). Over the years, several genes have been discussed as potential sex determining factors: *S. latifolia X/Y-gene 1* (*SlX/Y1*), encoding a WD-repeat protein and likely involved in cell proliferation and *SlX/Y4*, encoding a fructose-2,6-bisphosphatase (Atanassov et al. 2001); the floral organ identity gene *APETALA 3* (*SlAP3*) (Cegan et al. 2010)' [W]hich is specifically involved in the development of androecia, and orthologs of *SHOOT MERISTEMLESS* (*STM*) (named *SlSTM1* and *SlSTM2*) and *CUP-SHAPED COTYLEDON 1* (*CUC1*) and *CUC2* (denoted as *SlCUC*) (Zluvova et al. 2006), both activators of cytokinin biosynthesis (Yang et al. 2019). The function of either of these genes remains to be tested. Recent deletion mapping in *Silene* (Kazama et al. 2016) improved

the locations of the sex-determining loci on the Y chromosome and could help to identify candidate sex-determining genes for further testing.

For details including other species, Supplementary Table 1 presents a more complete list with respective references.

Summary of sex-determining genes in different dioecious plant species

Although there are still numerous species of which the molecular and physiological mechanisms of sex determination remain elusive, the recent progress described above is remarkable and finally allows first comparative analyses. Despite multiple origins of dioecy, there might still be similar genes and pathways employed repeatedly for the independent evolution of dioecy. Such similarities can only be identified now that several systems can be analyzed together.

Interestingly, a first tentative analysis revealed a remarkably high number of genes involved in cytokinin signaling (Table 1; Figure 4). Cytokinin is a plant hormone known to be important for gynoecium formation (Marsch-Martinez et al. 2012). This also becomes evident by exogenous application of cytokinin, which can initiate carpel development in several species, including grapes (Wang et al. 2013), persimmon (Yonemori et al. 1993) and kiwifruit (Akagi et al. 2018). In this regard it is noteworthy that in monoecious species plant hormones play an essential role in sex determination (West and Golenberg 2018). For instance, in maize, flower development is connected to the jasmonic acid signaling pathway (Acosta et al. 2009), whereas in melon, ethylene appears to be the major player (Boualem et al. 2015). In dioecious species, hormones appear to influence sexual differentiation as well. For example, the sex switch *ARR17* in poplar is a type-A cytokinin response regulator (Müller et al. 2020). Moreover, a network analysis identified the cytokinin pathway as an important component of flower development in persimmon (Yang et al. 2019). Interestingly, in species that determine sex via two genes, several of the genes encoding suppressors of female development (Su^F) are involved in cytokinin signaling as well. These genes include, among others, *SyGI* in kiwifruit, which encodes a cytokinin response regulator (Akagi et al. 2018), the candidate gene *LOG* in date palm (Torres et al. 2018), which encodes a cytokinin-activating enzyme, and the

candidate gene *APRT3* in grapes (Badouin et al. 2020), which encodes an enzyme involved in the inactivation of cytokinin. Finally, two cytokinin response regulators are located in the SDR of Ginkgo (Zhang et al. 2019). Together these data strongly suggest an important role of cytokinin signaling for sex determination (Figure 4). Despite the single-gene-based mechanism of sex determination in persimmon, it might be interesting to note that a cytokinin response regulator is located in the persimmon SDR as well (Akagi et al. 2014). Notably, the unified model (Figure 3) already indicates that the female suppressors may be closer to the genetic pathways found in monoecious species, thus explaining the potential excess of hormone-related genes.

The male-promoting factors, on the other hand, appear to include several genes that act in tapetum formation and are thus much more directly involved in floral organ development (Table 1; Figure 4), which again appears consistent with the unified model of sex determination (Figure 3). The genes encoding the male promoters in kiwifruit and asparagus, i.e. *FyBy* and *TDF1*, respectively, and the candidate genes *CYP* and *GPAT3* in date palm and *DYT1* in *Nepenthes* all function in the tapetum potentially influencing male fertility. The putative male promoter in grapes (*VviINP1*) functions in pollen development and is thus also very directly involved in floral organ functioning, rather than in more general hormone signaling.

Despite the small number of candidate sex-determining genes identified to date, there appear to be some overarching patterns. In species that determine sex via two genes, the male-promoting factors seem to act specifically in the androecium to allow functional male floral organ development. From the other candidate genes, including female suppressors and sex switches, several seem to function in similar hormone response pathways. The many cytokinin-related genes in different dioecious species (Table 1; Figure 4), suggest that cytokinin signaling may be especially important in the regulation of sex determination predestinating it for the evolution of dioecy. It will be exciting to extend the comparison with further sex-determining genes in the future to assess whether these first generalizations remain valid.

Table 1. Candidate genes for sex determination with or without functional validation in different dioecious plant species.

Species	Original gene name	Function of ortholog/original name	Species of ortholog	Reference
Actinidia deliciosa	Shy girl	*ARABIDOPSIS RESPONSE REGULATOR 24 (ARR24)*	*Arabidopsis thaliana*	Akagi et al. 2018, 2019
	Friendly boy	*FAS1 DOMAIN PROTEIN (FAS1)*	*Nicotiana benthaminana*	
Asparagus officinalis	*SOFF*	*DOMAIN OF UNKNOWN FUNCTION 247 (DUF247)*	*Medicago truncatula*	Harkess et al. 2017, 2020
	aspTDF1	*DEFECTIVE IN TAPETAL DEVELOPMENT AND FUNCTION 1 (TDF1)*	*Arabidopsis thaliana*	
Diospyros lotus	*MeGI*	*HOMEODOMAIN LEUCINE ZIPPER CLASS (HD-Zip I) PROTEIN (HB40)*	*Hordeum vulgare*	Akagi et al. 2014
Fragaria virginiana	*GMEW*	*GDP-MANNOSE 3,5-EPIMERASE 2 (GMEW)*	Not applicable	Tennessen et al. 2018
Fragaria chiloensis	*RPP0W*	*60S ACIDIC RIBOSOMAL PROTEIN P0 (RPP0W)*	Not applicable	
Ginkgo biloba	gb_15883	*ARABIDOPSIS RESPONSE REGULATOR 12 (ARR12)*	*Arabidopsis thaliana*	Zhang et al. 2019
	gb_15884	*ARABIDOPSIS RESPONSE REGULATOR 2 (ARR2)*	*Arabidopsis thaliana*	
	gb_15885	*EARLY FLOWERING 6 (ELF6)*	*Arabidopsis thaliana*	
	gb_15886	*BRASSINOSTEROID-RELATED ACYLTRANSFERASE1 (BAT 1)*	*Arabidopsis thaliana*	
	gb_28587	*AGAMOUS-LIKE 6 (AGL6)*	*Arabidopsis thaliana*	
Nepenthes gracilis	*DYT1*	*DYSFUNCTIONAL TAPETUM 1 (DYT1)*	*Arabidopsis thaliana*	Scharmann et al. 2019
Nepenthes rafflesiana	*SEP 1*	*SEPALLATA1 (SEP1)*	*Arabidopsis thaliana*	
Nepenthes pervillei				
Phoenix dactylifera	*CYP*	*CYTOCHROME P450 HYDROXYLASE (CYP703A3)*	*Oryza sativa*	Torres et al. 2018
	GPAT3-like	*GLYCEROL-3-PHOSPHATE ACYLTRANSFERASE 3-LIKE (GPAT3)*	*Oryza sativa*	
	LOG	*LONELY GUY (LOG)*	*Oryza sativa*	
Populus tremula	*ARR17*	*ARABIDOPSIS RESPONSE REGULATOR 17 (ARR17)*	*Arabidopsis thaliana*	Müller et al. 2020
Populus alba				
Silene latifolia	*SlAP3*	*APETALA 3 (AP3)*	*Arabidopsis thaliana*	Cegan et al. 2010
	SlSTM	*SHOOT MERISTEMLESS (STM)*	*Arabidopsis thaliana*	Zluvova et al 2006
	SlCUC	*CUP-SHAPED COTYLEDON 1 and 2 (CUC1 / CUC2)*	*Arabidopsis thaliana*	
Salix purpurea	*SpRR9 (ARR17)*	*ARABIDOPSIS RESPONSE REGULATOR 17 (ARR17)*	*Arabidopsis thaliana*	Zhou et al. 2020
Vitis vinifera	*VviINP1*	*INAPERTURATE POLLEN 1 (INP1)*	*Arabidopsis thaliana*	Massonnet et al. 2020
	APRT3	*ADENINEPHOSPHORIBOSYL TRANSFERASE (APT3)*	*Arabidopsis thaliana*	
	VviYABBY3	*YABBY DOMAIN CLASS TRANSCRIPTION FACTOR (YAB1)*	*Arabidopsis thaliana*	

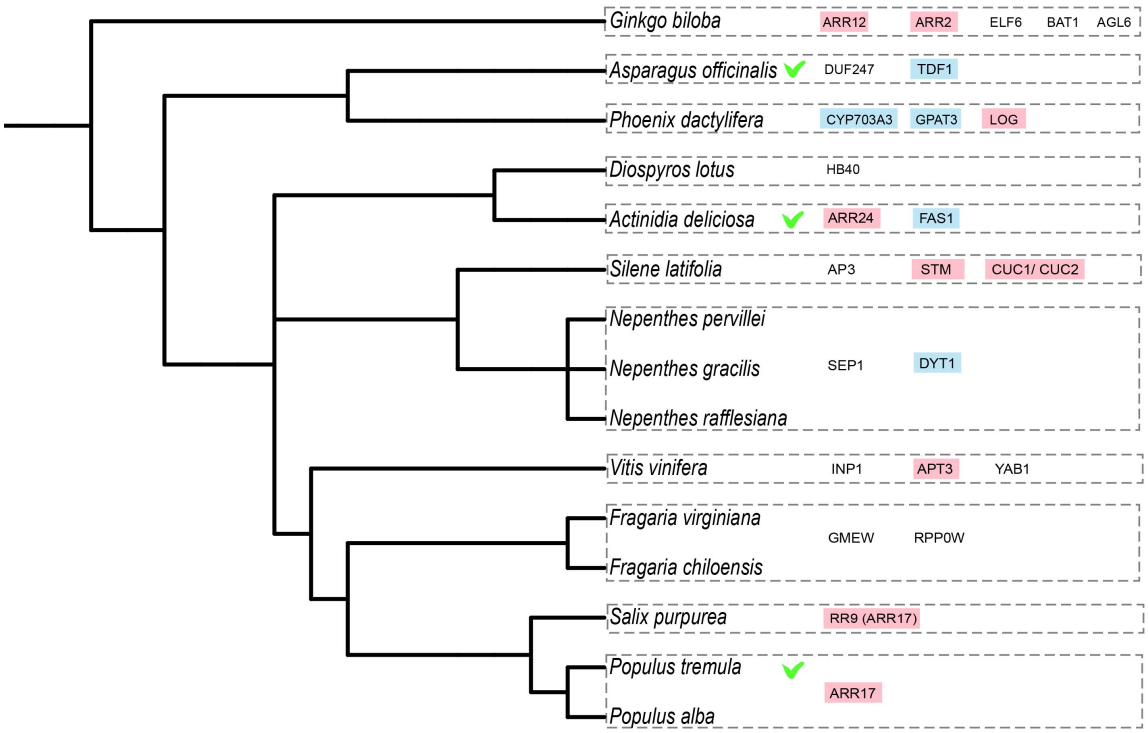


Figure 4. Evolutionary relationship of dioecious species with known sex-determining genes or strong candidate genes for sex determination (based on Table 1). Marked with a green tick the species with genes that were functionally validated in the species, in pink the cytokinin-related genes and in blue the tapetum-related genes. The phylogenetic tree was built using phyloT v2: a tree generator (based on NCBI taxonomy)[1].

Conclusions and outlook

Dioecy has evolved numerous times independently. Despite elegant theoretical models for the evolution of dioecy (Charlesworth and Charlesworth 1978; Renner 2016), only recently, powerful experimental work is providing empirical data for further assessing different possible trajectories (Harkess et al. 2017; Akagi et al. 2019; Harkess et al. 2020; Müller et al. 2020). These data highlight the diversity of sex-determining mechanisms and emphasize the need for considering more than just one theoretical model.

A first tentative comparative analysis of sex-determining and candidate genes in different dioecious species suggests that similar genes and pathways may be employed repeatedly for the independent evolution of dioecy. Cytokinin-related genes appear to be important in sex determination of several dioecious species, irrespective of whether

[1] https://phylot.biobyte.de/

sex is determined via one or two genes. Moreover, tapetum-related genes were identified as male-promoting factors in several two-gene systems.

The expanding number of studies related to sex chromosome evolution and sex-determining systems in crop plants may contribute to enhancing their agricultural value. Studies in model systems provide further important biological insights into chromosome evolution and the molecular mechanisms of flower development. Novel molecular techniques such as long read sequencing, transformation protocols or gene editing approaches are rapidly becoming available to support the identification of the sex-determining genes and the underlying genetic mechanisms leading to the evolution of dioecy. Thus, it is likely that several sex-determining systems will be resolved in the next couple of years. Furthermore, the pathway connecting these high-level regulators to floral phenotype is largely unknown, and work in this area is urgently required if we are to fully understand dioecy (Feng et al. 2020; Cronk and Müller, 2020). These data will allow exciting further generalizations and improve our understanding of the molecular control and the evolution of dioecy in flowering plants.

Conflict of interest

The authors declare that there are no conflicts of interest

Author contributions

Original draft: APLM. Inputs and revision from BK, MF, NAM.

Funding

We acknowledge funding from a grant of the Deutsche Forschungsgemeinschaft to NAM (DFG grant n° MU 4357/1-1) and core funding from the Thünen Institute.

Supplementary material

The Supplementary Information for this article can be found online at:

https://www.frontiersin.org/articles/10.3389/fpls.2020.580488/full#supplementary-material

References

Acosta, I.F., Laparra, H., Romero, S.P., Schmelz, E., Hamberg, M., Mottinger, J.P., Moreno, M.A., Dellaporta, S.L. (2009). *tasselseed1* is a lipoxygenase affecting jasmonic acid signaling in sex determination of maize. *Science*. 323 (5911): 262-265. https://doi.org/10.1126/science.1164645

Akagi, T., Henry, I.M.,Takashi K., Comai, L., Tao, R. (2016). Epigenetic regulation of the sex determination gene *MeGI* in polyploid persimmon. *Plant Cell*. 28 (12): 2905–2915. https://doi.org/10.1105/tpc.16.00532

Akagi, T., Henry, I.M., Ohtani, H., Morimoto, T., Beppu, K., Kataoka, I., Tao, R. (2018). A Y-encoded suppressor of feminization arose via lineage-specific duplication of a cytokinin response regulator in kiwifruit. *Plant Cell*. 30 (4): 780–795. https://doi.org/10.1105/tpc.17.00787

Akagi, T., Henry, I.M., Tao R., Comai, L. (2014). A Y-chromosome-encoded small RNA acts as sex determinant in persimmons. *Science*. 346 (6209): 646–650. https://doi.org/10.1126/science.1258705

Akagi, T., Pilkington, S.M., Varkonyi-Gasic, E., Henry, I.M., Sugano, S.S., Sonoda, M., et al. (2019). Two Y-chromosome-encoded genes determine sex in kiwifruit. *Nature Plants*. 5, 801-809. https://doi.org/10.1038/s41477-019-0489-6

Appelhagen, I., Huep, G., Lu, G.H., Strompen, G., Weisshaar, B., Sagasser, M. (2010). Weird fingers: functional analysis of *WIP* domain proteins. *FEBS Letters*. 584 (14): 3116–3122. https://doi.org/10.1016/j.febslet.2010.06.007

Atanassov, I., Delichère, C., Filatov, D.A., Charlesworth, D., Negrutiu, I., Monéger, F. (2001). Analysis and evolution of two functional Y-linked loci in a plant sex chromosome system. *Molecular Biology and Evolution*. 18 (12): 2162–2168. https://doi.org/10.1093/oxfordjournals.molbev.a003762

Bachtrog, D., Mank, J.E., Peichel, C.L., Kirkpatrick, M., Otto, S.P., Ashman, T.L., Hahn, M.W. et al. (2014). Sex determination: Why so many ways of doing it? *PLoS Biology*. 12 (7):e1001899. https://doi.org/10.1371/journal.pbio.1001899

Badouin, H., Velt, A., Gindraud, F., Flutre, T., Dumas, V., Vautrin, S., Marande, W. et al. (2020). The wild grape genome sequence provides insights into the transition from dioecy to hermaphroditism during grape domestication. *Genome biology* 21, 223. https://doi.org/10.1186/s13059-020-02131-y

Bawa, K.S. (1980). Evolution of dioecy in flowering plants. *Annual Review of Ecology and Systematics*. 11: 15-39. https://doi.org/10.1146/annurev.es.11.110180.000311

Bernasconi, G., Antonovics, J., Biere, A., Charlesworth, D., Delph, L.F., Filatov, D., Giraud, T. et al. (2009). *Silene* as a model system in ecology and evolution. *Heredity*. 103: 5-14. https://doi.org/10.1038/hdy.2009.34

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

Blaser, O., Neuenschwander, S., Perrin, N. (2014). Sex-chromosome turnovers: The hot-potato model. *The American Naturalist*. 183 (1): 140–146. https://doi.org/10.1086/674026

Boualem, A., Fergany, M., Fernandez, R., Troadec, C., Martin, A., Morin, H., Sari, M.A. et al. (2008). A conserved mutation in an ethylene biosynthesis enzyme leads to andromonoecy in melons. *Science*. 321(5890):836-838. https://doi.org/10.1126/science.1159023

Boualem, A., Troadec, C., Camps, C., Lemhemdi, A., Morin, H., Sari, M.A., Fraenkel-Zagouri, R. et al. (2015). A cucurbit androecy gene reveals how unisexual flowers develop and dioecy emerges. *Science*. 350 (6261): 688–691. https://doi.org/10.1126/science.aac8370

Bräutigam, K., and Cronk, Q. (2018). DNA methylation and the evolution of developmental complexity in plants. *Frontiers in Plant Science*. 9: 1–14. https://doi.org/10.3389/fpls.2018.01447

Capel, B. (2017). Vertebrate sex determination: evolutionary plasticity of a fundamental switch. *Nature Reviews Genetics*. 18: 675-689. https://doi.org/10.1038/nrg.2017.60

Cegan, R., Marais, G.A.B., Kubekova, H., Blavet, N., Widmer, A., Vyskot, B., Doležel, J., Šafář, J., Hobza, R.

(2010). Structure and evolution of *Apetala3*, a sex-linked gene in *Silene latifolia*. *BMC Plant Biology*. 10, 180. https://doi.org/10.1186/1471-2229-10-180
Charlesworth, D., and Charlesworth, B. (1978). Population genetics of partial male-sterility and the evolution of monoecy and dioecy. *Heredity*. 41: 137-153. https://doi.org/10.1038/hdy.1978.83
Chen, Y., Wang, T., Fang, L., Li, X., Yin, T. (2016). Confirmation of single-locus sex determination and female heterogamety in willow based on linkage analysis. *PLoS ONE*. 11 (2):e0147671. https://doi.org/10.1371/journal.pone.0147671
Cherif, E., S. Zehdi-Azouzi, A. Crabos, K. Castillo, N. Chabrillange, J. C. Pintaud, A. Salhi-Hannachi, S. Glémin, and F. Aberlenc-Bertossi. (2016). Evolution of sex chromosomes prior to speciation in the dioecious *Phoenix* species. *Journal of Evolutionary Biology*. 29 (8): 1513–1522. https://doi.org/10.1111/jeb.12887
Coito, J.L., Silva, H., Ramos, M.J.N., Montez, M., Cunha, J., Amâncio, S., Costa, M.M.R., and Rocheta, M. (2018). *Vitis* flower sex specification acts downstream and independently of the ABCDE model genes. *Frontiers in Plant Science*. 9:1029. https://doi.org/10.3389/fpls.2018.01029
Cronk, Q. and Müller, N.A. (2020). Default sex and single gene sex determination in dioecious plants. *Frontiers in Plant Science*. 11: 1162. https://doi.org/10.3389/fpls.2020.01162
Diggle, P.K., Di Stilio, V.S., Gschwend, A.R., Golenberg, E.M., Moore, R.C., Russel, J.R., Sinclair, J.P. (2011). Multiple developmental processes underlie sex differentiation in angiosperms. *Trends in Genetics*. 27 (9): 368-376. https://doi.org/10.1016/j.tig.2011.05.003
Dobritsa, A.A., and Coerper, D. (2012). The Novel plant protein *INAPERTURATE POLLEN 1* marks distinct cellular domains and controls formation of apertures in the *Arabidopsis* pollen exine. *Plant Cell*. 24 (11): 4452–4464. https://doi.org/10.1105/tpc.112.101220
Doorn, G.S.V., and Kirkpatrick, M. (2007). Turnover of sex chromosomes induced by sexual conflict. *Nature*. 449: 909-912. https://doi.org/10.1038/nature06178
Fechter, I., Hausmann, L., Daum, M., Sörensen, T.R., Viehöver, P., Weisshaar, B., Töpfer, R. (2012). Candidate genes within a 143 kb region of the flower sex locus in *Vitis*. *Molecular Genetics and Genomics*. 287: 247-259. https://doi.org/10.1007/s00438-012-0674-z
Feng, G., Sanderson, B.J., Keefover-Ring, K., Liu, J., Ma, T., Yin, T., Smart, L.B., DiFazio, S.P., Olson, M.S. (2020). Pathways to sex determination in plants: how many roads lead to Rome? *Current Opinion in Plant Biology*. 54:61-68. https://doi.org/10.1016/j.pbi.2020.01.004
Gaudet, M., Jorge, V., Paolucci, I., Beritognolo, I., Mugnozza, G.S., Sabatti, M. (2008). Genetic linkage maps of *Populus nigra* L. including AFLPs, SSRs, SNPs, and sex trait. *Tree Genetics and Genomes*. 4:25-36. https://doi.org/10.1007/s11295-007-0085-1
Geraldes, A., Hefer, C.A., Capron, A., Kolosova, N., Martinez-Nuñez, F., Soolanayakanahally, R.Y., Stanton, B. et al. (2015). Recent Y chromosome divergence despite ancient origin of dioecy in poplars (*Populus*). *Molecular Ecology*. 24 (3): 3243-3256. https://doi.org/10.1111/mec.13126
Goldberg, M.T., Spigler, R.B., Ashman, T.L. (2010). Comparative genetic mapping points to different sex chromosomes in sibling species of wild strawberry (*Fragaria*). *Genetics*. 186 (4): 1425–33. https://doi.org/10.1534/genetics.110.122911
Golenberg, E.M., and West, N.W. (2013). Hormonal interactions and gene regulation can link monoecy and environmental plasticity to the evolution of dioecy in plants. *American Journal of Botany*. 100 (6): 1022-1037. https://doi.org/10.3732/ajb.1200544
Guan, R., Zhao, Y., Zhang, H., Fan, G., Liu, X., Zhou, W., Shi, C. et al. (2016). Draft genome of the living fossil *Ginkgo biloba*. *GigaScience*. 5 (1): s13742–016–0154–1. https://doi.org/10.1186/s13742-016-0154-1
Gubbay, J., Collignon, J., Koopman, P., Capel, B., Economou, A., Münsterberg, A., Nigel, V., Goodfellow, P., Lovell-Badge, R. (1990). A gene mapping to the sex-determining region of the mouse Y chromosome is a member of a novel family of embryonically expressed genes. *Nature*. 346: 245-250. https://doi.org/10.1038/346245a0
Harkess, A., Huang, K., Van Der Hulst, R., Tissen, B., Caplan, J.L., Koppula, A., Batish, M., Meyers, B.C., Leebens-Mack, J. (2020). Sex determination by two Y-linked genes in garden asparagus. *Plant Cell*.32 (6): 1790-1796. https://doi.org/10.1105/tpc.19.00859
Harkess, A., Zhou, J., Xu, C., Bowers, J.E., Van Der Hulst, R., Ayyampalayam, S., Mercati, F. et al. (2017). The asparagus genome sheds light on the origin and evolution of a young Y chromosome. *Nature Communications*. 8, 1279. https://doi.org/10.1038/s41467-017-01064-8
Hartwig, T., Chuck, G.S., Fujioka, S., Klempien, A., Weizbauer, R., Potluri, D.P.V., Choe, S., Johal, G.S., Schulz, B. (2011). Brassinosteroid control of sex determination in maize. *Proceedings of the National Academy of Sciences of the United States of America*. 108 (49): 19814–19819. https://doi.org/10.1073/pnas.1108359108
Hazzouri, K.M., Gros-Balthazard, M., Flowers, J.M., Copetti, D., Lemansour, A., Lebrun, M. et al. (2019).

Genome-wide association mapping of date palm fruit traits. *Nature Communications.* 10, 4680. https://doi.org/10.1038/s41467-019-12604-9

Hodgkin, J. (2002). The remarkable ubiquity of DM domain factors as regulators of sexual phenotype: Ancestry or aptitude?. *Genes and Development*. 16: 2322-2326. https://doi.org/10.1101/gad.1025502

Hou, J., Ye, N., Zhang, D., Chen, Y., Fang, L., Dai, X., Yin, T. (2015). Different autosomes evolved into sex chromosomes in the sister genera of *Salix* and *Populus*. *Scientific Reports*. 5, 9076. https://doi.org/10.1038/srep09076

Jeffries, D.L., Lavanchy, G., Sermier, R., Sredl, M.J., Miura, I., Borzée, A., Barrow, L.N. et al. (2018). A rapid rate of sex-chromosome turnover and non-random transitions in true frogs. *Nature Communications*. 9, 4088. https://doi.org/10.1038/s41467-018-06517-2

Jones, D.F. (1934). Unisexual maize plants and their bearing on sex differentiation in other plants and in animals. *Genetics.*19 (6):552-567. Online ISSN 1943-2631.

Kashimada, K., and Koopman, P. (2010). Sry: The master switch in mammalian sex determination. *Development*. 137 (23): 3921–3930. https://doi.org/10.1242/dev.048983

Kazama, Y., Ishii, K., Aonuma, W. et al. (2016). A new physical mapping approach refines the sex-determining gene positions on the *Silene latifolia* Y-chromosome. Scientific Reports 6, 18917. https://doi.org/10.1038/srep18917

Kejnovsky, E., and Vyskot, B. (2010). *Silene latifolia*: The classical model to study heteromorphic sex chromosomes. *Cytogenetic and Genome Research*. 129 (1–3): 250–262. https://doi.org/10.1159/000314285

Kersten, B., Pakull, B., Groppe, K., Lueneburg, J., Fladung, M. (2014). The sex-linked region in *Populus tremuloides* Turesson 141 corresponds to a pericentromeric region of about two million base pairs on *P. trichocarpa* chromosome 19. *Plant Biology*. 16 (2):411-418. https://doi.org/10.1111/plb.12048

Komatsuda, T., Pourkheirandish, M., He, C., Azhaguvel, P., Kanamori, K., Perovic, D., Stein, N. et al. (2007). Six-rowed barley originated from a mutation in a homeodomain-leucine zipper I-class homeobox gene. *Proceedings of the National Academy of Sciences of the United States of America*. 104 (4):1424-1429. https://doi.org/10.1073/pnas.0608580104

Kottler, V.A., Feron, R., Nanda, I., Klopp, C., Du, K., Kneitz, S., Helmprobst, F. et al. (2020). Independent origin of XY and ZW sex determination mechanisms in mosquitofish sister species. *Genetics*. 214 (1): 193–209. https://doi.org/10.1534/genetics.119.302698

Kurakawa, T., Ueda, N., Maekawa, M. et al. (2007). Direct control of shoot meristem activity by a cytokinin-activating enzyme. *Nature* 445, 652–655. https://doi.org/10.1038/nature05504

Li, H., Liang, W., Jia, R., Yin, C., Zong, J., Kong, H., Zhang, D. (2010) The *AGL6*-like gene *OsMADS6* regulates floral organ and meristem identitites in rice. *Cell Research*, 20: 299-313. https://doi.org/10.1038/cr.2009.143

Li, W., Wu, H., Li, X., Chen, Y., Yin, T. (2020). Fine mapping of the sex locus in *Salix triandra* confirms a consistent sex determination mechanism in genus *Salix*. *Horticulture Research*. 7, 64. https://doi.org/10.1038/s41438-020-0289-1

Li, Y., Zheng, M., Lau, Y.F.C. (2014). The sex-determining factors *SRY* and *SOX9* regulate similar target genes and promote testis cord formation during testicular differentiation. *Cell Reports*. 3 (7):723-733. https://doi.org/10.1016/j.celrep.2014.06.055

Massonnet, M., Cochetel, N., Minio, A., Vondras, A.M., Lin J., Muyle, A. et al. (2020). The genetic basis of sex determination in grapes. *Nature communications*. 11, 2902. https://doi.org/10.1038/s41467-020-16700-z

Marsch-Martinez, N., Ramos-Cruz, D., Reyes-Olalde, I., Lozano-Sotomayor, P., Zuniga-Mayo, V.M., de Folter, S. (2012). The role of cytokinin during *Arabidopsis* gynoecia and fruit morphogenesis and patterning. The Plant Journal. 72, 222-234. https://doi.org/10.1111/j.1365-313X.2012.05062.x

McKown, A.D., Klápště, J., Guy, R.D., Soolanayakanahally, R.Y., La Mantia, J., Porth, I., Skyba, O. et al. (2017). Sexual homomorphism in dioecious trees: extensive tests fail to detect sexual dimorphism in *Populus*. *Scientific Reports*. 7, 1831. https://doi.org/10.1038/s41598-017-01893-z

Men, X., Shi, J., Liang, W., Zhang, Q., Lian, G., Quan, S., Zhu, L., Luo, Z., Chen, M., Zhang, D. (2017). *Glycerol-3-Phosphate Acyltransferase 3* (*OsGPAT3*) is required for anther development and male fertility in rice. *Journal of Experimental Botany*. 68 (3): 513–526. https://doi.org/10.1093/jxb/erw445

Ming, R., Bendahmane, A., Renner, S.S. (2011). Sex chromosomes in land plants. *Annual Review of Plant Biology*. 62:485-514. https://doi.org/10.1146/annurev-arplant-042110-103914

Morant, M., Jørgensen, K., Schaller, H., Pinot, F., Møller, B.L., Werck-Reichhart, D., Bak, S. (2007). *CYP703* is an ancient Cytochrome P450 in land plants catalyzing in-chain hydroxylation of lauric acid to provide building blocks for sporopollenin synthesis in pollen. *Plant Cell*. 19 (5): 1473–1487.

https://doi.org/10.1105/tpc.106.045948
Muller H. J., 1932 Some genetic aspects of sex. *The American Naturalist*. 66: 703. 118–138. https://doi.org/10.1086/280418
Müller, N.A., Kersten, B., Leite Montalvao, A.P., Mähler, N., Bernhardsson, C., Bräutigam, K., Lorenzo, Z.C. et al. (2020). A Single gene underlies the dynamic evolution of poplar sex determination. *Nature Plants*, 6:630-637. https://doi.org/http://doi.org/10.1038/s41477-020-0672-9
Murase, K., Shigenobu, S., Fujii, S., Ueda, K., Murata, T., Sakamoto, A. Wada, Y. et al. (2017). *MYB* transcription factor gene involved in sex determination in *Asparagus officinalis*. *Genes to Cells*. 22 (1):115-123. https://doi.org/10.1111/gtc.12453
Muyle, A., Shearn, R., Marais, G.A.B. (2017). The evolution of sex chromosomes and dosage compensation in plants. *Genome Biology and Evolution*. 9 (3):627-645. https://doi.org/10.1093/gbe/evw282
Muyle, A., Zemp, N., Deschamps, C., Mousset, S., Widmer, A., Marais, G.A.B. (2012). Rapid *de novo* evolution of X chromosome dosage compensation in *Silene latifolia*, a plant with young sex chromosomes. *PLoS Biology*. 10 (4):e1001308. https://doi.org/10.1371/journal.pbio.1001308
Paolucci, I.,Gaudet, M., Jorge, V., Beritognolo, I., Terzoli, S., Kuzminsky, E., Muleo, R., Mugnozza, G.S., Sabatti, M. (2010). Genetic linkage maps of *Populus alba* L. and comparative mapping analysis of sex determination across *Populus* species. *Tree Genetics & Genomes*. 6:863-875. https://doi.org/10.1007/s11295-010-0297-7
Pennell, M.W., Mank, J.E., Peichel, C.L. (2018). Transitions in sex determination and sex chromosomes across vertebrate species. *Molecular Ecology*. 27 (19):3950-3963. https://doi.org/10.1111/mec.14540
Picq, S., Santoni, S., Lacombe, T., Latreille, M., Weber, A., Ardisson, M., Ivorra, S. et al. (2014). A small XY chromosomal region explains sex determination in wild dioecious *V. vinifera* and the reversal to hermaphroditism in domesticated grapevines. *BMC Plant Biology*. 14, 229. https://doi.org/10.1186/s12870-014-0229-z
Pucholt, P., Rönnberg-Wästljung, A.C., Berlin, S. (2015). Single locus sex determination and female heterogamety in the basket willow (*Salix viminalis* L.). *Heredity*. 114:575–583. https://doi.org/10.1038/hdy.2014.125
Renner, S.S. (2014). The relative and absolute frequencies of angiosperm sexual systems: dioecy, monoecy, gynodioecy, and an updated online database. *American Journal of Botany*. 101 (10):1588-1596. https://doi.org/10.3732/ajb.1400196
Renner, S.S. (2016). Pathways for making unisexual flowers and unisexual plants: Moving beyond the 'two mutations linked on one chromosome' model. *American Journal of Botany*. 103 (4):587-589. https://doi.org/10.3732/ajb.1600029
Renner, S.S., Heinrichs, J., Sousa, A. (2017). The Sex chromosomes of bryophytes: Recent insights, open questions, and reinvestigations of *Frullania dilatata* and *Plagiochila asplenioides*. *Journal of Systematics and Evolution*. 55 (4): 333–339. https://doi.org/10.1111/jse.12266
Renner, S.S., and Ricklefs, R.E. (1995). Dioecy and its correlates in the flowering. *American Journal of Botany*. 82 (5): 596–606. https://doi.org/10.1002/j.1537-2197.1995.tb11504.x
Sanderson, B.J., Feng, G., Hu, N., Carlson, C.H., Smart, L.B., Keefover-Ring, K., Yin, T. et al. (2020). "Sex determination through X-Y heterogamety in *Salix nigra*." *BioRxiv* [Preprint]. Available at: https://doi.org/10.1101/2020.03.23.000919
Scharmann, M.T., Grafe, U., Metali, F., Widmer, A. (2019). Sex is determined by XY chromosomes across the radiation of dioecious *Nepenthes* pitcher plants. *Evolution Letters*. 3 (6): 586–597. https://doi.org/10.1002/evl3.142
Scott, M.F., Osmond, M.M., Otto, S.P. (2018). Haploid selection, sex ratio bias, and transitions between sex-determining systems. *PLoS Biology*. 16 (6): e2005609. https://doi.org/10.1371/journal.pbio.2005609
Smith, C.A., Roeszler, K.N., Ohnesorg, T., Cummins, D.M., Farlie, P.G., Doran, T.J., Sinclair, A.H. (2009). The avian Z-Linked gene *DMRT1* is required for male sex determination in the chicken. *Nature*. 461:267-271. https://doi.org/10.1038/nature08298
Spigler, R.B., Lewers, K.S., Main, D.S., Ashman, T.L. (2008). Genetic mapping of sex determination in a wild strawberry, *Fragaria virginiana*, reveals earliest form of sex chromosome. *Heredity*. 101:507–17. https://doi.org/10.1038/hdy.2008.100
Tan, H., Liang, W., Hu, J., and Zhang, D. (2012). MTR1 Encodes a Secretory Fasciclin Glycoprotein Required for Male Reproductive Development in Rice. Developmental Cell. 22, 1127–1137. doi: 10.1016/j.devcel.2012.04.011
Tennessen, J.A., Govindarajulu, R., Liston, A., Ashman, T.L. (2016). Homomorphic ZW chromosomes in a wild strawberry show distinctive recombination heterogeneity but a small sex-determining region.

The New Phytologist. 211 (4):1412-1423. https://doi.org/10.1111/nph.13983
Tennessen, J.A., Wei, N., Straub, S.C.K., Govindarajulu, R., Liston, A., Ashman, T.L. (2018). Repeated translocation of a gene cassette drives sex-chromosome turnover in strawberries. *PLoS Biology*. 16 (8): e2006062. https://doi.org/10.1371/journal.pbio.2006062
Torres, M.F., Mathew, L.S., Ahmed, I., Al-Azwani, I.K., Krueger, R., Rivera-Nuñez, D., Mohamoud, Y.A, Clark, A.G., Suhre, K., Malek, J.A. (2018). Genus-wide sequencing supports a two-locus model for sex-determination in *Phoenix*. *Nature Communications*, 9, 3969. https://doi.org/10.1038/s41467-018-06375-y
Tsugama, D., Matsuyama, K., Ide, M., Hayashi, M., Fujino, K., Masuda, K. (2017). A putative *MYB35* ortholog is a candidate for the sex-determining genes in *Asparagus officinalis*. *Scientific Reports*. 7, 41497. https://doi.org/10.1038/srep41497
Vicoso, B., Kaiser, V.B., Bachtrog, D. (2013). Sex-biased gene expression at homomorphic sex chromosomes in emus and its implication for sex chromosome evolution. *Proceedings of the National Academy of Sciences of the United States of America* 110 (16):6453-6458. https://doi.org/10.1073/pnas.1217027110
Villanueva, J.M., Broadhvest, J., Hauser, B.A., Meister, R.J., Schneitz, K., Gasser, C.S. (1999). *INNER NO OUTER* Regulates abaxial-adaxial patterning in *Arabidopsis* ovules. *Genes and Development*. 13:3160-3169. https://doi.org/10.1101/gad.13.23.3160
Villarreal, J.C., and Renner, S.S. (2013). Correlates of monoicy and dioicy in hornworts, the apparent sister group to vascular plants. *BMC Evolutionary Biology*. 13, 239. https://doi.org/10.1186/1471-2148-13-239
Walas, Ł., Mandryk, W., Thomas, P.A., Tyrała-Wierucka, Z., Iszkuło, G. (2018). Sexual systems in gymnosperms: A review. *Basic and Applied Ecology*. 31: 1–9. https://doi.org/10.1016/j.baae.2018.05.009
Wang, Z., Jiao, Z., Xu, P., Chen, L., Ai, J., Liu, X. and Yang, Y. (2013) Bisexual flower ontogeny after chemical induction and berry characteristics evaluation in male *Vitis amurensis* Rupr. *Scientia Horticulturae*. 162, 11– 19. https://doi.org/10.1016/j.scienta.2013.07.038
Wei, N., Govindarajulu, R., Tennessen, J.A., Liston, A., Ashman, T.L. (2017). Genetic mapping and phylogenetic analysis reveal intraspecific variation in sex chromosomes of the virginian strawberry. *Journal of Heredity*. 1-9. https://doi.org/10.1093/jhered/esx077.
West, N.W., and Golenberg, E.M. (2018). Gender-specific expression of *GIBBERELLIC ACID INSENSITIVE* is critical for unisexual organ initiation in dioecious *Spinacia oleracea*. *New Phytologist*. 217 (3):1322-1334. https://doi.org/10.1111/nph.14919
Westergaard, M. (1958). The mechanism of sex determination in dioecious flowering plants. *Advances in Genetics*. 9:217-281. https://doi.org/10.1016/S0065-2660(08)60163-7
Yamaki, S., Nagato, Y., Kurata, N., Nonomura, K.I. (2011). Ovule is a lateral organ finally differentiated from the terminating floral meristem in rice. *Developmental Biology*. 351 (1): 208–216. https://doi.org/10.1016/j.ydbio.2010.12.006
Yang, X., Wu, D., Shi, J., He, Y., Pinot, F., Grausem, B., Yin, C. et al. (2014). Rice *CYP703A3*, a cytochrome P450 hydroxylase, is essential for development of anther cuticle and pollen exine. *Journal of Integrative Plant Biology*. 56 (10):979-994. https://doi.org/10.1111/jipb.12212
Yang, H.W., Akagi, T., Kawakatsu, T., Tao, R. (2019). Gene networks orchestrated by *MeGI*: a single-factor mechanism underlying sex determination in persimmon. *The Plant Journal*. 98(1):97-111. https://doi.org/10.1111/tpj.14202
Yonemori, K., Sugiura, A., Tanaka, K. and Kameda, K. (1993) Floral ontogeny and sex determination in monoecious-type persimmons. *Journal of the American Sociecty for Horticultural Science*. 118, 293–297. https://doi.org/10.21273/JASHS.118.2.293
Zhang, H., Zhang, R., Yang, X., Gu, K.J., Chen, W., Chang, Y., Xu, Q. et al. (2019). Recent origin of an XX/XY Sex-determination system in the ancient plant lineage *Ginkgo biloba*. *BioRxiv* [Preprint]. Available at: https://doi.org/10.1101/517946
Zhou, Q., Zhang, J., Bachtrog, D., An, N., Huang, Q., Jarvis, E.D., Gilbert, M.T.P., Zhang, G. (2014). Complex evolutionary trajectories of sex chromosomes across bird taxa. *Science*. 346 (6215):1246338. https://doi.org/10.1126/science.1246338
Zhou, R., Macaya-Sanz, D., Rodgers-Melnick, E., Carlson, C.H., Gouker, F.E., Evans, L.M., Schmutz, J. et al. (2018). Characterization of a large sex determination region in *Salix purpurea* L. (Salicaceae). *Molecular Genetics and Genomics.* 293:1437-1452. https://doi.org/10.1007/s00438-018-1473-y
Zhou, R., Macaya-Sanz, D., Carlson, C.H., Schmutz, J., Jenkins, J.W., Kudrna, D., Sharma, A. et al. (2020a). A willow sex chromosome reveals convergent evolution of complex palindromic repeats. *Genome Biology*. 21, 38. https://doi.org/10.1186/s13059-020-1952-4

Zhou, R.; Macaya-Sanz, D.; Schmutz, J.; Jenkins, J.W.; Tuskan, G.A.; DiFazio, S.P. (2020b). Sequencing and Analysis of the Sex Determination Region of *Populus trichocarpa*. *Genes*, 11(8), 843. https://doi.org/10.3390/genes11080843

Zhou, Y., Minio, A., Massonnet, M., Solares, E., Lv, Y., Beridze, T., Cantu, D., Gaut, B.S. (2019). The population genetics of structural variants in grapevine domestication. *Nature Plants*, 5:965-979. https://doi.org/10.1038/s41477-019-0507-8

Zluvova, J., Nicolas, M., Berger, A., Negrutiu, I., 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 of the United States of America*. 103 (49): 18854–18859. https://doi.org/10.1073/pnas.0606622103

Chapter 2

A single gene underlies the dynamic evolution of poplar sex chromosome

Niels A. Müller[1,8*], Birgit Kersten[1,8], Ana P. Leite Montalvão[1], Niklas Mähler[2], Carolina Bernhardsson[3], Katharina Bräutigam[4], Zulema Carracedo Lorenzo[2], Hans Hoenicka[1], Vikash Kumar[2], Malte Mader[1], Birte Pakull[1], Kathryn Robinson[2], Maurizio Sabatti[5], Cristina Vettori[6], Pär K. Ingvarsson[3], Quentin Cronk[7], Nathaniel R. Street[2] & Matthias Fladung[1*]

[1] Thünen Institute of Forest Genetics, Grosshansdorf, Germany.
[2] Department of Plant Physiology, Umeå Plant Science Centre, Umeå, Sweden.
[3] Department of Plant Biology, Linnean Centre for Plant Biology, Swedish University of Agricultural Sciences, Uppsala, Sweden.
[4] Department of Biology, University of Toronto Mississauga, Mississauga ON, Canada.
[5] Department for Innovation in Biological, Agro-food and Forestry systems, University of Tuscia, Viterbo, Italy.
[6] Institute of Biosciences and BioResources, Division of Florence, National Research Council, Sesto Fiorentino, Italy.
[7] Department of Botany, University of British Columbia, Vancouver BC, Canada.
[8] These authors contributed equally
[*] Corresponding author

Published in *Nature Plants* (2020)
doi: 10.1038/s41477-020-0672-9

Abstract[1]

Although hundreds of plant lineages have independently evolved dioecy, i.e. separation of the sexes, the underlying genetic basis remains largely elusive[1]. Here we show that diverse poplar species carry partial duplicates of the *ARABIDOPSIS RESPONSE REGULATOR 17* (*ARR17*) ortholog within the male-specific region of the Y chromosome. These duplicates give rise to small RNAs apparently causing male-specific DNA methylation and silencing of the *ARR17* gene. Excitingly, CRISPR/Cas9-induced mutations demonstrate that *ARR17* functions as a sex switch triggering female development when 'on' and male development when 'off'. Despite repeated turnover events, including a transition from the XY to a ZW system, sex-specific regulation of *ARR17* is conserved across the poplar genus and likely beyond. Our data reveal how a single-gene-based mechanism of dioecy can enable highly dynamic sex-linked regions and contribute to maintaining sex chromosome recombination and integrity.

Main text

Sexual reproduction is pervasive throughout nature and enhances the rate of adaptation[2]. Whereas gonochorism (unisexuality) is most common in animals, most plant species are hermaphrodites[3]. Although the separation of female and male flowers on different individuals, called dioecy, has evolved independently in diverse plant lineages hundreds if not thousands of times[4], the genetic basis of the implied transitions remains largely elusive. The principal model assumes that two linked mutations on one chromosome, one affecting female and one male function, are necessary and initiate the formation of suppressed recombination and sex chromosome divergence[5]. Recent experimental data support this classic view[6,7]. A single-gene-based mechanism represents an intriguing alternative[8] but has been difficult to unambiguously test[9-11]. Here we employ poplar (*Populus*) natural variation to resolve the genetic basis and the evolutionary trajectory of an ancient but highly variable system of sex determination.

The genus *Populus* exhibits at least two distinct sex-determining regions (SDRs). While the SDR of the aspens *P. tremuloides* and *P. tremula* comprises the pericentromeric region of chromosome 19 (ref. 12) (Supplementary Fig. 1), the SDR of the balsam poplars *P. trichocarpa* and *P. balsamifera* spans only a small part of the distal end of the same chromosome[13]. Genome-wide association studies (GWAS) in balsam poplars[13,14] identified the highest number of sex-linked single nucleotide polymorphisms (SNPs) in a type-A cytokinin response regulator (Potri.019G133600), orthologous to the *Arabidopsis* gene pair *ARABIDOPSIS RESPONSE REGULATOR 16* (*ARR16*) / *ARR17*. We subsequently refer to this gene as *ARR17* for consistency with previous work. Using genome re-sequencing data[13], we identified a male-specific duplication of exons 1-4 (Fig. 1a), exactly coinciding with the sex-linked SNPs (Supplementary Fig. 2). Since the available *P. trichocarpa* reference genome is derived from a female individual[15], reads from the male-specific partial *ARR17* duplicates can only map to the *ARR17* gene. Strikingly, *P. tremula* DNA-seq data[16] also revealed male-specific partial *ARR17* duplicates (Fig. 1b). This opens the exciting possibility of a shared mechanism of sex determination in aspens and balsam poplars despite 15 million years of independent evolution.

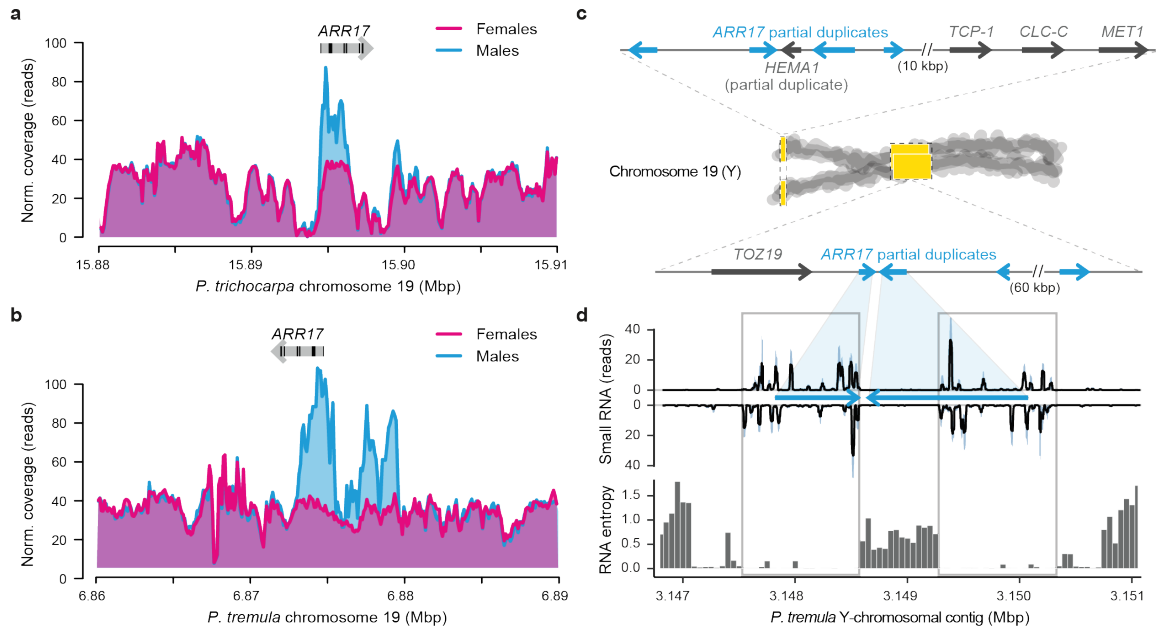


Figure 1. Male-specific partial *ARR17* duplicates arranged as inverted repeats are present in distantly related poplar species. (a, b) Mean normalized DNA-seq coverage of female (magenta line and shading) and male (cyan line and shading) (a) balsam poplars (*P. trichocarpa* and *P. balsamifera*) (n=52 biologically independent female samples and n=34 biologically independent male samples) and (b) European aspens (*P. tremula*) (n=12 biologically independent female samples and n=12 biologically independent male samples) in 100 bp windows along the *ARR17* gene region of the respective reference genomes[15,44]. *ARR17* is depicted by a gray arrow; exons are indicated by black vertical lines. (c) Schematic representation of

the Y-chromosomal sex-determining regions (indicated in yellow) and the comprised male-specific partial *ARR17* duplicates (blue arrows) derived from long-read assemblies of representative cottonwood (*P. deltoides*, upper panel) and aspen (*P. tremula*, lower panel) (Methods). In both cases, the male-specific partial *ARR17* duplicates are arranged as inverted repeats and located immediately adjacent to previously reported sex-linked genes, i.e. *HEMA1* (Potri.009G080600) in the cottonwood and *TOZ19* (Potri.019G047300) in the aspen. (d) Upper part: mean small RNA-seq coverage (solid black line) of n=7 biologically independent male *P. tremula* samples along the Y-chromosomal *ARR17* inverted repeat region of the new male *P. tremula* genome assembly (Methods). Shaded areas indicate the standard deviation. Alignments for the forward and reverse strands are shown above and below the line, respectively. Data are from two independent experiments. Lower part: estimated positional entropy for secondary RNA structures (Supplementary Fig. 4). Low entropy regions expected to form double-stranded RNA are marked by gray boxes.

To elucidate the genomic structure of the male-specific partial *ARR17* duplicates, we employed single-molecule long-read sequencing data, allowing us to assemble highly contiguous Y-chromosomal contigs. To this end we used publically available sequencing data of a representative cottonwood species (*P. deltoides*), which shares the genetic basis of sex determination with balsam poplars[13], but also generated Oxford Nanopore data for a male aspen (*P. tremula*) clone. These assemblies revealed that the male-specific partial *ARR17* duplicates are located immediately adjacent to established sex-linked genes[13,17] and are arranged as inverted repeats (Fig. 1c). The inversely repeated structure suggested the possibility that double-stranded RNA (dsRNA) may be formed if the inverted repeats are actively transcribed. This dsRNA could then be processed into small interfering RNAs (siRNAs) that may affect expression of the *ARR17* gene. In line with this, we identified male-specific non-coding transcripts of the *ARR17* inverted repeat in aspens (Extended Data Fig. 1 and Supplementary Fig. 3). RNA folding prediction highlighted two regions of about 1,000 bp, corresponding to the two arms of the inverted repeat, expected to form dsRNA as indicated by very low entropy levels (Fig. 1d and Supplementary Fig. 4). Remarkably, small RNAs, and in particular 24-nt RNAs, precisely mapped to those two regions (Fig. 1d, Extended Data Fig. 2 and Supplementary Fig. 5,6). These small RNAs provide a potential link to the male-specific DNA methylation patterns of the *ARR17* locus, which have been previously reported for *P. balsamifera*[18] (Extended Data Fig. 3a). Newly generated bisulfite sequencing data of *P. tremula* males and females revealed male-specific methylation of the *ARR17* locus in aspens as well (Extended Data Fig. 3b). Although the causal relationship remains to be experimentally tested in future studies, our data indicate that Y-chromosomal *ARR17*

inverted repeats apparently trigger RNA-directed DNA methylation (RdDM), which is known for mediating epigenetic gene silencing[19].

To explore whether silencing of *ARR17* may in fact influence sex determination, we first analyzed *ARR17* expression. Expression was minimal across various *P. tremula* tissues (Supplementary Note 1) but present in developing flower buds (Extended Data Fig. 4). Notably, this expression was limited to female clones, supporting the hypothesis that silencing of *ARR17* may be important for male flower development. To experimentally test the role of ARR17 in poplar sex expression, we generated CRISPR/Cas9-mediated *arr17* mutations in early-flowering male and female aspen lines. Excitingly, all female *arr17* CRISPR plants developed functional stamens and mostly failed to develop carpels (Fig. 2, Extended Data Fig. 5 and Supplementary Table 1), confirming the feminizing role of ARR17. Additionally, these data demonstrate that ARR17 functions as a master sex switch triggering female development when 'on' and male development when 'off'. In line with *ARR17* being naturally silenced in males, the flowers of the male *arr17* CRISPR lines did not show any difference compared to the controls (Supplementary Fig. 7). Although occasional hermaphrodite flowers occurred on masculinized *arr17* lines (Supplementary Table 1 and Supplementary Fig. 7), similar to previously reported early flowering aspen males[20], the complete sex switch of the majority of flowers rather than a reversion to hermaphrodites or neuters, as observed in *Silene*, asparagus or kiwi[6,7,21], strongly argues for a single-gene sex-determining system.

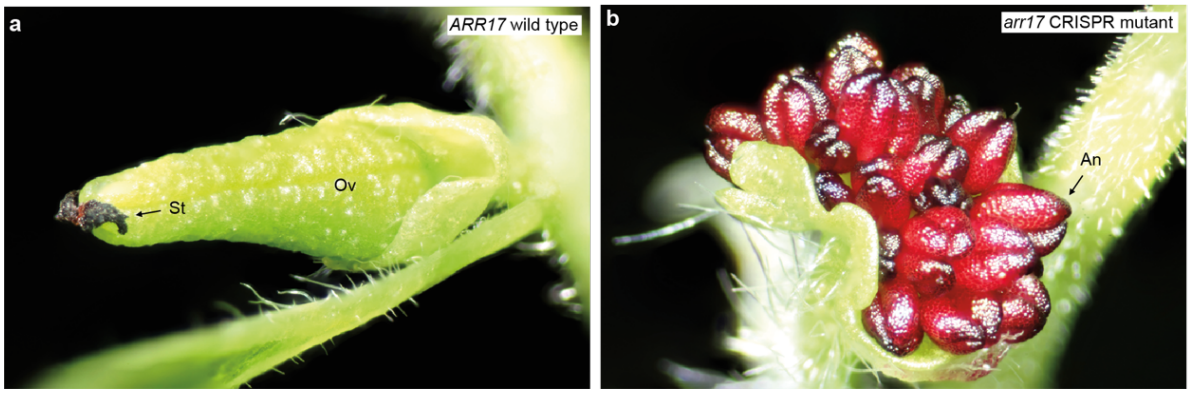


Figure 2. *ARR17* represents a sex switch triggering female development when 'on' and male development when 'off'. (a, b) Flowers of a female *P. tremula* ('Brauna11') early-flowering HSP::FT transgenic line with (a) the wild type *ARR17* gene or (b) a CRISPR/Cas9-induced *arr17* mutation. Photos show 40× magnifications of selected flowers of representative four-month-old plants. Floral organs are indicated: St = stigma, Ov = ovary, An = anther. The experiment included five independent *arr17* lines and was repeated twice with similar results.

White poplar (*P. alba*) represents an exception in the genus *Populus* as it exhibits female heterogamety, i.e. a ZW system of sex determination[22]. A sex chromosome turnover event must have occurred during the evolution of this species, changing the XY to a ZW system. We wondered whether a new gene acquired control of sex determination during this transition, or whether *ARR17* retained its sex-determining function. To address this question, we set out to precisely localize the SDR of white poplar and to define the W chromosome-specific sequence. We first sequenced a group of female and a group of male *P. alba* clones. Scanning these data for hemizygous regions, i.e. W chromosome-specific sequences only present in ZW females but not ZZ males, highlighted a single genomic region on chromosome 19 missing in white poplar males (Extended Data Fig. 6). Remarkably, *ARR17* is located exactly in the middle of this region. Sequencing and *de novo* assembly of a female *P. alba* clone yielded a W-chromosomal contig comprising three female-specific complete copies of *ARR17* (Fig. 3a). Comparison of that W-chromosomal contig with the corresponding Z-chromosomal sequence (i.e. part of chromosome 19) showed that precisely this *ARR17* region is deleted from the Z chromosome (Fig. 3b). Finally, SNPs differentiating the Z and W chromosomes indicated that the white poplar SDR may be just as small as the one of balsam poplars and may comprise less than 200 kbp (Fig. 3c). The male-specific deletion of *ARR17*, which we confirmed in 53 additional *P. alba* individuals (Supplementary Table 2), provides a simple explanation for the switch from male to female heterogamety. Instead of dominantly silencing *ARR17* in trans by Y-chromosomal inverted repeats, loss of the *ARR17* gene from the Z chromosome leads to a recessive inactivation in cis.

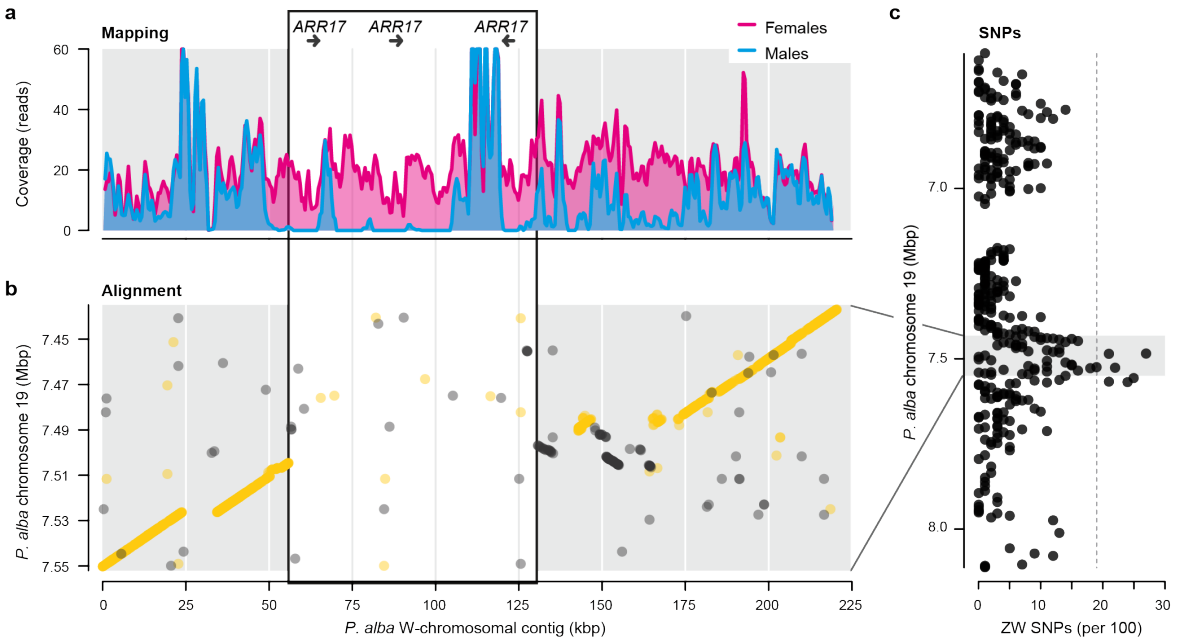


Figure 3. The *ARR17* gene is deleted in white poplar males causing a switch from an XY to a ZW system. (a) DNA-seq coverage of a pool of seven female (magenta line and shading) and seven male (cyan line and shading) white poplar (*P. alba*) individuals in sliding windows (window=1,000 bp, step=500 bp) along the W-chromosomal contig of a new female *P. alba* genome assembly (Methods). Copies of the complete *ARR17* gene are depicted as arrows. (b) Alignment of the W-chromosomal contig and the corresponding Z-chromosomal region of pseudo-chromosome 19 of *P. alba*. Forward and reverse alignments are shown as yellow and gray dots, respectively. The region entirely deleted from the Z chromosome is indicated by white background. (c) Number of ZW SNPs, exhibiting heterozygosity (i.e. allele frequencies between 40% and 60%) in the female pool but homozygosity in the male pool, in 100 SNP windows along chromosome 19. The Z-chromosomal region from (b) is marked by gray shading. The dashed line indicates the chromosome-wide top 1% cutoff.

Finally, having determined *ARR17* as the master regulator of dioecy in *Populus*, we wanted to clarify the evolutionary trajectory of the different sex-determining regions and the turnover events that occurred during the more than 15 million years of poplar evolution. We therefore analyzed the sequences of the *ARR17* gene and the *ARR17* inverted repeat in seven poplar species. For this we employed publically available genomic data but also our *de novo* assembled male *P. tremula* and female *P. alba* genomes and a *P. tremuloides* BAC clone sequence. A phylogenetic analysis revealed two independent origins of the *ARR17* inverted repeat, one of which appears to predate poplar speciation (Fig. 4a). Additionally, one transition from male to female heterogamety (XY to ZW) can be observed (Fig. 4a,b). The results from the phylogenetic analysis are further supported by the genomic structure of the inverted repeats. The distantly related poplar species *P. trichocarpa*, *P. deltoides*, *P. euphratica* and *P. pruinosa* all harbor the partial *HEMA1* sequence in the middle of the *ARR17* inverted repeat, as

depicted in Fig. 1c. This indicates a common evolutionary origin dating far back in the evolution of the genus. The *ARR17* inverted repeat in the aspens, on the other hand, has no similarity besides the inversely repeated arrangement of the partial *ARR17* sequences, making a shared origin unlikely even in the potential event of gene conversion. The sex chromosome turnover events in *Populus* thus appear to be characterized by newly formed SDRs taking over sex determination, unlike the 'jumping' SDR of strawberries[23]. Nevertheless, the underlying molecular mechanism is preserved. Although independent formation of *ARR17* inverted repeats seems unlikely at first, evidence for a special predisposition of this sequence to evolve is provided by the genome of the sister species *Salix purpurea*. An inverted repeat of the first half of *ARR17* is also present in the *Salix* genome (Fig. 4a). However, this sequence is not sex-linked and likely non-functional[24]. Mechanistically this predisposition may be explained by retrotransposons within and around the *ARR17* gene (Supplementary Note 2), which are known for their ability to mediate gene duplication[25]. Our results add to the growing evidence of the importance of sequence duplications in the evolution of sex determination[6,26,27].

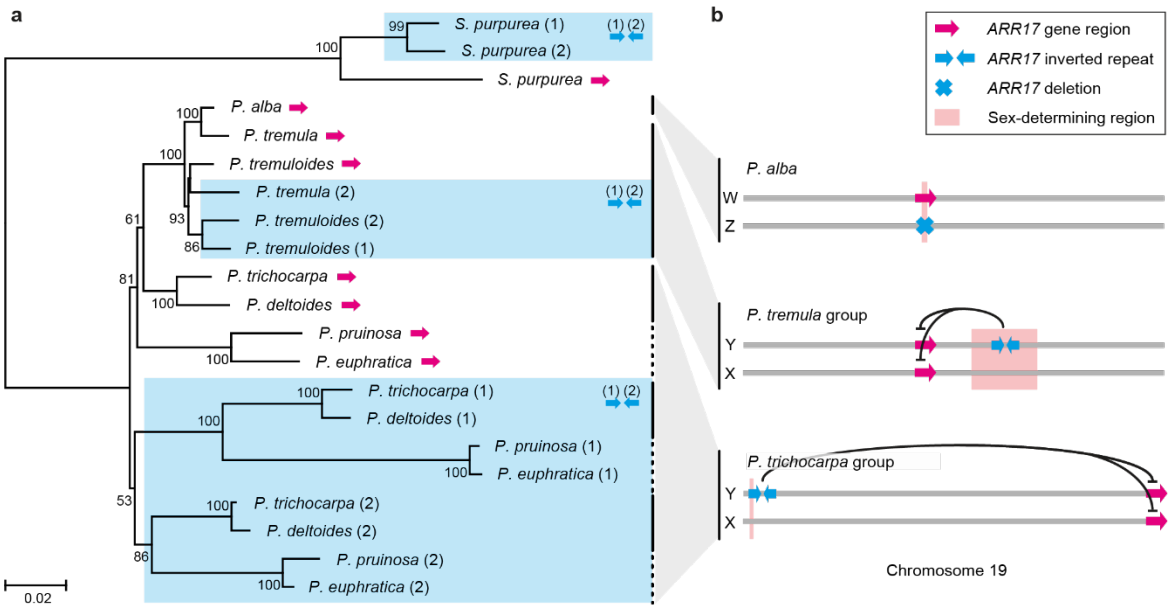


Figure 4. Independently evolved sex-determining regions regulate *ARR17* expression across the *Populus* genus. (a) Phylogeny of the *ARR17* gene and *ARR17* inverted repeat sequences from seven *Populus* and one *Salix* species analyzed by the Maximum Likelihood method. The two inverted repeat sequences are numbered (1) and (2) and marked by blue shading; *ARR17* gene sequences are marked by magenta arrows. The left arm (1) of the *ARR17* inverted repeat of *P. tremula* was excluded from the analysis because it is truncated. Numbers at the nodes indicate support values based on 1,000 bootstrap replications. Black vertical bars mark groups of species with different independently evolved sex-determining regions. (b) Schematic representation of the sex chromosomes of the three groups shown in (a). *ARR17* inverted repeats on the male Y chromosome dominantly repress *ARR17* expression in the *P. trichocarpa* and *P. tremula* groups. Deletion of the ARR17 gene from the male Z chromosome recessively inactivates *ARR17* and female development in *P. alba* causing the switch to a ZW system.

In summary, our data reveal *ARR17* as a master regulator of sex determination across the poplar genus (Fig. 4b). *ARR17* acts as a switch, reminiscent of the master regulators of sex in mammals and birds[28]. Whenever *ARR17* is 'on', female development is initiated. Only when *ARR17* is 'off', male development occurs. This switch can be either controlled directly, by a presence/absence mutation, or remotely via a putative sRNA/RdDM-based mechanism. Our results demonstrate that not only can a single segregating mutation be employed for the artificial generation of dioecy[29,30] but also for the evolution of a natural system. *ARR17* is a type-A response regulator implicated in negative regulation of the plant hormone cytokinin[31]. Plant hormones have well established roles in floral development and the regulation of monoecy[32]. A transition from monoecy to dioecy while maintaining the underlying genetic pathways for unisexual flower development would be an intriguing explanation for the single-gene based mechanism and the apparent lack of suppressed recombination.

Remarkably, genomic data of basket and purple willows (*Salix viminalis* and *S. purpurea*, respectively) also highlighted W chromosome-specific *ARR17* homologs[24,33], indicating that the molecular basis of sex determination may predate the split of the two genera *Populus* and *Salix*. Considering their evolutionary distance[34], the initial sex chromosomes would have had more than 60 million years to evolve suppressed recombination and genetic divergence as predicted by evolutionary theory[5,35]. Instead, Salicaceae sex chromosomes are highly dynamic and appear to stay ever young. The single-gene-based mechanism facilitating frequent sex chromosome turnover likely represents part of their 'fountain-of-youth'. Together with the lack of sexual conflict minimizing selection for sexual dimorphism[14,36], poplar sex chromosomes circumvent runaway expansion of the sex-determining region. Similar evolutionary patterns may more generally contribute to the surprising abundance of 'young' plant sex chromosomes[6,13,26,37-40]. Recent sex chromosome divergence and even different heterogametic systems do not exclude conservation of the underlying molecular mechanism. The highly dynamic sex-determining regions and the lack of suppressed recombination in *Populus* highlight a striking alternative for sex chromosome evolution compared to the established theoretical frameworks[35]. The genetic elucidation of dioecy in further species will allow exciting comparisons and will reveal how different systems may impact species evolution.

Methods

Plant material, growth conditions and phenotyping

For phenotyping sexual floral organs in plants with and without CRISPR-induced arr17 mutations, we used the early-flowering male *P. tremula* clone 'W52' (ref. 41). Additionally we transformed the female *P. tremula* clone 'Brauna11' with the previously described HSP::FT transgene[20] kindly provided by O. Nilsson. To induce early flowering, we used a protocol described before[41]. In short, in vitro-grown plants were transferred to soil and cultivated under 16/8 hours light/dark and 22/17 °C temperature cycles for 1.5 months. FT expression was then activated by heat-shocking the plants for 2 hours at 40 °C every morning for four weeks. Finally, plants were grown under cold conditions (16/8 hours light/dark and 12/6 °C temperature cycles) for up to two months to allow development of fertile flowers. Especially anther development largely fails under heat shock conditions. All plants were randomized and shuffled within the growth chambers. They were watered daily and repotted and fertilized when necessary. Flowers were phenotyped using a Zeiss binocular microscope. Pollen viability was assessed using Fluorescein diacetate (FDA) staining as described previously[41,42].

Mapping re-sequencing data of balsam poplars and aspens

Re-sequencing data from *P. trichocarpa*/*P. balsamifera* and *P. tremula* were downloaded from NCBI's SRA BioProjects PRJNA276056 and PRJNA297202, respectively (see Supplementary Table 3 for detailed information)[13,16]. Reads were trimmed using Trimmomatic v0.35 (ref.43) with the following parameters: ILLUMINACLIP: TruSeq3-PE-2.fa:2:30:10 LEADING:3 TRAILING:3 SLIDINGWINDOW:4:20 MINLEN:50. Trimmed paired reads were mapped to the according reference genome (Potri v3.0 and Potra v2.2, available on phytozome.jgi.doe.gov and popgenie.org, respectively)[15,44] using Bowtie2 v2.3.2 with default parameters[45].

Re-sequencing of female and male aspen and white poplar pools

Pool-sequencing of the two aspen species has been described before[17]. For the white poplar *P. alba* pool-seq we followed the same strategy. DNA of seven female and seven male genotypes was pooled in equal amounts. Illumina sequencing libraries were prepared and 150 bp paired-end sequencing was performed on the Illumina HiSeq2000 platform by GATC (GATC Biotech GmbH, Konstanz, Germany).

Genome-wide association study (GWAS) for sex in the Swedish Aspen collection

The sex of 75 genotypes of the Swedish Aspen (SwAsp) collection[46] was determined using the second *TOZ19* PCR marker previously reported[17] (Supplementary Table 4). Briefly, leaf material was collected from mature leaves of three clonal replicates per genotype. The test was confirmed as accurate in our aspen populations by examining flowers under a hand lens from 39 aspen trees in comparison to the same replicated PCR test in an outgroup of mature, flowering aspen trees sampled around Umeå, Sweden. Re-sequencing data[47] was mapped against the new male *P. tremula* genome assembly (see below) and SNP calling and filtering was performed as detailed before[47]. Genome-wide association mapping was conducted using GEMMA v 0.98.1 (ref. 48) following the methods described previously[47] and with sex used as a binary phenotype. Significant associations were determined using a 5% false discovery rate (qvalue) calculated using the function in the qvalue package in R[49].

Genome sequencing and de novo assembly

1) *Male P. tremula 'W52' (with HSP::FT transgene)*

DNA for Oxford Nanopore long-read sequencing was extracted using the CTAB method with a sorbitol pre-wash[50]. Importantly, lysis was done for one hour at 56 °C instead of 65 °C. Phase separation with chloroform:isoamyl alcohol was performed twice, precipitation was done by adding 0.66 volumes of ice-cold isopropanol. The extracted DNA was further processed by a subsequent cleanup step using an optimized SPRI

beads solution[51]. As a final step, we removed short DNA fragments by employing the Short Read Eliminator Kit (Circulomics Inc, Baltimore, MD, USA) according to the manufacturer's instructions. DNA for Illumina sequencing was extracted as previously described[52]. Illumina 350 bp insert sequencing libraries were prepared and sequenced on the Illumina HiSeq2000 platform by Novogene (Novogene Europe, Cambridge, UK).

All long-read sequencing was done with Oxford Nanopore kits and devices (Oxford Nanopore Technologies, Oxford, UK). The sequencing library was prepared using the Ligation Sequencing Kit (SQK-LSK109) following the manufacturer's instructions, with 1 µg of DNA as input and an extended adapter ligation step of one hour. We used R9.4.1 MinION Flow Cells primed with the Flow Cell Priming Kit (EXP-FLP002). Sequencing was performed on a MinION Mk1B device (MIN-101B) connected to a MinIT computer (MNT-001). Each flow cell was run for 48 hours. We used the MinIT software version 19.06.8 and selected accurate basecalling using the guppy 'flip-flop' mode, which produces higher read and consensus accuracy[53]. Reads generated with the older MinIT software version (19.01.1) not including the accurate basecalling option were re-basecalled on the MinIT GPUs using guppy v3.0.6.

Sequencing data (fastq-files) of two runs were assembled with Canu v1.8 (ref. 54) using default settings, except 'useGrid=false'. The input data consisted of 1.13 million reads totaling 21.06 Gbp. The initial assembly was polished by two rounds of Pilon[55] using 60× 150 bp paired-end Illumina data of the same individual, which was trimmed using Trimmomatic v0.35 (ref. 43) with the following parameters: ILLUMINACLIP: TruSeq3-PE-2.fa:2:30:10 LEADING:3 TRAILING:3 SLIDINGWINDOW:4:20 MINLEN:50. The resulting genome assembly was slightly inflated with a size of 605 Mbp. Estimates based on flow cytometry are around 480 Mbp[34]. This inflation appears to be due to haplotype splitting. A coverage analysis with the Purge Haplotigs pipeline[56] revealed a substantial amount of contigs representing single haplotypes. A BUSCO v3.0.2 analysis[57] with AUGUSTUS v 3.3.2 corroborated this indication but also demonstrated a high level of genome completeness: C:95.4% [S:51.0%,D:44.4%], F:0.8%, M:3.8%, n:1440. Also, the assembly contiguity was relatively high with a contig N50 of 1.26 Mbp.

We removed haplotypic contigs with the Purge Haplotigs pipeline[56]. To produce reference-guided pseudochromosomes we employed RaGOO[58] together with the chromosome-level female *P. tremula* 'Asp201' assembly[44].

2) *Female P. alba '14P11'*

DNA extraction and sequencing was performed as described above for the male *P. tremula* individual. This time we additionally performed a nuclease flush after 24 hours of sequencing following the protocol available from Oxford Nanopore Technologies. This generated 1.15 million reads totaling 18.66 Gbp, which were again assembled using Canu v1.8 (ref. 54) and polished by two rounds of Pilon[55]. This assembly was 548 Mbp with a contig N50 of 836 kbp. Since heterozygous SNP frequencies along the genome will be confounded by split haplotypes, we ran the Purge Haplotigs pipeline[56] to remove haplotypic contigs. To produce reference-guided pseudochromosomes we employed RaGOO[58] together with the chromosome-level female *P. tremula* 'Asp201' assembly[44]. The resulting assembly had a size of 365 Mbp with a contig N50 of 1.89 Mbp. Interestingly, the gene space completeness, as assessed by BUSCO, remained almost identical – initial assembly: C:95.4% [S:66.6%,D:28.8%], F:0.8%, M:3.8%, n:1440; purged assembly: C:94.6% [S:81.5%,D:13.1%], F:1.1%, M:4.3%, n:1440.

For visualizing the global alignment of two sequences, e.g. the Z- and W-contigs, we used MUMmer 3 (ref. 59) employing the mummer and mummerplot programs.

3) *Male P. deltoides 'ILL101'*

To explore the genomic structure of the male-specific region of the Y chromosome (MSY) of *P. deltoides*, we downloaded PacBio and Illumina sequencing data of clone ILL-101 from NCBI's SRA (BioProject: PRJNA443297). These sequence data were produced by the US Department of Energy Joint Genome Institute http://www.jgi.doe.gov/ in collaboration with the user community. We assembled and polished the genome as described above, using Canu and Pilon. Here we only describe the contig containing the MSY (Supplementary Data 1), since whole-genome analyses are reserved under the JGI data usage policy.

Bisulfite-sequencing and methylation analysis

Whole-genome bisulfite sequencing of *P. balsamifera* has been described before[18]. Briefly, 125 bp paired-end sequences were quality and adapter trimmed, aligned to the *P. trichocarpa* reference genome (v3.0), and 5-methyl cytosine methylation was calculated as percentage relative to all cytosines per position. Only reads with a mapping quality (MQ) above 30 were considered to ensure that only the actual *ARR17* locus and not the partial duplicates are represented. Reads originating from the partial duplicates exhibit mapping quality scores below 30.

For *P. tremula*, methylation levels were estimated using bisulfite sequencing data of six individuals from the Swedish Aspen (SwAsp) collection[46]. All individuals were bisulfite sequenced using two biological replicates per individual on an Illumina HiSeq X instrument using paired-end libraries with 300 bp insert sizes. The sequencing was performed at the National Genomics Infrastructure facility at Science for Life Laboratory in Uppsala, Sweden. Samples were sequenced to an average depth of 60× as approximately 50% of the bisulfite sequencing data cannot be mapped uniquely due to excessive damage induced by the bisulfite treatment. Raw sequencing reads were trimmed using trimGalore v. 0.4.4 (https://github.com/FelixKrueger/TrimGalore), a wrapper around Cutadapt[60] and FastQC[61], using a paired-end trimming mode and otherwise default settings. Polymorphic substituted versions of the *P. tremula* v2.2 assembly were created for all samples separately to ensure as accurate methylation calls as possible. Trimmed reads for all individuals were mapped against the corresponding converted reference genomes using Bowtie2 (ref. 45) and optical duplicates were subsequently removed from the BAM files. Finally, methylation levels were called and extracted using Bismark[62]. Coverage-files were filtered for low (< 5) and high (> 44) coverage observations to remove spurious results due to low coverage or collapsed duplicate genomic regions, respectively. Following filtering, coverage files for all samples were merged and the different contexts of methylation (GpG, CHG and CHH) were extracted from the Bismark context files.

sRNA- and mRNA-sequencing

Plant material for sRNA and mRNA analysis was obtained from tissue culture-grown plants. Three to five plants per jar were collected and snap frozen in liquid nitrogen after cleaning any media and stored in -80 °C until further use. The frozen material was pulverized using a pestle and mortar in liquid nitrogen and the powder was immediately used for RNA extraction. RNA was extracted using a modified version of Chang et al.[63] as detailed previously[64], and the Qiagen RNeasy Kit. The final elution of the modified Chang et al. protocol was performed using RLT buffer from the RNeasy kit and the kit instructions were then followed to perform additional RNA cleanup, including DNase treatment, to obtain total RNA to be used for RNA-Seq library construction. The first elution was kept as this contained the short RNA fraction. This first elution flow through was added with 0.65 volumes of absolute ethanol, mixed thoroughly and loaded onto Qiagen MinElute columns. The manufacturer's instructions were then followed to obtain clean short RNAs, which were subsequently used to construct sequencing libraries. RNA was quantified using a Qubit fluorometer (Thermo Fisher Scientific), purity was checked using a Nanodrop and RNA integrity was determined using an Agilent Bioanalyzer using the plant specific protocol for calculating an RNA Integrity Number (RIN). All samples had RIN > 8.

Illumina mRNA-Seq and sRNA-Seq library construction was performed by Novogene (Novogene Europe, Cambridge, UK). Strand-specific mRNA-Seq libraries were produced using standardized procedures that included mRNA isolation, subsequent degradation, first strand cDNA synthesis using random hexamer primers, second strand cDNA synthesis and subsequent terminal repair, sequencing adapter ligation and size selection. The prepared library was sequenced as 2x150 bp reads with an insert size of ~300 bp on a Novaseq 6000 to a target depth of 10 million paired-end reads per sample. sRNA-Seq libraries were produced using the Illumina Small RNA Sample Pre kit and sequenced as 50 bp reads with an insert size of ~40 bp on a Novaseq 6000 to a target depth of 10 million reads per sample. For both datasets sequencing reads were filtered before alignment using the Novogene standard pipeline that, briefly, removes reads containing adapters (mRNA-Seq) or trims adapters (sRNA-Seq), removes reads containing > 10% N and reads where > 50% of bases have Qscore < 5. For sRNA-Seq,

additionally reads with 5' primer contaminants, reads without 3' primer or the insert tag and reads with polyA/T/G/C are removed.

mRNA-sequencing data was aligned to the male P. tremula 'W52' genome assembly (see above) using STAR v2.6.1d (ref. 65). Gene level read counts were estimated using HTSeq v0.11.2 and coverage values were obtained from STAR by using the `--outWigType bedGraph` flag. sRNA-sequencing data was aligned to the male *P. tremula* 'W52' genome assembly (see above) using Bowtie v1.2.3 allowing for zero mismatches against the reference and a maximum of 100 multimapping locations for each read. RNAfold v2.4.14 from the Vienna RNA package[66] was used to analyze the folding dynamics of the *ARR17* inverted repeat. The CENSOR tool[67] on the giri Repbase website was used for repeat annotation of the "*TOZ19*/*ARR17* inverted repeat region". The Integrative Genomics Viewer (IGV)[68] was used for visualization of RNA-Seq mappings.

Counts for sequenced and mapped read are given in Supplementary Table 5 and 6 for mRNA and sRNA-sequencing, respectively.

Expression analysis of *ARR17* in aspen flower buds

Young aspen (*P. tremula*) flower and leaf buds were collected in the field at the Thünen Institute of Forest Genetics on September 12, 2019, from four female and four male genotypes. Flower buds could be easily distinguished from leaf buds by their round shape. Bracts, which severely compromised subsequent RNA extraction, were quickly removed before freezing the samples in liquid nitrogen. The dissected flower buds yielded tiny catkins (with a length of about 5-10 mm) with the individual flowers and even the floral organs already visible under a binocular microscope.

RNA was extracted with the Spectrum Plant Total RNA Kit (Sigma-Aldrich Chemie GmbH, Munich, Germany) according to the manufacturer's manual using 'Protocol B' including the On-Column DNase I Digestion Set step. RNA was quantified using a Nanodrop 1000 spectrophotometer (Peqlab Biotechnologie GmbH, Erlangen, Germany), and RNA quality was assessed by OD260/OD280 and OD260/OD230 ratios and by native agarose gels. cDNA was synthesized from 2 µg of RNA using Oligo(dT)^{20} primers and SuperScript

IV reverse transcriptase (Invitrogen) following the manufacturer's user guide, using 10 µl reactions without RNaseOUT.

qRT-PCR reactions were run in triplicates on a CFX96 Touch Real-Time PCR Detection System (Bio-Rad Laboratories GmbH, Feldkirchen, Germany) using the SsoAdvanced Universal SYBR Green Supermix and a two-step PCR program with an annealing/extension temperature of 60 °C. Melting curve analysis was carried out to confirm the specificity of the amplification. Calibration curves using a dilution series of the cDNA were made to confirm that PCR amplification efficiencies for each quantified gene were between 95 and 100%. Relative expression levels were calculated using the delta CT method[69]. Primers are given in Supplementary Table 7.

CRISPR/Cas-induced mutations

We employed the CRISPR/Cas9 system to generate *arr17* mutations. Guide RNAs (gRNAs) were designed using the CRISPR-P v2.0 web tool[70]. We selected two highly ranked gRNAs mapping to two different *ARR17* exons: gRNA1: GCAGTGACCACAGCAGAAAA; gRNA2: TACCATGAAGGAGATACCGG. These gRNAs were cloned by DCS (DNA Cloning Service, Hamburg, Germany) into a binary vector with a 35S-driven Cas9 nuclease under the control of the *Arabidopsis thaliana* U6 promoter (p6i0R-35sCasWT). This construct was transformed into the *Agrobacterium tumefaciens* strain GV3101-pMP90RK, which we then used for plant transformation. We transformed poplar explants following the established leaf disc method[71]. CRISPR mutations were determined by Sanger sequencing of the two target sites. Primers are given in Supplementary Table 7. In all but one line both alleles were mutated. The type of mutations ranged from single nucleotide insertions or deletions to more complicated rearrangements (Supplementary Table 8).

BAC clone sequencing

Isolation of BAC clone 1D3, containing the *ARR17* inverted repeat, was based on screening of a *P. tremuloides* BAC library[72]. The BAC library (55,296 clones, 8× haploid genome equivalents) was spotted on Hybond N+ membranes (GE Healthcare, Freiburg, Germany). Screening of the library was performed using a DIG-dUTP PCR partial-labelled

probe. Probe amplification was carried out using the Roche DIG DNA Labelling Kit (Merck, Darmstadt, Germany) according to manufacturer's instructions with genomic DNA of Turesson141 and the *TOZ19*-specific primers (Supplementary Table 7). Seven BAC clones were identified, four of which could be confirmed via *TOZ19* PCR. BAC DNA was isolated from overnight cultures using the Ψ Clone BAC DNA-Kit (Princeton Separations, Adelphia, USA).

The BAC DNA was sequenced on an Ion Torrent Personal Genome Machine (PGM) Sequencer. All used devices and consumables were from Life Technologies, USA. 100 ng of DNA were used to prepare an IonXpress_001 (CTAAGGTAAC)-barcoded sequencing library with the Ion Xpress Plus Fragment Library Kit following Ion Torrent PGM protocol. An equimolar library pool of 25 pM was amplified and enriched on the Ion OneTouch 2 System using the Ion PGM Template OT2 400 Kit. The sample was loaded onto one PGM Ion 314 Chip v2 and sequenced using the Ion PGM Hi-Q Sequencing Kit according to manufacturer's protocol.

Bioinformatic analyses were performed using CLC Genomics Workbench v.7.0.4 (QIAGEN Bioinformatics, Aarhus, Denmark). After removing duplicates with the "remove duplicates reads"-plugin, adapter sequences were trimmed using the "trim sequences"-tool with default parameters. Resulting reads were assembled using the "de novo assembly"-tool with default parameters. The resulting contig that included part of *TOZ19* and two inversely repeated fragments of *ARR17* was further extended based on terminal sequence overlaps with two other contigs. The resulting sequence was refined by three rounds of mappings of the trimmed reads using the "map reads to reference"-tool with default parameters but with a length fraction of 0.97 and a similarity fraction of 0.99 and subsequent consensus sequence calling with the "extract consensus sequence"-tool.

Variant calling

To call putatively heterozygous SNPs in the *P. tremula*, *P. tremuloides* and *P. alba* pool-sequencing data, which are expected to differentiate the two sex chromosomes in the heterogametic sex, we marked duplicates with the MarkDuplicates (Picard) tool and produced gVCF files using the GATK HaplotypeCaller. The gVCF files were jointly called

using the CombineGVCFs and GenotypeGVCFs tools[73]. The final VCF file was filtered using the R package vcfR74. Only bi-allelic variants with a minor allele frequency of 5% were kept. Additionally, all variants with a coverage below the 0.1 or above the 0.9 quantile of the respective pool were excluded. The remaining variants were screened for putative sex-specific heterozygous variants with allele frequencies between 40% and 60% in the heterogametic and 0% or 100% in the homogametic sex.

Phylogenetic analysis

For the phylogenetic analysis, we extracted the *ARR17* gene sequences from published genome assemblies: Potra v1.1 for *P. tremula*[34], PopEup v1.0 for *P euphratica*[75], draft assembly for *P. pruinosa*[76], Potri v3.0 for *P. trichocarpa*[15], WV94 v2.1 for *P. deltoides* and v1.0 (Clone 94006) Scaffold0922 for *Salix purpurea* (the *P. deltoides* and *S. purpurea* sequence data were produced by the US Department of Energy Joint Genome Institute http://www.jgi.doe.gov/ in collaboration with the user community). The *ARR17* gene sequence of *P. tremuloides* is represented by the consensus sequence of the female pool-seq data (see above) mapped against Potri v3.0. The *ARR17* gene sequences of *P. alba* is represented by the consensus of the three W-chromosomal copies of our new '14P11' genome assembly (see above). We extracted the *ARR17* inverted repeat sequences from the same published genome assemblies for *P. euphratica* and *P. pruinosa*. The *P. tremuloides* sequences were extracted from the BAC clone described above. The *P. tremula* sequence was taken from our new male 'W52' assembly and the *P. deltoides* from an ILL-101 assembly (see above). Finally, the *P. trichocarpa* inverted repeat was extracted by mapping all male re-sequencing data[13] (Supplementary Table 3) against the Potri v3.0 genome complemented with the contig representing the *P. deltoides* MSY (Supplementary Data 1). From these mappings we called SNPs as described above and filtered the resulting vcf file for variants with an alternative allele frequency above 50%. We then used the bcftools consensus command to apply those variants to the *P. deltoides* sequence.

All sequences were trimmed to the left half of *ARR17* of *P. trichocarpa* (v3.0) starting with the ATG and ending at position 1500. The resulting sequences can be found in Supplementary Data 2. The subsequent alignment with MUSCLE and phylogenetic

analysis inferred by the Maximum Likelihood method based on the Tamura-Nei model[77] were done using MEGA7 with 1,000 bootstrap replications and including indels ("use all sites" option) with otherwise default parameters[78].

Data analysis and visualization

All additional data analyses and visualization were done in R[79]. The following packages were used: Rsamtools, GenomicAlignements[80], seqinr[81] and ggplot2[82].

Data availability

DNA- and RNA-seq data have been deposited in NCBI's SRA under the accession number PRJNA542603. Genome assemblies have been deposited at ftp://plantgenie.org/Publications/Muller2019/.

Acknowledgements

We thank members of the Thünen Institute of Forest Genetics and S. DiFazio, M. Olson and G. Tuskan for helpful comments and discussions, J. Mank and T. Brügmann for critical reading of the manuscript, S. Müller for language editing, and K. Groppe, A. Eikhof, D. Ebbinghaus, G. Wiemann, D. Boedecker, M. Wellern, A. Worm, M. Spauszus, L. Lierke and J. Lüneburg for technical assistance. We acknowledge funding from a grant of the Deutsche Forschungsgemeinschaft to NAM (DFG grant number MU 4357/1-1) and MF (DFG grant number FL 263/15-1). NRS, NM, ZCL, VK and KMR acknowledge funding from the Trees for the Future (T4F), Knut and Alice Wallenberg Foundation, the Umeå Plant Science Centre Berzelii Centre, the Stiftelsen för Strategisk Forskning Centre for Plant Developmental Biology, the Kempe Foundation and the Swedish Research Council Vetenskapsrådet. We thank the Swedish National Genomics Infrastructure hosted at SciLifeLab, the National Bioinformatics Infrastructure Sweden (NBIS), for providing computational assistance and the Uppsala Multidisciplinary Center for Advanced Computational Science for providing computational infrastructure. We further acknowledge funding from the Natural Sciences and Engineering Research Council of

Canada (NSERC) to KB (grant no RGPIN-2017-06552) and QC (grant no RGPIN-2019-04041).

Author contributions

Conceptualization: NAM, BK, QC, NRS, MF. Data curation: NAM, BK, KB, NRS. Formal analysis: NAM, BK, APLM, NM, CB, KB, MM, KMR, PKI, NRS. Funding acquisition: NAM, BK, NRS, MF. Investigation: NAM, BK, APLM, NM, CB, KB, ZCL, HH, VK, BP, NRS. Methodology: NAM, BK, APLM, NM, CB, KB, ZCL, MM, BP, KMR, NRS, MF. Resources: MS, CV. Supervision: NAM, BK, PKI, QC, NRS, MF. Validation: NAM, NRS. Visualization: NAM, NM, KB, KMR, PKI. Writing – original draft: NAM. Writing – review & editing: All authors.

Competing interests statement

The authors declare that they have no competing interests.

Supplementary information

The extended data, supplementary notes 1 and 2 and supplementary figures, tables and data for this article can be found online at: https://doi.org/10.1038/s41477-020-0672-9

References

1. Henry, I.M., Akagi, T., Tao, R. & Comai, L. one hundred ways to invent the sexes: theoretical and observed paths to dioecy in plants. *Annu Rev Plant Biol* **69**, 553-575 (2018).
2. McDonald, M.J., Rice, D.P. & Desai, M.M. Sex speeds adaptation by altering the dynamics of molecular evolution. *Nature* **531**, 233-6 (2016).
3. Bachtrog, D. *et al.* Sex determination: why so many ways of doing it? *PLoS Biol* **12**, e1001899 (2014).
4. Renner, S.S. The relative and absolute frequencies of angiosperm sexual systems: dioecy, monoecy, gynodioecy, and an updated online database. *Am J Bot* **101**, 1588-96 (2014).
5. Charlesworth, B. & Charlesworth, D. A Model for the evolution of dioecy and gynodioecy. *American Naturalist* **112**, 975-997 (1978).
6. Akagi, T. *et al.* Two Y-chromosome-encoded genes determine sex in kiwifruit. *Nat Plants* **5**, 801-809 (2019).
7. Harkess, A. *et al.* sex determination by two Y-linked genes in garden asparagus. *Plant Cell* (2020).
8. Renner, S.S. Pathways for making unisexual flowers and unisexual plants: Moving beyond the "two mutations linked on one chromosome" model. *Am J Bot* **103**, 587-9 (2016).

9. Akagi, T., Henry, I.M., Tao, R. & Comai, L. A Y-chromosome-encoded small RNA acts as a sex determinant in persimmons. *Science* **346**, 646-50 (2014).
10. West, N.W. & Golenberg, E.M. Gender-specific expression of *GIBBERELLIC ACID INSENSITIVE* is critical for unisexual organ initiation in dioecious *Spinacia oleracea*. *New Phytol* **217**, 1322-1334 (2018).
11. Yang, H.W., Akagi, T., Kawakatsu, T. & Tao, R. Gene networks orchestrated by MeGI: a single-factor mechanism underlying sex determination in persimmon. *Plant J* **98**, 97-111 (2019).
12. Kersten, B., Pakull, B., Groppe, K., Lueneburg, J. & Fladung, M. The sex-linked region in *Populus tremuloides* Turesson 141 corresponds to a pericentromeric region of about two million base pairs on *P. trichocarpa* chromosome 19. *Plant Biol (Stuttg)* **16**, 411-8 (2014).
13. Geraldes, A. *et al.* Recent Y chromosome divergence despite ancient origin of dioecy in poplars (*Populus*). *Mol Ecol* **24**, 3243-56 (2015).
14. McKown, A.D. *et al.* Sexual homomorphism in dioecious trees: extensive tests fail to detect sexual dimorphism in *Populus*. *Sci Rep* **7**, 1831 (2017).
15. Tuskan, G.A. *et al.* The genome of black cottonwood, *Populus trichocarpa* (Torr. & Gray). *Science* **313**, 1596-604 (2006).
16. Wang, J., Street, N.R., Scofield, D.G. & Ingvarsson, P.K. Variation in linked selection and recombination drive genomic divergence during allopatric speciation of European and American aspens. *Mol Biol Evol* **33**, 1754-67 (2016).
17. Pakull, B., Kersten, B., Lüneburg, J. & Fladung, M. A simple PCR-based marker to determine sex in aspen. *Plant Biol (Stuttg)* **17**, 256-61 (2015).
18. Bräutigam, K. *et al.* Sexual epigenetics: gender-specific methylation of a gene in the sex determining region of *Populus balsamifera*. *Sci Rep* **7**, 45388 (2017).
19. Aufsatz, W., Mette, M.F., van der Winden, J., Matzke, A.J. & Matzke, M. RNA-directed DNA methylation in *Arabidopsis*. *Proc Natl Acad Sci U S A* **99 Suppl 4**, 16499-506 (2002).
20. Zhang, H. *et al.* Precocious flowering in trees: the *FLOWERING LOCUS T* gene as a research and breeding tool in *Populus*. *J Exp Bot* **61**, 2549-60 (2010).
21. Kazama, Y. *et al.* A new physical mapping approach refines the sex-determining gene positions on the *Silene latifolia* Y-chromosome. *Sci Rep* **6**, 18917 (2016).
22. Paolucci, I. *et al.* Genetic linkage maps of *Populus alba* L. and comparative mapping analysis of sex determination across *Populus* species. *Tree Genetics & Genomes* **6**, 863-875 (2010).
23. Tennessen, J.A. *et al.* Repeated translocation of a gene cassette drives sex-chromosome turnover in strawberries. *PLoS Biol* **16**, e2006062 (2018).
24. Zhou, R. *et al.* A willow sex chromosome reveals convergent evolution of complex palindromic repeats. *Genome Biology* **21**, 38 (2020).
25. Xiao, H., Jiang, N., Schaffner, E., Stockinger, E.J. & van der Knaap, E. A retrotransposon-mediated gene duplication underlies morphological variation of tomato fruit. *Science* **319**, 1527-30 (2008).
26. Harkess, A. *et al.* The asparagus genome sheds light on the origin and evolution of a young Y chromosome. *Nat Commun* **8**, 1279 (2017).
27. Akagi, T. *et al.* The persimmon genome reveals clues to the evolution of a lineage-specific sex determination system in plants. *PLoS Genet* **16**, e1008566 (2020).
28. Capel, B. Vertebrate sex determination: evolutionary plasticity of a fundamental switch. *Nat Rev Genet* **18**, 675-689 (2017).
29. Boualem, A. *et al.* A cucurbit androecy gene reveals how unisexual flowers develop and dioecy emerges. *Science* **350**, 688-91 (2015).
30. Jones, D.F. Unisexual maize plants and their bearing on sex differentiation in other plants and in animals. *Genetics* **19**, 552-67 (1934).
31. To, J.P. *et al.* Type-A Arabidopsis response regulators are partially redundant negative regulators of cytokinin signaling. *Plant Cell* **16**, 658-71 (2004).
32. Golenberg, E.M. & West, N.W. Hormonal interactions and gene regulation can link monoecy and environmental plasticity to the evolution of dioecy in plants. *Am J Bot* **100**, 1022-37 (2013).
33. Almeida, P. *et al.* Single-molecule genome assembly of the basket willow, *Salix viminalis*, reveals earliest stages of sex chromosome expansion. *bioRxiv*, 589804 (2019).
34. Lin, Y.C. *et al.* Functional and evolutionary genomic inferences in *Populus* through genome and population sequencing of American and European aspen. *Proc Natl Acad Sci U S A* **115**, E10970-E10978 (2018).
35. Charlesworth, D. Plant sex chromosomes. *Annu Rev Plant Biol* **67**, 397-420 (2016).

36. Robinson, K.M. *et al. Populus tremula* (European aspen) shows no evidence of sexual dimorphism. *BMC Plant Biol* **14**, 276 (2014).
37. Pucholt, P., Wright, A.E., Liu Conze, L., Mank, J.E. & Berlin, S. Recent sex chromosome divergence despite ancient dioecy in the willow *Salix viminalis*. *Mol Biol Evol* (2017).
38. Okazaki, Y., Takahata, S., Hirakawa, H., Suzuki, Y. & Onodera, Y. Molecular evidence for recent divergence of X- and Y-linked gene pairs in *Spinacia oleracea* L. *PLoS One* **14**, e0214949 (2019).
39. Veltsos, P. *et al.* Early sex-chromosome evolution in the diploid dioecious plant *Mercurialis annua*. *Genetics* **212**, 815-835 (2019).
40. Sousa, A., Fuchs, J. & Renner, S.S. Cytogenetic comparison of heteromorphic and homomorphic sex chromosomes in *Coccinia* (Cucurbitaceae) points to sex chromosome turnover. *Chromosome Research* **25**, 191-200 (2017).
41. Hoenicka, H., Lehnhardt, D., Briones, V., Nilsson, O. & Fladung, M. Low temperatures are required to induce the development of fertile flowers in transgenic male and female early flowering poplar (*Populus tremula* L.). *Tree Physiol* **36**, 667-77 (2016).
42. Widholm, J.M. The use of fluorescein diacetate and phenosafranine for determining viability of cultured plant cells. *Stain Technology* **47**, 189-194 (1972).
43. Bolger, A.M., Lohse, M. & Usadel, B. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* **30**, 2114-20 (2014).
44. Schiffthaler, B. *et al.* An improved genome assembly of the European aspen *Populus tremula*. Preprint at https://www.biorxiv.org/content/10.1101/805614v1 (2019).
45. Langmead, B. & Salzberg, S.L. Fast gapped-read alignment with Bowtie 2. *Nat Methods* **9**, 357-9 (2012).
46. Luquez, V. *et al.* Natural phenological variation in aspen (*Populus tremula*): the SwAsp collection. *Tree Genetics & Genomes* **4**, 279-292 (2008).
47. Wang, J. *et al.* A major locus controls local adaptation and adaptive life history variation in a perennial plant. *Genome Biol* **19**, 72 (2018).
48. Zhou, X. & Stephens, M. Genome-wide efficient mixed-model analysis for association studies. *Nat Genet* **44**, 821-4 (2012).
49. Storey, J.D., Bass, A.J., Dabney, A. & Robinson, D. qvalue: Q-value estimation for false discovery rate control. *R package version 2.18.0*, http://github.com/jdstorey/qvalue (2019).
50. Inglis, P.W., Pappas, M.C.R., Resende, L.V. & Grattapaglia, D. Fast and inexpensive protocols for consistent extraction of high-quality DNA and RNA from challenging plant and fungal samples for high-throughput SNP genotyping and sequencing applications. *PLoS One* **13**, e0206085 (2018).
51. Schalamun, M. *et al.* Harnessing the MinION: An example of how to establish long-read sequencing in a laboratory using challenging plant tissue from *Eucalyptus pauciflora*. *Mol Ecol Resour* **19**, 77-89 (2019).
52. Dumolin, S., Demesure, B. & Petit, R.J. Inheritance of chloroplast and mitochondrial genomes in pedunculate oak investigated with an efficient PCR method. *Theor Appl Genet* **91**, 1253-6 (1995).
53. Wick, R.R., Judd, L.M. & Holt, K.E. Performance of neural network basecalling tools for Oxford Nanopore sequencing. *Genome Biol* **20**, 129 (2019).
54. Koren, S. *et al.* Canu: scalable and accurate long-read assembly via adaptive k-mer weighting and repeat separation. *Genome Res* **27**, 722-736 (2017).
55. Walker, B.J. *et al.* Pilon: an integrated tool for comprehensive microbial variant detection and genome assembly improvement. *PLoS One* **9**, e112963 (2014).
56. Roach, M.J., Schmidt, S.A. & Borneman, A.R. Purge Haplotigs: allelic contig reassignment for third-gen diploid genome assemblies. *BMC Bioinformatics* **19**, 460 (2018).
57. Simao, F.A., Waterhouse, R.M., Ioannidis, P., Kriventseva, E.V. & Zdobnov, E.M. BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs. *Bioinformatics* **31**, 3210-2 (2015).
58. Alonge, M. *et al.* RaGOO: fast and accurate reference-guided scaffolding of draft genomes. *Genome Biol* **20,** 224 (2019).
59. Kurtz, S. *et al.* Versatile and open software for comparing large genomes. *Genome Biol* **5**, R12 (2004).
60. Martin, M. Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet.journal* **17**, 3 (2011).
61. Andrews, S. FastQC: a quality control tool for high throughput sequence data. Available online at: http://www.bioinformatics.babraham.ac.uk/projects/fastqc (2010).

62. Krueger, F. & Andrews, S.R. Bismark: a flexible aligner and methylation caller for Bisulfite-Seq applications. *Bioinformatics* **27**, 1571-2 (2011).
63. Chang, S., Puryear, J. & Cairney, J. A simple and efficient method for isolating RNA from pine trees. *Plant Molecular Biology Reporter* **11**, 113-116 (1993).
64. Street, N.R. *et al.* The genetics and genomics of the drought response in *Populus*. *Plant J* **48**, 321-41 (2006).
65. Dobin, A. *et al.* STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* **29**, 15-21 (2013).
66. Lorenz, R. *et al.* ViennaRNA Package 2.0. *Algorithms Mol Biol* **6**, 26 (2011).
67. Kohany, O., Gentles, A.J., Hankus, L. & Jurka, J. Annotation, submission and screening of repetitive elements in Repbase: RepbaseSubmitter and Censor. *BMC Bioinformatics* **7**, 474 (2006).
68. Robinson, J.T. *et al.* Integrative genomics viewer. *Nat Biotechnol* **29**, 24-6 (2011).
69. Schmittgen, T.D. & Livak, K.J. Analyzing real-time PCR data by the comparative C(T) method. *Nat Protoc* **3**, 1101-8 (2008).
70. Liu, H. *et al.* CRISPR-P 2.0: an improved CRISPR-Cas9 tool for genome editing in plants. *Mol Plant* **10**, 530-532 (2017).
71. Bruegmann, T., Polak, O., Deecke, K., Nietsch, J. & Fladung, M. Poplar transformation. *Methods Mol Biol* **1864**, 165-177 (2019).
72. Fladung, M., Kaufmann, H., Markussen, T. & Hoenicka, H. Construction of a Populus tremuloides Michx. BAC library. *Silvae Genetica* **57**, 65-69 (2008).
73. Poplin, R. *et al.* Scaling accurate genetic variant discovery to tens of thousands of samples. Preprint at https://www.biorxiv.org/content/10.1101/201178v3 (2018).
74. Knaus, B.J. & Grunwald, N.J. vcfr: a package to manipulate and visualize variant call format data in R. *Mol Ecol Resour* **17**, 44-53 (2017).
75. Ma, T. *et al.* Genomic insights into salt adaptation in a desert poplar. *Nat Commun* **4**, 2797 (2013).
76. Yang, W. *et al.* The draft genome sequence of a desert tree *Populus pruinosa*. *Gigascience* **6**, 1-7 (2017).
77. Tamura, K. & Nei, M. Estimation of the number of nucleotide substitutions in the control region of mitochondrial DNA in humans and chimpanzees. *Mol Biol Evol* **10**, 512-26 (1993).
78. Kumar, S., Stecher, G. & Tamura, K. MEGA7: molecular evolutionary genetics analysis version 7.0 for bigger datasets. *Mol Biol Evol* **33**, 1870-4 (2016).
79. R Core Team. R: A language and environment for statistical computing. (2019).
80. Lawrence, M. *et al.* Software for computing and annotating genomic ranges. *PLoS Comput Biol* **9**, e1003118 (2013).
81. Charif, D. & Lobry, J.R. SeqinR 1.0-2: A Contributed Package to the R Project for Statistical Computing Devoted to Biological Sequences Retrieval and Analysis. in *Structural Approaches to Sequence Evolution: Molecules, Networks, Populations* (eds. Bastolla, U., Porto, M., Roman, H.E. & Vendruscolo, M.) 207-232 (Springer Berlin Heidelberg, Berlin, Heidelberg, 2007).
82. Wickham, H. *ggplot2: Elegant Graphics for Data Analysis*, 216 (Springer Publishing Company, Incorporated, 2009).

Chapter 3

Recombination and segregation analysis of the white poplar SDRs

Ana P. Leite Montalvão, Annika Eikhof & Niels A. Müller

Thünen Institute of Forest Genetics, Großhansdorf, Germany

In preparation

Abstract

Suppressed recombination if often defined as a hallmark of sex chromosome evolution. However, its role in sex chromosomes with low levels of differentiation and small sex-determining regions (SDR) remains largely unexplored. In this study, we assessed the potential suppression of recombination along the white poplar SDR. To this end, cleaved amplified polymorphic sequences (CAPS) markers were placed flanking the edges of the SDR and analyzed in an F1 mapping population. This analysis highlights a possible allelic incompatibility between Z_1 and W. Hemizygosity of the W-specific poplar sex-determining gene *ARR17* may cause the recombination suppression between Z and W in the single-gene sex-determining system.

Introduction

More than a century has passed since genetic sex determination (i.e. sex chromosomes) was discovered (Allen, 1917; Stevens, 1905; Wilson, 1905). Yet, interesting questions about sex chromosome evolution remain open, especially how they evolve over time and what drives this process.

Sex chromosomes are chromosomes that carry sex-determining regions (SDRs) of varied differentiation (Hobza et al., 2018). Theory indicates that selection against recombination results in a nonrecombining region necessary to maintain the coinheritance of two sex-linked genes, which in turn are responsible for sequential mutations affecting female and male fertility, as per the two-gene model for the evolution of dioecy (Charlesworth and Charlesworth, 1978). The linkage of the two sex-determining genes is the starting point for recombination suppression in the SDR, which blocks the exchange of genetic material between the heterogametic chromosomes (Stapley et al., 2017). Recombination suppression is expected to spread over time and lead to chromosome differentiation (i.e. sequence divergence and accumulation of repetitive sequences) and degeneration (i.e. deletion of genes). However, empirical data show that SDRs and sex chromosomes can be highly variable (Furman et al., 2020). In case of single-gene sex determination, the SDR might differ by only a single nucleotide polymorphism (SNP) (Kamiya et al., 2012). On the other hand, SDRs can also present larger differences, for example, due to chromosomal rearrangements such as

translocations, inversions, duplications or transposable elements (TEs) that accumulate near the sex-determining locus (Charlesworth et al., 2005; Furman et al., 2020; Stapley et al., 2017). There are different scenarios in which suppressed recombination may develop and the rates of recombination may vary between species and between and within sex chromosomes (Rifkin et al., 2021; Sardell and Kirkpatrick, 2020; Stapley et al., 2017). Sex-determining regions recently formed are often small, but they can occur within larger non-recombining regions (Charlesworth, 2021). These regions can remain small over long evolutionary times (Renner and Müller, 2021), or evolve to large regions where recombination is suppressed, following the theoretical prediction (Charlesworth, 2021).

Experimental evidence is highlighting an increasing number of exceptions to the classic views. There are various dioecious plant species (e.g. asparagus, grapevine, poplar, strawberry or willow) that exhibit a near lack of genetic differentiation between the two sex chromosomes, which is often interpreted as evidence for the evolutionary young age of these sex chromosomes (Fechter et al., 2012; Geraldes et al., 2015; Pucholt et al., 2017; Tennessen et al., 2018). While this may be true in some cases, e.g. the garden asparagus (Harkess et al., 2017), a meta-analysis revealed that there is no correlation between the size and the age of plant SDRs (Renner and Müller, 2021). In fact, an alternative evolutionary trajectory of sex chromosome evolution where a single gene suffices to control dioecy (Akagi et al., 2014; Müller et al., 2020), suggests that selection against recombination may not always be necessary – "the two-gene model with male sterility followed by a female suppressor was proposed in order to understand why recombination often becomes suppressed in genome regions carrying sex-determining loci" (Charlesworth, 2019). Single-gene systems, however, rely on co-expression instead of coinheritance to define sexual development (Golenberg and West, 2013; Leite Montalvão et al., 2021; Renner and Müller, 2021; Yang et al., 2019).

The genus *Populus* is comprised of dioecious tree species with homomorphic sex chromosomes. Genome sequencing and linkage mapping revealed a strikingly diverse genetic architecture of sex determination in *Populus*, with independently evolved SDRs with different sizes and in different genomic locations (Geraldes et al., 2015; Kersten et al., 2014; Müller et al., 2020; Yang et al., 2021). For instance, although balsam poplars and aspens exhibit male heterogamety (XY system), the SDR of *P. trichocarpa* Torr. & A. Gray

and *P. balsamifera* L. is small (100 kb) and positioned at the beginning of chromosome 19, (Geraldes et al., 2015), while the aspens *P. tremula* L. and *P. tremuloides* Michx. present a larger SDR (~1 Mb) located in a pericentromeric region of chromosome 19 (Kersten et al., 2014; Müller et al., 2020). Conversely, a turnover event gave rise to a female heterogametic system (ZW system) in the white poplar *P. alba* L. which presents a small (~200 kb) centrally located SDR also on chromosome 19 (Müller et al., 2020; Paolucci et al., 2010; Yang et al., 2021). Importantly, sex determination in both the XY and ZW systems rely on *ARR17* expression (Müller et al., 2020). In poplars with XY systems, Y-specific partial *ARR17* duplicates apparently trigger RNA-directed DNA methylation (RdDM) mediating silencing of the *ARR17* gene (Müller et al., 2020). On the other hand, in ZW systems, the male-specific absence of *ARR17* from the Z chromosome mediates sex determination. Comparing these contrasting systems from closely related species can provide insights regarding the evolution of the SDRs and suppression of recombination in the genus *Populus*.

Here, we explore the potential recombination suppression along the SDR of the white poplar *P. alba*. Using re-sequencing data from the male and female parents and a reference genome assembly we developed markers flanking the SDR to determine the recombination frequencies and segregation patterns in an F1 mapping population. We also examine how these correlate with the evolution of the SDR in this species. Our results highlight that suppressed recombination occurs in the heterogametic parent (ZW) only in a small region, where the W-specific sex-determining gene *ARR17* is located.

Material and methods

Plant material

Genomic DNA (gDNA) from the F1 population was used to assess recombination frequencies and the potential of recombination suppression in white poplar SDR. The F1 population is a result from the cross between 14P11 (female parent) × 6K3 (male parent) and consisted of 60 individuals, which have their sex phenotypically identified as either male or female.

Re-sequencing of white poplar and SNP calling

The genomes of the female (14P11) and male (6K3) parents of white poplar were re-sequenced using next generation sequencing (NGS). Libraries were generated by Novogene (Novogene (UK) Company Ltd., Cambridge, UK) and whole-genome sequenced to a depth of 60x and 30x, respectively, using the Illumina HiSeq platform and paired end reads (PE 2 x 150 bp). The reads were filtered to remove ffffadapters, reads containing bases that cannot be determined (N>10%) and low-quality reads (Qscore ≤ 5). The quality of the raw reads was assessed using FASTQC (https://www.bioinformatics.babraham.ac.uk/projects/fastqc/). Following that, reads were trimmed using Trimmomatic (Bolger et al., 2014) with the following parameters: ILLUMINACLIP: TruSeq3-PE-2.fa:2:30:10 LEADING:3 TRAILING:3 SLIDINGWINDOW:4:15 MINLEN:36. The reads from 14P11 and 6K3 were mapped to a female *P. alba* 14P11 assembly (Müller et al., 2020). The SNPs were identified using GATK (McKenna et al., 2010) and filtered based on read depth (DP) to exclude low coverage variants. To reach a normal distribution, the DP was set to a minimum of 30 and a maximum of 100 for both clones using the R package vcfR (Knaus and Grünwald, 2016).

Marker development, PCR, and agarose gel genotyping

Cleaved amplified polymorphic sequences (CAPS) markers were designed to flank the SDRs of each parent using a custom R script to select heterozygous (i.e. 1/0 or 0/1) single nucleotide polymorphisms (SNPs) for one parent and homozygous (i.e. 1/1 and 0/0) for the other. The CAPS markers allow simple genotyping of the progeny using PCR, enzymatic digestion, and agarose gels. Essentially, in the case of the heterozygote SNP, two (sometimes three) different PCR products will be obtained, because they are cleaved by the restriction enzymes in specific sites. When fractionated by gel electrophoresis, the PCR products digested by the restriction enzyme gives readily distinguishable patterns. Comparing the markers placed on the same chromosome for a given sample means that whenever the two markers show differences a recombination event took place between the two markers.

The PCR experiments were performed in 10 µl reactions. Primer sequences and respective restriction enzymes are given in Table 1. For all PCR reactions, 100 ng of gDNA,

200 µM of dNTP, primers at 0.10 µM and 1 unit of Taq polymerase were used. The PCR started with a 3 min denaturation at 95 °C, followed by 40 cycles of 20 sec denaturation at 95 °C, 30 sec annealing at 53 °C or 55 °C (depending on the primer pair), 1 min extension at 72°C and a final extension of 10 min at 72 °C. Restriction enzyme digestion for each primer pair was performed at 37 °C for 1 hour. Following that, all PCR products were checked by gel electrophoresis at 70 volts for 2 hours using a 2% agarose gel with 0.5 x TAE as a running buffer. The DNA was stained with RotiSafe from Roth (Karlsruhe, Germany).

Table 1. PCR primer sequences designed and used to determine recombination frequencies in the white poplar SDR.

Marker	ID	Sequence (5'→ 3')	Tm (°C)	Amplicon size (bp)	Restriction enzyme	Expected fragment size (bp)
1	#3159	GAGGGGAAGGAAAAGGTACA	55.88	397	BamHI	275,122
	#3160	TTTGAAGCACAACATTAGCACA	57.21			
2	#3087	CCAGGGAGTGATAGGAGTAGAAAA	60.02	466	EcoRI	366,100
	#3088	TCAGGAAGATGGTGTGTGGA	60.09			
3	#3034	GTCTCATTTTTATGACCATGTTCACC	61.20	353	HindIII	184,169
	#3035	GCACAAAAACCACGAAAACTTG	61.18			
4	#2954	CAAGGAAGAAGACATGCTTGA	58.09	216	BamHI	154,62
	#2955	TTTTGGTTCGCATAGGTGAG	58.27			

Results and discussion

Genetic mapping and molecular markers are commonly used approaches employed to detect sex-linked regions (Charlesworth, 2021). A genetic map revealed the sex-linked region in white poplar to be positioned on linkage group XIX and suggested a ZW system of sex determination (Paolucci et al., 2010). Recently, a single sex-determining gene, the poplar response regulator *ARR17*, was recognized as the sex-determining gene in both XY and ZW systems (Müller et al., 2020; Yang et al., 2021). Nevertheless, insights regarding the evolution of the SDR in white poplar remained largely unexplored. Thus, in this study, we assessed the potential suppression of recombination along the white poplar SDR. We screened a subset of an F1 population utilizing cleaved amplified polymorphism sequences (CAPS) markers.

The F1 population was generated from a *P. alba* cross (14P11 x 6K3). Two markers per genomic position were placed at the edges of the SDR, for each chromosomal

combination, that is Z_1/Z_2 and WZ_3 (Figure 1a). This resulted in a set of four markers, hereafter referred to as markers 1 to 4 (Figure 1b).

The recombination frequencies along the white poplar SDR in male and female meiosis were estimated by analysing the segregation patterns resulting in four possible gametic combinations (Figure 2).

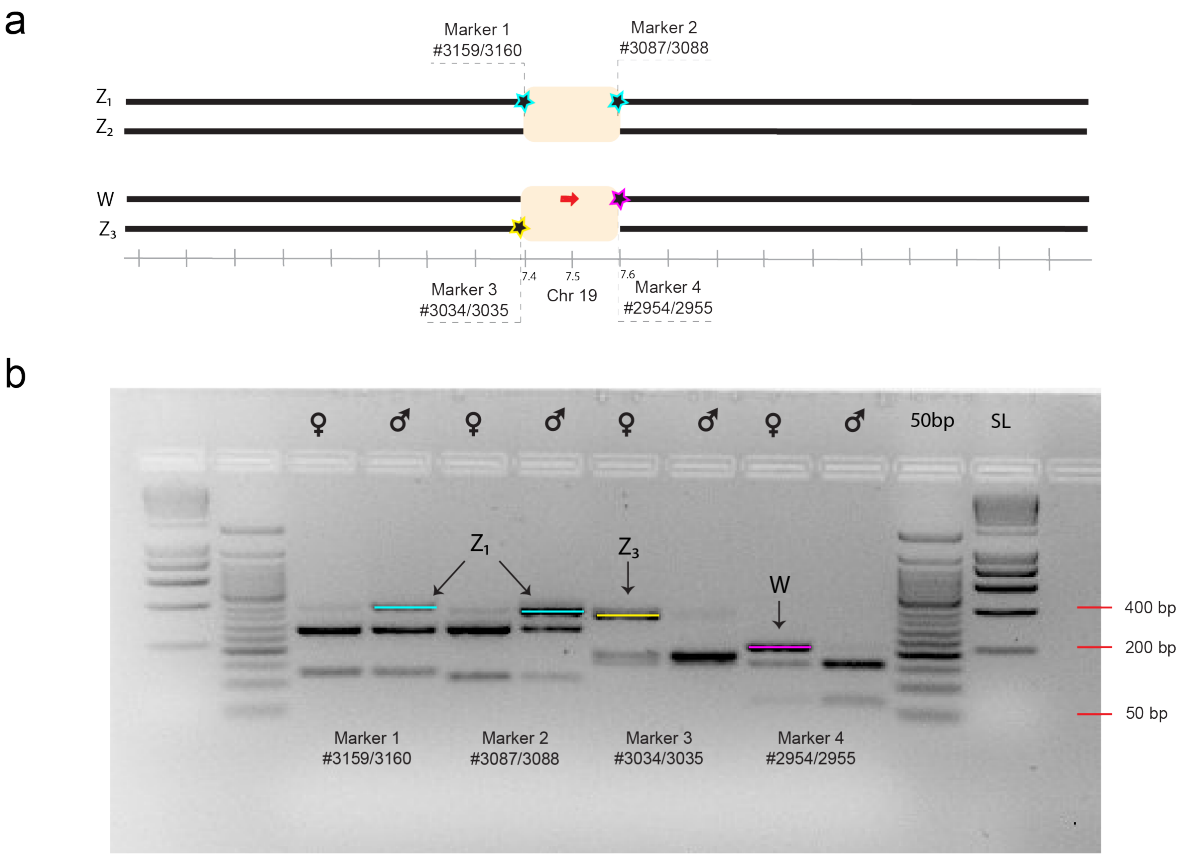


Figure 1. Molecular markers to assess recombination along the white poplar SDR. (a) Marker locations flanking the white polar SDR (~200 kb) located in the middle of chromosome 19. The red arrow represents the W-specific *ARR17* gene. **(b)** The parents have two markers each that can be differentiated after PCR, restriction enzyme digestion and visualized on agarose gels. Ladders: 50 bp and SL (smart ladder). The unique fragments of the haplotypes were highlighted: cyan for Z_1, yellow for Z_3 and magenta for W.

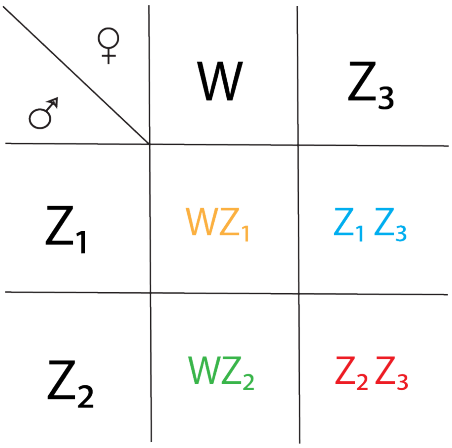


Figure 2. Possible chromosomal combinations in the offspring from a white polar Z_1/Z_2 and WZ_3 cross. *Populus alba* ZZ-males crossed with *P. alba* WZ-females should produce 50% female and 50% male offspring. The segregation patterns can produce four different gametic combinations: Females: WZ_1, WZ_2 and males Z_1Z_3 and Z_2Z_3.

Thus, we can follow the different chromosomes into the progeny and look for recombination between WZ and ZZ (Figure 3).

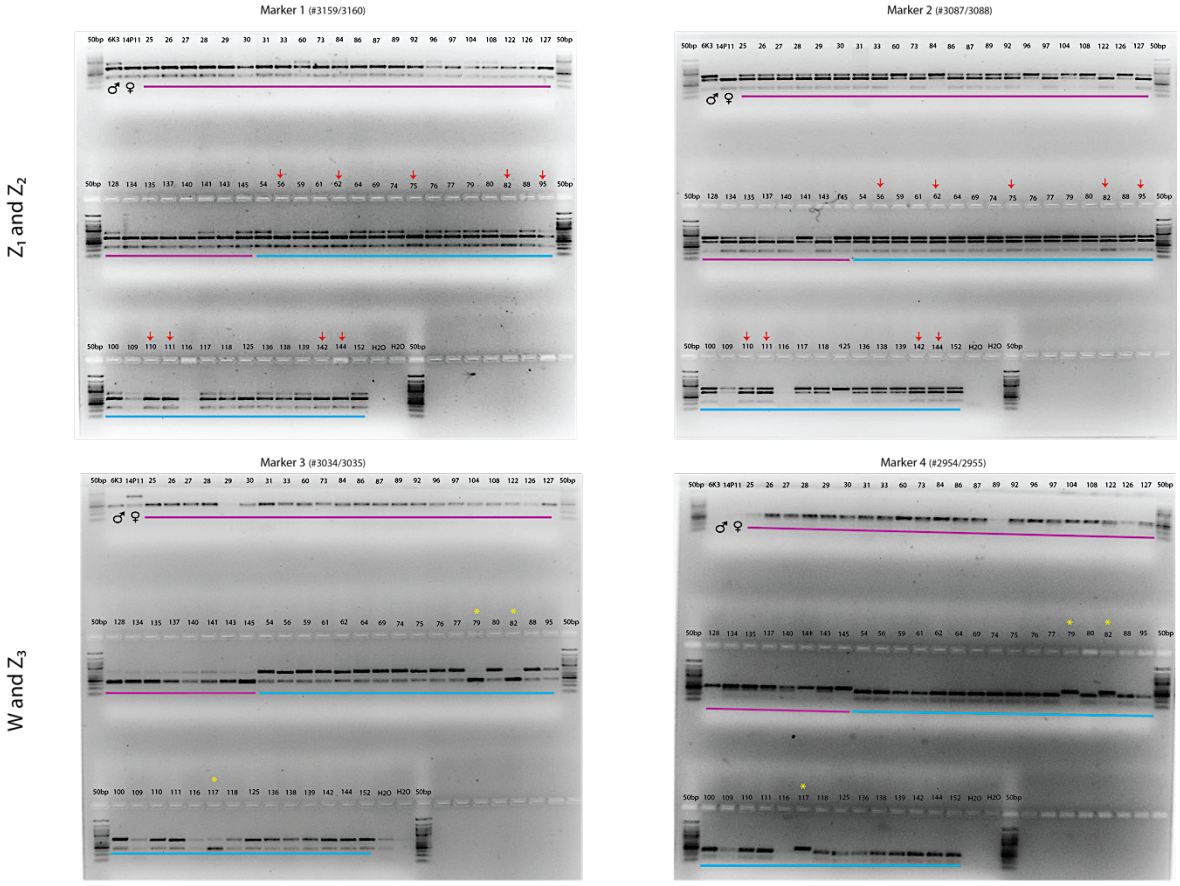


Figure 3. Genotyping after enzymatic digestion in the *Populus alba* F1 population. Markers 1 and 2 assess recombination of Z_1 and Z_2, while markers 3 and 4 of W and Z_3. A set of 60 samples, 30 females (pink) and 30 males (blue) were genotyped. ZZ recombinants are highlighted with a red arrow. No recombinants were found for WZ. Outlier individuals are marked with a yellow asterisk.

From a total of 60 individuals (30 females and 30 males), it was possible to analyze the recombination frequencies and segregation patterns in 56. Our results show 9 samples as possible recombinants of Z_1 and Z_2, while no recombination was found among the W and Z (Figure 3). The absence of *ARR17* from the Z chromosome characterizes a small hemizygous SDR and results in recombination suppression between Z and W.

Suppression of recombination can facilitate the evolution of the sex chromosomes (Stapley et al., 2017). For instance, the translocation of a hemizygous sex-determining region (SDR) results in recombination repression in strawberries (Tennessen et al., 2018). In persimmon, a duplication event resulted in supressed recombination and the maintenance of a small SDR (Akagi et al., 2014). Here, the turnover event in *P. alba* facilitated the change in heterogamety and the evolution of a new sex-determining region (Charlesworth, 2021) favouring a precise and flexible sex-determining mechanism, and the use of the W-specific *ARR17* as a robust molecular sex marker (Kim et al., 2021). Moreover, since there is no antagonistic sex-linked gene, a larger SDR in white poplar is not likely to evolve, similarly to what has been noted in persimmon (Akagi and Charlesworth, 2019).

Additionally, based on the fragments unique to one of the haplotypes (Figure 1b) we can identify the gametic segregation patterns of Z_1/Z_2 and WZ_3 in the samples (Table 2).

Table 2. Resulting gametic combinations of each sample (Figure 2) based on the experimental results (Figure 3). The column 'sex' represents the phenotype of each individual, where F means female and M male. Outlier samples are marked with an asterisk.

F1 sample	Sex	Z_1 or Z_2	W or Z_3
14P11	F	-	WZ
6K3	M	ZZ	-
25	F	Z_2	W
26	F	Z_2	W
27	F	Z_2	W
28	F	Z_2	W
29	F	Z_2	W
30	F	Z_2	W

31	F	Z_2	W
33	F	Z_2	W
73	F	Z_2	W
84	F	Z_2	W
86	F	Z_2	W
87	F	Z_2	W
92	F	Z_2	W
97	F	Z_2	W
122	F	Z_2	W
127	F	Z_2	W
134	F	Z_2	W
135	F	Z_2	W
137	F	Z_2	W
140	F	Z_2	W
143	F	Z_2	W
60	F	Z_1	W
96	F	Z_1	W
104	F	Z_1	W
108	F	Z_1	W
126	F	Z_1	W
128	F	Z_1	W
141	F	Z_1	W
145	F	Z_1	W
89	F	-	-
54	M	Z_1	Z_3
59	M	Z_1	Z_3
61	M	Z_1	Z_3
64	M	Z_1	Z_3
69	M	Z_1	Z_3
74	M	Z_1	Z_3
76	M	Z_1	Z_3
77	M	Z_1	Z_3
80	M	Z_1	Z_3
88	M	Z_1	Z_3
100	M	Z_1	Z_3

118	M	Z_1	Z_3
136	M	Z_1	Z_3
138	M	Z_1	Z_3
139	M	Z_1	Z_3
152	M	Z_1	Z_3
79*	M	Z_1	W
117*	M	Z_1	W
56	M	Z_2/Z_1	Z_3
62	M	Z_2/Z_1	Z_3
75	M	Z_2/Z_1	Z_3
82*	M	Z_2/Z_1	W
95	M	Z_2/Z_1	Z_3
110	M	Z_2/Z_1	Z_3
111	M	Z_2/Z_1	Z_3
142	M	Z_2/Z_1	Z_3
144	M	Z_2/Z_1	Z_3
116	M	-	-
125	M	-	-
109	M	-	-

In markers 1 and 2, the (undigested) upper band is defined as Z_1 (Figure 1b). Based on this, any offspring with the Z_1 fragment in both markers does not recombine. Recombination occurs if the Z_1 fragment is present for one of the markers but not for the other marker, which can either be Z_1/Z_2 or Z_2/Z_1 (Table 2). Notably, all 9 recombinant haplotypes are Z_2/Z_1. Theoretically, both recombination patterns should be equally frequent. Upon a closer look, it is possibly that an artifact in the female Z_3 allele results in the upper band not being specific to Z_1. Thus, we cannot yet draw conclusions whether recombination is suppressed or not in the region between Z_1 and Z_2.

In marker 3, the unique fragment represents Z_3 (Figure 1b). Thus, only males in marker 3 present an upper band (Figure 3). All females only have the lower band, and no Z_3, meaning that only Z_1 or Z_2 comes with the W. On the other hand, in marker 4 the upper fragment represents the W chromosome (Figure 1b). Clearly, all females present the

upper band for this marker (Figure 3). There would be only a recombinant if a given sample presents W and Z_3 together.

Interestingly, the individuals 79, 82 and 117 represents exceptions. Even though these individuals were phenotypically identified as males, our analysis demonstrated an opposite situation since they carry a W chromosome (upper band for marker 4). While we cannot ignore the possibility of a technical mistake with the DNA, it is also conceivable that in this population the genotype is not totally fixed by the genetics, and inconstant males occur. This phenomenon has been previously reported in different dioecious plant species, including *Populus.* Recently, a long-term study revealed F1 individuals that started as males, and gradually turned females (Sabatti et al., 2020).

Apart from the identification of recombinants, the segregation patterns also give insights regarding incompatibility between maternal and paternal alleles. Among the females, 21 individuals presented the Z_2W pattern compared to only 8 individuals with Z_1W, suggesting a potential quantitative incompatibility of Z_1 and W. Among the males, 16 individuals presented Z_1. However, no Z_2 chromosome was observed together with the Z_3 inherited from the mother, which again appears to be caused by a potential problem with marker 2. Incompatibilities between gametes can also lead to biased sex ratios, which has been reported also for our mapping population (Tuskan et al., 2012). For instance, allelic incompatibility has been shown to be responsible for a biased sex ratio in *Salix viminalis* L. (Pucholt et al., 2017). Even though our work considered a subset of the F1 population of this *P. alba* cross in equal numbers of males and females, a male-sex ratio bias was reported, with 2 times more males than females (Pakull et al., 2011). Further screenings utilizing these CAPS markers in larger populations are necessary to confirm these hypotheses.

Finally, further research about the evolution of the SDRs in the genus *Populus* should also target the aspen SDR, in which the recombination is suppressed in similar rates between the X chromosomes in females and between X and Y in males (Kersten et al., 2014). Preliminary attempts to design CAPS markers resulted in inconsistent patterns so far, probably, due to the central position which shows reduced mapping efficiency, likely because the proximity to the centromere (Stapley et al., 2017). It will be exciting to investigate whether recombination suppression was indeed the ancestral state in the region where the SDR in aspens evolved (Charlesworth, 2021; Rifkin et al., 2021).

References

Akagi, T., Charlesworth, D., 2019. Pleiotropic effects of sex-determining genes in the evolution of dioecy in two plant species. Proc. R. Soc. B Biol. Sci. 286, 20191805. https://doi.org/10.1098/rspb.2019.1805

Akagi, T., Henry, I.M., Tao, R., Comai, L., 2014. A Y-chromosome–encoded small RNA acts as a sex determinant in persimmons. Science 346, 646–650. https://doi.org/10.1126/science.1257225

Allen, C.E., 1917. A chromosome difference correlated with sex differences in Sphærocarpos. Science. https://doi.org/10.1126/science.46.1193.466

Bolger, A.M., Lohse, M., Usadel, B., 2014. Trimmomatic: a flexible trimmer for Illumina sequence data. Bioinformatics 30, 2114–2120. https://doi.org/10.1093/bioinformatics/btu170

Charlesworth, B., Charlesworth, D., 1978. A model for the evolution of dioecy and gynodioecy. Am. Nat. 112, 975–997. https://doi.org/10.1086/283342

Charlesworth, D., 2021. When and how do sex-linked regions become sex chromosomes? Evolution 75, 569–581. https://doi.org/10.1111/evo.14196

Charlesworth, D., Charlesworth, B., Marais, G., 2005. Steps in the evolution of heteromorphic sex chromosomes. Heredity 95, 118–128. https://doi.org/10.1038/sj.hdy.6800697

Fechter, I., Hausmann, L., Daum, M., Rosleff Sörensen, T., Viehöver, P., Weisshaar, B., Töpfer, R., 2012. Candidate genes within a 143 kb region of the flower sex locus in *Vitis*. Mol. Genet. Genomics 287, 247–259. https://doi.org/10.1007/s00438-012-0674-z

Furman, B.L.S., Metzger, D.C.H., Darolti, I., Wright, A.E., Sandkam, B.A., Almeida, P., Shu, J.J., Mank, J.E., 2020. Sex chromosome evolution: so many exceptions to the rules. Genome Biol. Evol. 12, 750–763. https://doi.org/10.1093/gbe/evaa081

Geraldes, A., Hefer, C.A., Capron, A., Kolosova, N., Martinez-Nuñez, F., Soolanayakanahally, R.Y., Stanton, B., Guy, R.D., Mansfield, S.D., Douglas, C.J., Cronk, Q.C.B., 2015. Recent Y chromosome divergence despite ancient origin of dioecy in poplars (*Populus*). Mol. Ecol. 24, 3243–3256. https://doi.org/10.1111/mec.13126

Golenberg, E.M., West, N.W., 2013. Hormonal interactions and gene regulation can link monoecy and environmental plasticity to the evolution of dioecy in plants. Am. J. Bot. 100, 1022–1037. https://doi.org/10.3732/ajb.1200544

Harkess, A., Zhou, J., Xu, C., Bowers, J.E., Van der Hulst, R., Ayyampalayam, S., Mercati, F., Riccardi, P., McKain, M.R., Kakrana, A., Tang, H., Ray, J., Groenendijk, J., Arikit, S., Mathioni, S.M., Nakano, M., Shan, H., Telgmann-Rauber, A., Kanno, A., Yue, Z., Chen, Haixin, Li, W., Chen, Y., Xu, X., Zhang, Y., Luo, S., Chen, Helong, Gao, J., Mao, Z., Pires, J.C., Luo, M., Kudrna, D., Wing, R.A., Meyers, B.C., Yi, K., Kong, H., Lavrijsen, P., Sunseri, F., Falavigna, A., Ye, Y., Leebens-Mack, J.H., Chen, G., 2017. The asparagus genome sheds light on the origin and evolution of a young Y chromosome. Nat. Commun. 8, 1279. https://doi.org/10.1038/s41467-017-01064-8

Hobza, R., Hudzieczek, V., Kubat, Z., Cegan, R., Vyskot, B., Kejnovsky, E., Janousek, B., 2018. Sex and the flower – developmental aspects of sex chromosome evolution. Ann. Bot. 122, 1085–1101. https://doi.org/10.1093/aob/mcy130

Kamiya, T., Kai, W., Tasumi, S., Oka, A., Matsunaga, T., Mizuno, N., Fujita, M., Suetake, H., Suzuki, S., Hosoya, S., Tohari, S., Brenner, S., Miyadai, T., Venkatesh, B., Suzuki, Y., Kikuchi, K., 2012. A Trans-species missense SNP in *Amhr2* is associated with sex determination in the Tiger Pufferfish, Takifugu rubripes (Fugu). PLOS Genet. 8, e1002798. https://doi.org/10.1371/journal.pgen.1002798

Kersten, B., Pakull, B., Groppe, K., Lueneburg, J., Fladung, M., 2014. The sex-linked region in *Populus tremuloides* Turesson 141 corresponds to a pericentromeric region of about two million base pairs on *P. trichocarpa* chromosome 19. Plant Biol. 16, 411–418. https://doi.org/10.1111/plb.12048

Kim, G., Leite Montalvão, A.P., Kersten, B., Fladung, M., Müller, N.A., 2021. The genetic basis of sex determination in *Populus* provides molecular markers across the genus and indicates convergent evolution. Silvae Genet. 70, 145–155. https://doi.org/10.2478/sg-2021-0012

Knaus, B.J., Grünwald, N.J., 2016. VcfR: a package to manipulate and visualize VCF format data in R (preprint). Bioinformatics. https://doi.org/10.1101/041277

Leite Montalvão, A.P., Kersten, B., Fladung, M., Müller, N.A., 2021. The diversity and dynamics of sex determination in dioecious plants. Front. Plant Sci. 11, 1-12. https://doi.org/10.3389/fpls.2020.580488

McKenna, A., Hanna, M., Banks, E., Sivachenko, A., Cibulskis, K., Kernytsky, A., Garimella, K., Altshuler, D., Gabriel, S., Daly, M., DePristo, M.A., 2010. The Genome Analysis Toolkit: A MapReduce framework

for analyzing next-generation DNA sequencing data. Genome Res. 20, 1297–1303. https://doi.org/10.1101/gr.107524.110

Müller, N.A., Kersten, B., Leite Montalvão, A.P., Mähler, N., Bernhardsson, C., Bräutigam, K., Carracedo Lorenzo, Z., Hoenicka, H., Kumar, V., Mader, M., Pakull, B., Robinson, K.M., Sabatti, M., Vettori, C., Ingvarsson, P.K., Cronk, Q., Street, N.R., Fladung, M., 2020. A single gene underlies the dynamic evolution of poplar sex determination. Nat. Plants 6, 630–637. https://doi.org/10.1038/s41477-020-0672-9

Paolucci, I., Gaudet, M., Jorge, V., Beritognolo, I., Terzoli, S., Kuzminsky, E., Muleo, R., Scarascia Mugnozza, G., Sabatti, M., 2010. Genetic linkage maps of *Populus alba* L. and comparative mapping analysis of sex determination across *Populus* species. Tree Genet. Genomes 6, 863–875. https://doi.org/10.1007/s11295-010-0297-7

Pucholt, P., Wright, A.E., Conze, L.L., Mank, J.E., Berlin, S., 2017. Recent sex chromosome divergence despite ancient dioecy in the willow *Salix viminalis*. Mol. Biol. Evol. 34, 1991–2001. https://doi.org/10.1093/molbev/msx144

Renner, S.S., Müller, N.A., 2021. Plant sex chromosomes defy evolutionary models of expanding recombination suppression and genetic degeneration. Nat. Plants 7, 392–402. https://doi.org/10.1038/s41477-021-00884-3

Rifkin, J.L., Hnatovska, S., Yuan, M., Sacchi, B.M., Choudhury, B.I., Gong, Y., Rastas, P., Barrett, S.C.H., Wright, S.I., 2021. Recombination landscape dimorphism contributes to sex chromosome evolution in the dioecious plant *Rumex hastatulus*. bioRxiv 29. https://doi.org/10.1101/2021.11.03.466946

Sabatti, M., Gaudet, M., Müller, N.A., Kersten, B., Gaudiano, C., Scarascia Mugnozza, G., Fladung, M., Beritognolo, I., 2020. Long-term study of a subdioecious *Populus* × *canescens* family reveals sex lability of females and reproduction behaviour of cosexual plants. Plant Reprod. 33, 1–17. https://doi.org/10.1007/s00497-019-00378-5

Sardell, J.M., Kirkpatrick, M., 2020. Sex differences in the recombination landscape. Am. Nat. 195, 361–379. https://doi.org/10.1086/704943

Stapley, J., Feulner, P.G.D., Johnston, S.E., Santure, A.W., Smadja, C.M., 2017. Variation in recombination frequency and distribution across eukaryotes: patterns and processes. Philos. Trans. R. Soc. Lond. B. Biol. Sci. 372, 20160455. https://doi.org/10.1098/rstb.2016.0455

Stevens, N.M., 1905. Studies in spermatogenesis. Carnegie Institution of Washington.

Tennessen, Jacob A., Wei, N., Straub, S.C.K., Govindarajulu, R., Liston, A., Ashman, T.-L., 2018. Repeated translocation of a gene cassette drives sex-chromosome turnover in strawberries. PLOS Biol. 16, e2006062. https://doi.org/10.1371/journal.pbio.2006062

Tuskan, G.A., DiFazio, S., Faivre-Rampant, P., Gaudet, M., Harfouche, A., Jorge, V., Labbé, J.L., Ranjan, P., Sabatti, M., Slavov, G., Street, N., Tschaplinski, T.J., Yin, T., 2012. The obscure events contributing to the evolution of an incipient sex chromosome in *Populus*: a retrospective working hypothesis. Tree Genet. Genomes 8, 559–571. https://doi.org/10.1007/s11295-012-0495-6

Wilson, E.B., 1905. Studies on chromosomes. I. The behavior of the idiochromosomes in hemiptera. J. Exp. Zool. 2, 371–405. https://doi.org/10.1002/jez.1400020305

Yang, H.-W., Akagi, T., Kawakatsu, T., Tao, R., 2019. Gene networks orchestrated by *MeGI*: a single-factor mechanism underlying sex determination in persimmon. Plant J. 98, 97–111. https://doi.org/10.1111/tpj.14202

Yang, W., Wang, D., Li, Y., Zhang, Z., Tong, S., Li, M., Zhang, X., Zhang, L., Ren, L., Ma, X., Zhou, R., Sanderson, B.J., Keefover-Ring, K., Yin, T., Smart, L.B., Liu, J., DiFazio, S.P., Olson, M., Ma, T., 2021. A general model to explain repeated turnovers of sex determination in the Salicaceae. Mol. Biol. Evol. 38, 968–980. https://doi.org/10.1093/molbev/msaa261

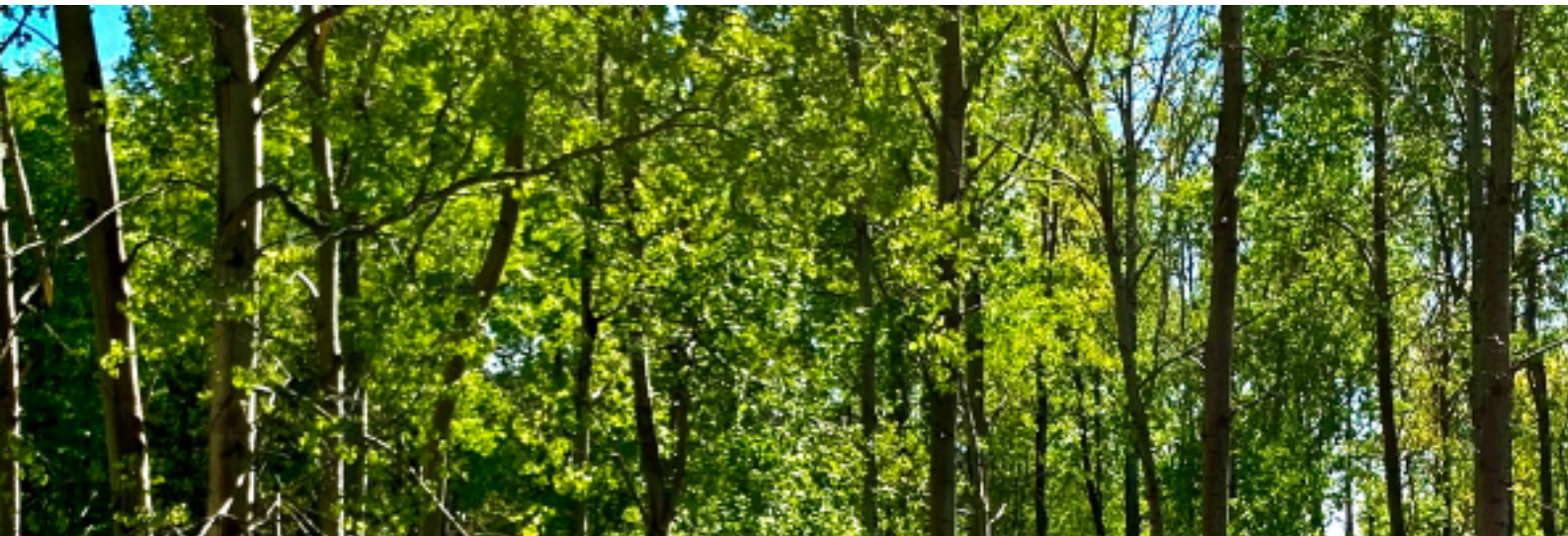
Chapter 4

The genetic basis of sex determination in *Populus* provides markers across the genus and indicates convergent evolution

Gihwan Kim[1], Ana P. Leite Montalvão[1], Birgit Kersten, Matthias Fladung & Niels A. Müller[*]

Thünen Institute of Forest Genetics, Großhansdorf, Germany
[1] These authors contributed equally
[*] Corresponding author

Published in *Silvae Genetica* (2021)
doi:10.2478/sg-2021-0012



Abstract

Many dioecious angiosperms are trees, which only flower after years of vegetative development and do not usually exhibit marked secondary sexual dimorphism. Nevertheless, if the genetic basis of sex determination is known, the sex of an individual can be determined using molecular markers. Here, we report that in the genus *Populus* sect. Populus an XY system of sex determination, which is found in *P. tremula* and *P. tremuloides*, likely re-evolved from a ZW system present in *P. alba*, *P. adenopoda* and *P. qiongdaoensis*. Strikingly, this new XY system is mechanistically identical to the older system found in several species of the *Populus* sections Tacamahaca, Aigeiros and Turanga demonstrating a remarkable example of convergent evolution. In both XY systems, male-specific inversely repeated sequences appear to silence the *ARR17* gene, which functions as a sex switch, via small interfering RNAs and DNA methylation. In the ZW system, female-specific copies of *ARR17* appear to regulate dioecy. With this detailed information on the genetic basis of sex determination it was possible to develop molecular markers that can be utilized to determine the sex in seedlings and non-flowering trees of different poplar species. We used the female specific *ARR17* gene to develop a sex marker for *P. alba* and *P. adenopoda*. For *P. grandidentata*, we employed the male specific *ARR17* inverted repeat. Finally, we summarize previously described markers for *P. tremula*, *P. tremuloides*, *P. trichocarpa*, *P. deltoides* and *P. nigra*. These markers can be useful for poplar ecologists, geneticists and breeders.

Keywords: *Populus*, poplar breeding, dioecy, sex-determining region, ARR17, sex chromosome evolution, female and male heterogamety, sex markers, convergent evolution.

Introduction

Unisexuality in plants can be regulated through different molecular mechanisms and has been used as model to study flower development and plant sex determination. Recent studies in different dioecious plant species, such as kiwifruit (Akagi et al., 2019, 2018), garden asparagus (Harkess et al., 2020, 2017), persimmon (Akagi et al., 2016, 2014) or poplar (Müller et al., 2020), have started to clarify some of the molecular mechanisms

of sex determination, improving our knowledge about the evolution of dioecy (Carey et al., 2021; Feng et al., 2020; Leite Montalvão et al., 2021; Renner and Müller, 2021). *Populus,* an economically and ecologically important genus of the *Salicaceae* family, is composed of dioecious species that are collectively called poplars (Slavov and Zhelev, 2010), however, *P. tremula* L., *P. grandidentata* Michx. and *P. tremuloides* Michx. are often termed aspens. Sex determination in poplars and aspens is genetically controlled by sex-determining regions (SDRs) on different chromosomes and with different heterogamety (Geraldes et al., 2015; Kersten et al., 2014; Müller et al., 2020; Yang et al., 2021) and it was demonstrated that at least in aspens it relies on the expression of a single gene (named *ARR17* for consistency with previous reports) that functions as a sex switch (Müller et al., 2020).

Interestingly, transitions between heterogametic systems (XY ↔ ZW) are documented for some genera, including *Populus* (Bhat and Bindroo, 1980; Cormier et al., 2019; He et al., 2021; Martin et al., 2019; Müller et al., 2020; Paolucci et al., 2010; Sanderson et al., 2021; Zhou et al., 2018). The same molecular mechanism of sex determination based on the *ARR17* gene appears to be present in *Populus* species with both, XY and ZW systems. *Populus* species presenting an XY system (Müller et al., 2020; Zhou et al., 2020), e.g. *P. tremula* and *P. tremuloides,* have partial *ARR17*-duplicates arranged as inverted repeats located in the sex-determining region (SDR) of the Y chromosome and, apparently, through small RNAs and DNA methylation, the *ARR17* gene is dominantly silenced in males. In contrast, in the poplar ZW system, e.g. *P. alba* L., sex determination is based on a presence/absence variant of the *ARR17* gene (Müller et al., 2020; Yang et al., 2021).

Revealing the genetic basis of sex determination in different *Populus* species allows to explore the possibilities for sex markers, which can be used to assess sex in trees that are not yet sexually mature or not flowering. For species with long life cycles that exhibit no or minimal sexual dimorphism, such as poplars (McKown et al., 2017; Renner and Müller, 2021; Robinson et al., 2014), molecular markers are important tools for research and breeding programs.

In this work we aimed to (i) establish the W-specific *ARR17* as a female-specific sex marker applicable in poplars with ZW system (*P. alba* and *P. adenopoda* Maxim.), (ii) employ the Y-specific *ARR17* inverted repeat as a male-specific marker in *P.*

grandidentata, and (iii) summarize the previously described sex markers in *P. tremula* and *P. tremuloides* (Y-specific *TOZ19*) (Pakull et al., 2015) and *P. trichocarpa* Torr. & A. Gray ex Hook., *P. deltoides* Marshall and *P. nigra* L. (Y-specific *HEMA1* partial duplicate) (Geraldes et al., 2015). The molecular sex markers that we describe here can reliably discriminate males from females in non-flowering trees of different species across the *Populus* genus which may be useful for ecologists, geneticists, and breeders.

Material and methods

Re-sequencing data analysis

Raw sequencing reads were downloaded from NCBI's SRA (Bioproject PRJNA612655 – accessions SRR11308211 to SRR11308214 and SRR11308190) (Shang et al., 2020) and NGDC's GSA (Bioproject PRJACA001334 – accessions CRR050666 to CRR050668) (Wang et al., 2020) for *P. adenopoda*. From NCBI's SRA (Bioproject PRJNA612655 – accessions SRR11308208 to SRR11308210) (Shang et al., 2020) and NGDC's GSA (Bioproject PRJACA001334 – accessions CRR050678 to CRR050680) (Wang et al., 2020) for *P. qiongdaoensis.* And from NGDC's GSA (Bioproject PRJACA002485 – accessions CRR176864 to CRR176875 and CRR176894 to CRR176905) for *P. alba* (Yang et al., 2021). Reads were trimmed using Trimmomatic (Bolger et al., 2014) with the following parameters: ILLUMINACLIP: TruSeq3-PE-2.fa:2:30:10 LEADING:3 TRAILING:3 SLIDINGWINDOW:4:15 MINLEN:36, and subsequently mapped to the reference genome of female *P. tremula* (Schiffthaler et al., 2019), with bowtie2 (Langmead and Salzberg, 2012) using default parameters. Read coverage was visualized for the previously described sex-linked region using R ("R Core Team," 2018) with the packages Rsamtools (Arora et al., 2021), GenomicRanges (Lawrence et al., 2013), zoo (Zeileis and Grothendieck, 2005), ggplot2 (Wickham, 2016) and tidyverse (Wickham et al., 2019).

Sex-linked markers

PCR experiments were performed in 25 µl reactions, except the PCR-RFLP assay, which was done in 10 µl. The PCR control TOZ13-1, previously described by Pakull et al. (2015), was used as PCR positive control in combination with the sex markers ARR17, ARR17_IR

or TOZ19-2, to confirm the successful PCR amplification. Primer sequences are given in Table 1. Genomic DNA (gDNA) from *Populus alba* (10 individuals), *P. adenopoda* (5), *P. × tomentosa* Carrière (6), *P. grandidentata* (10), *P. tremula* (10), *P. tremuloides* (10), *P. trichocarpa* (9), *P. deltoides* (10) and *P. nigra* (10) was extracted from leaves. For all PCR reactions, 100ng of gDNA, 200 µM of dNTP, primers at 0.24 µM (ARR17, ARR17_IR and TOZ9-2) and 0.16 µM (TOZ13-1) and 0.2 units of Taq polymerase were used with the exception that proofreading DNA polymerase (i.e., Phusion™ High-Fidelity DNA Polymerase) and primers at 0.5 µM were used for the PCR-RFLP assay.

P. alba, P. adenopoda and P. × tomentosa

For the ARR17 marker, which can be used to determine the female-specific presence of the W chromosome, the PCR started with a 3 min denaturation at 95 °C, followed by 30 cycles of 30 sec denaturation at 95 °C, 30 sec annealing at 55 °C, 1 min extension at 72 °C and a final extension of 2 min at 72 °C.

P. grandidentata

For the *ARR17* inverted repeat marker (ARR17_IR), which is based on the Y chromosome specific *ARR17* inverted repeat sequence of *P. grandidentata*, the following program was used: 3 min denaturation at 94 °C, followed by 38 cycles of 30 sec denaturation at 94 °C, 45 sec annealing at 55 °C, 90 sec extension at 72 °C and a final extension of 10 min at 72 °C.

P. tremula and P. tremuloides

For the previously described *TOZ19* marker (Pakull et al., 2015), the PCR program followed a 3 min denaturation step at 94 °C, 42 cycles of 30 sec denaturation at 94 °C, then annealing at 50 °C for 45 sec, extension for 1 min at 72 °C and a final extension at 72 °C for 10 min.

P. trichocarpa, P. deltoides and P. nigra

The PCR-RFLP assay (Geraldes et al., 2015), here named HEMA1_*Tsp*RI, was performed using the following program: denaturation at 98 °C for 30 sec followed by 35 cycles of 10 sec denaturation at 98 °C, annealing at 58 °C for 30 sec, extension at 72 °C for 30 sec and then a final extension for 5 min at 72 °C. Amplicons were digested with the enzyme *Tsp*RI (New England Biolabs, Ipswich, MA), using 5 µL of PCR product, 0.5 µL of *Tsp*RI, 1 µL of buffer and water to reach a final volume of 10 µL. The reaction was incubated at 37 °C overnight. On an agarose gel, females (XX) are expected to present PCR products at 382 bp and 178 bp, while males (XY) additionally have Y-specific fragments at 279 bp and 281 bp.

All PCR products were checked by gel electrophoresis at 100 volts for 60 minutes using a 1% or 1.5% agarose gel with 0.1 x TAE as a running buffer. The DNA was stained with RotiSafe from Roth (Karlsruhe, Germany).

Table 1. PCR primer sequences designed and used to determine sex in different *Populus* species. The marker TOZ13-1 was used as a PCR control for *P. alba*, *P. adenopoda*, *P.* × *tomentosa*, *P. grandidentata*, *P. tremula* and *P. tremuloides*.

Primer name	Forward primer (5' → 3')	Reverse primer (5' → 3')	Size of PCR product (bp)	*Populus* species
ARR17	TCAGTGTACATGTCTAATGACAAGC	AGCCCTAGAATTACGCCTCC	826	*P. alba*, *P. adenopoda*, *P.* × *tomentosa* (W chromosome)
ARR17_IR	AGAGAGCATTGGAGTATTTGGG	GTTGAGGTGGTTAGACATTGTGG	400	*P. grandidentata* (Y chromosome)
TOZ19-2	GACGCCATCAAGATTGTGGATTCACCA	GTATCAGGATGGAACATGAGAGTAGTTACG	500	*P. tremula*, *P. tremuloides* (Y chromosome)
HEMA1_*Tsp*RI	TGATCATGGGCATTATAACCAA	TGGACAATGGTCAAACAGTCC	559	*P. trichocarpa*, *P. deltoides*, *P. nigra* (Y chromosome)
TOZ13-1	TTAGGTGCTGATGGTTTGGTAAAGCTA	CTTGCATGTGGATCATCAACTCAAGATCA	260	PCR control (autosome)

Results

P. adenopoda and *P. qiongdaoensis* appear to feature a ZW system of sex determination

To date, the only described poplar species that exhibits female heterogamety is white poplar (*P. alba*) (Müller et al., 2020; Paolucci et al., 2010; Yang et al., 2021). Sex determination in *P. alba* relies on a presence/absence mutation of the *ARR17* gene. In females (ZW), three complete copies of *ARR17* can be found in the SDR of the W chromosome, while *ARR17* is absent from the Z chromosome. Males (ZZ) thus do not carry any *ARR17* sequence (Müller et al., 2020; Yang et al., 2021).

To test whether the same ZW system could be present in sister species of *P. alba* we downloaded publicly available next generation sequencing (NGS) data of *P. alba*, *P. adenopoda* and *P. qiongdaoensis* and mapped it to the *P. tremula* v2.2 genome (Schiffthaler et al., 2019), which represents a high-quality reference genome assembly for the section Populus. This allowed us to specifically analyze the region where the master regulator of sex determination, *ARR17*, is located. The normalized read coverage of several *P. alba* samples demonstrated female-specific coverage in the *ARR17* region as reported before (Figure 1a) (Müller et al., 2020). Strikingly, a near identical pattern can be observed in *P. adenopoda* (Figure 1b) suggesting that the origin of the sex-determining systems of *P. adenopoda* and *P. alba* may be shared. Interestingly, the outgroup species of *P. alba* and *P. adenopoda*, *P. qiongdaoensis* also appears to feature a ZW system of sex determination (Figure 1c). Putative males only carry a small part of the *ARR17* gene. Additional female-specific hemizygous regions around the *ARR17* gene are shared with *P. alba* and *P. adenopoda*.

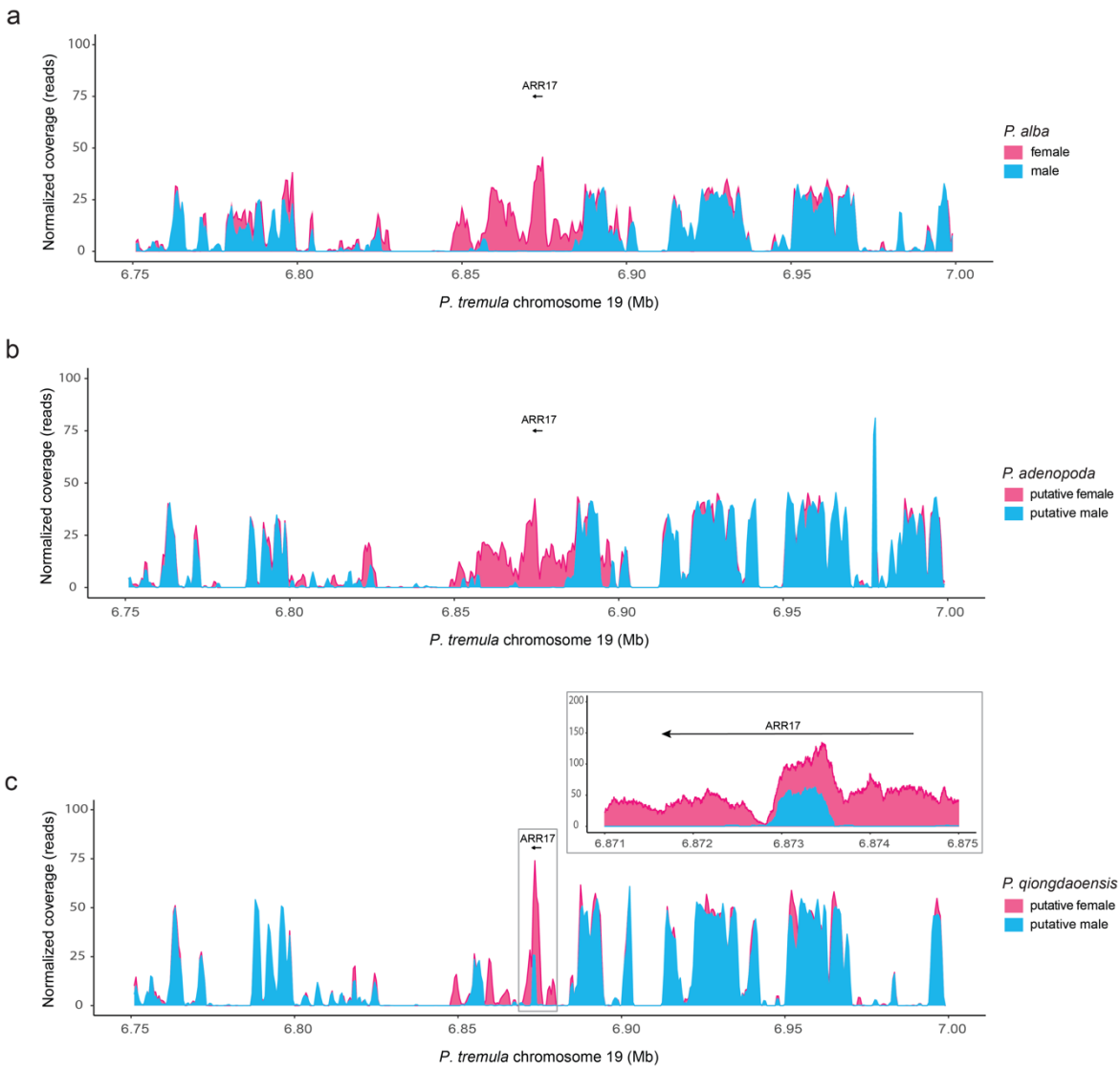


Figure 1. Mean normalized DNA-seq coverage in 1000-bp windows (with a step size of 500 bp) along the *ARR17* genomic region of the *P. tremula* reference genome (Schiffthaler et al., 2019). Sequencing reads were filtered for a mapping quality >10. *ARR17* is represented by a black arrow. (a) White poplar (*P. alba*) (12 female samples and 12 male samples). (b) Chinese aspen (*P. adenopoda*) (3 putative females and 5 putative males, grouped by the presence or absence of *ARR17*). (c) *P. qiongdaoensis* (4 putative female samples and 2 putative male samples, grouped by the presence or absence of *ARR17*). Zoom of *ARR17* region in panel (c) presents the raw sequencing coverage instead of an average in sliding windows.

To provide further evidence for the putative female heterogametic system in *P. adenopoda*, we tested a fragment of the *ARR17* gene as a female sex-marker (ARR17), while employing a fragment of the *TOZ13* gene (TOZ13-1) as PCR control (Pakull et al., 2015). All samples are expected to show a PCR product for TOZ13-1 and only females should present a PCR product for *ARR17*. We tested the marker on individuals of *P. alba* (Figure 2a) and *P. adenopoda* (Figure 2b), and on individuals of the Chinese white poplar *P.* × *tomentosa* (Figure 2b), which is considered a hybrid between *P. alba* and *P.*

adenopoda (Wang et al., 2019) . While the marker perfectly distinguishes between males and females based on presence/absence of *ARR17* in *P. alba* and *P. × tomentosa* (Figure 2), all of the *P. adenopoda* individuals, which were of unknown sex, failed to show an *ARR17* PCR product. Since the genomic region amplified by the ARR17 marker shows no indication of structural variation and no SNPs or indels in the primer binding sites in the *P. adenopoda* re-sequencing data, our samples are probably all males. Taken together, our results indicate that the ZW systems of sex determination of *P. alba*, *P. adenopoda* and *P. qiongdaoensis* might have a shared origin. Nevertheless, further phylogenetic and synteny analyses should be performed to precisely characterize the SDRs of *P. adenopoda* and *P. qiongdaoensis*. In any case, our results demonstrate that *ARR17* can be used as a sex marker in *P. alba*, *P. adenopoda* and their hybrids.

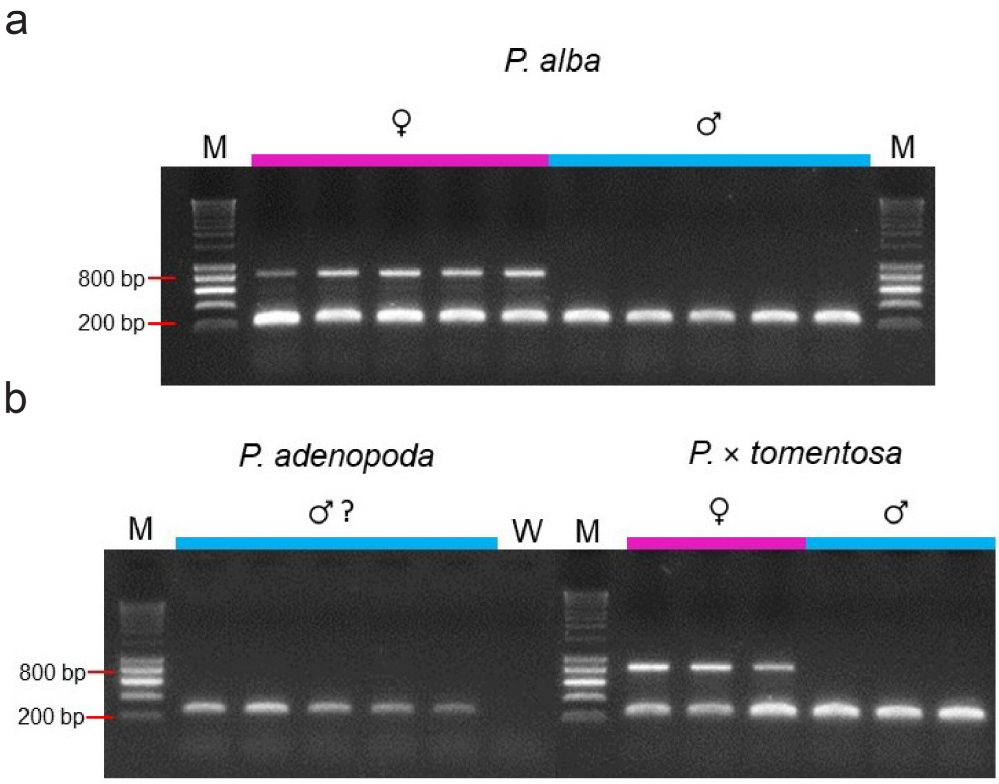


Figure 2. PCR amplifications of the gene *ARR17* in *Populus* species with a ZW system of sex determination. (a) *P. alba*, (b) *P. adenopoda* and *P. × tomentosa*. Samples were run on a 1% agarose gel. The upper PCR product represents ARR17 and the lower PCR product TOZ13-1. M = Smart Ladder (Eurogentech, Cologne, Germany), W = water control. Fragment sizes of the DNA ladder are given on the left side. The *P. adenopoda* were of unknown sex, while the sex of all other individuals is documented.

The Y chromosome specific *ARR17* inverted repeat as a sex marker in *P. grandidentata*

Few studies have been carried out in the bigtooth aspen *Populus grandidentata*. The described male-specific aspen sex-linked marker TOZ19 (Pakull et al., 2015) did not yield successful PCR amplification in this species. Therefore, to develop an alternative molecular sex marker for *P. grandidentata*, we decided to test the *ARR17* inverted repeat sequences (ARR17_IR), which are located in the male-specific region of the Y chromosome in the aspen species *P. tremula* and *P. tremuloides*. Sanger sequencing of a part of the *P. grandidentata* ortholog of Potri.019G047600, a gene located close to *TOZ19*, identified a SNP (at position 6,721,236 bp of chromosome 19 of the poplar reference genome Potri v3.0 (Tuskan et al., 2006)) that was heterozygous (A/G) in all 19 male samples tested but homozygous (G/G) in all 48 females. These results indicated an XY system of sex determination. Since the male-specific heterozygous variant is located within the aspen SDR based on the aspen reference genome, *P. grandidentata* may share the same genetic basis of sex determination with *P. tremula* and *P. tremuloides*. We therefore tested several *ARR17* primer pairs comprising two forward primers, that can only work on an inversely repeated sequence. One of these primer pairs, here termed ARR17_IR, successfully amplified part of the male-specific *ARR17* inverted repeat sequence in male *P. grandidentata* individuals but did not yield any PCR product in *P. tremula* or *P. tremuloides*. This male-specific amplification can be utilized to discriminate female and male *P. grandidentata* individuals (Figure 3).

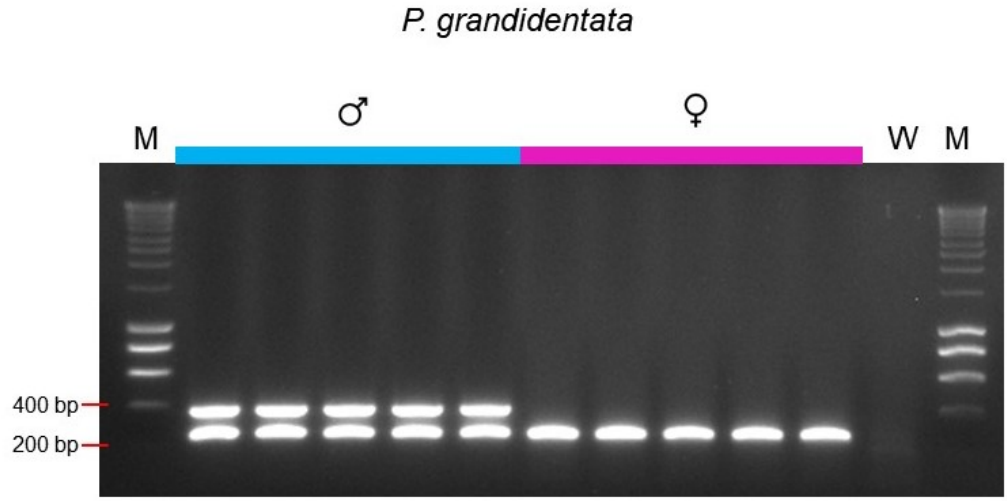


Figure 3. PCR amplification of Y chromosome-specific *ARR17* inverted repeat sequences as a molecular marker to determine sex in *P. grandidentata*. Samples were run on a 1% agarose gel. The upper PCR product represents the *ARR17* inverted repeat (ARR17_IR) and the lower PCR product the TOZ13-1 control. M = Smart Ladder (Eurogentech, Cologne, Germany), W = water control. Fragment sizes of the DNA ladder are given on the left side.

Previously described sex markers in different *Populus* species

The sex of non-flowering aspens can be determined using the male-specific sex-linked gene *TOZ19*, which has been validated in over 80 samples in previous work (Pakull et al., 2015). Through a PCR amplification, sex in *P. tremula* and *P. tremuloides* can be determined. Here we reproduced these results, where all samples presented a control PCR product (TOZ13-1) and only males presented a PCR product for the *TOZ19* sequence (ToZ19-2) (Figure 4).

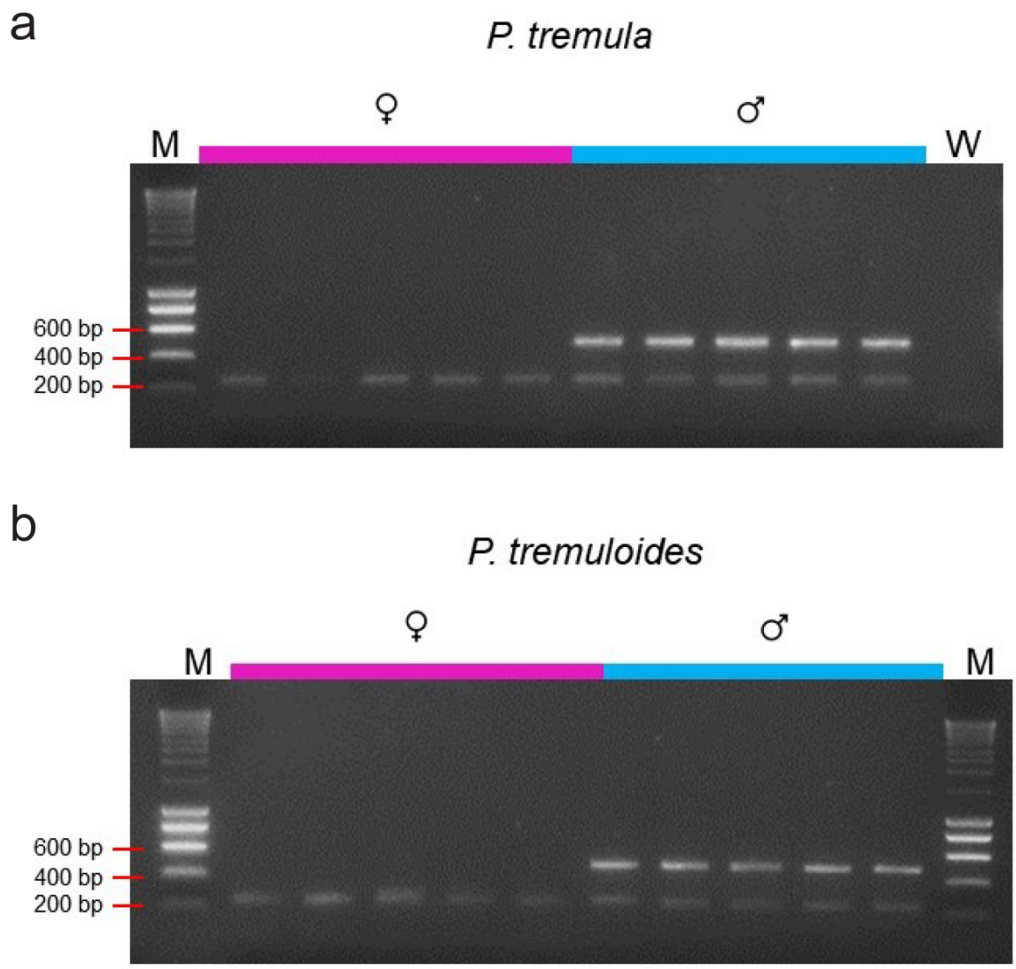


Figure 4. PCR amplification of *TOZ19* in the aspens (a) *P. tremula* and (b) *P. tremuloides*. Samples were run on a 1% agarose gel. The upper PCR product represents the TOZ19-2 and the lower PCR product the TOZ13-1 control. M = Smart Ladder (Eurogentech, Cologne, Germany), W = water control. Fragment sizes of the DNA ladder are given on the left side.

Another sex-linked marker has been developed for genotyping of a male-specific single nucleotide polymorphisms (SNPs), which is located in a Y-chromosomal partial duplicate of the *HEMA1* gene, in *P. trichocarpa, P. deltoides and P. nigra* (Geraldes et al., 2015). We validated this PCR-RFLP marker in independent samples (Figure 5). The fragments match to the *in silico* prediction and the overall pattern is completely reproducible.

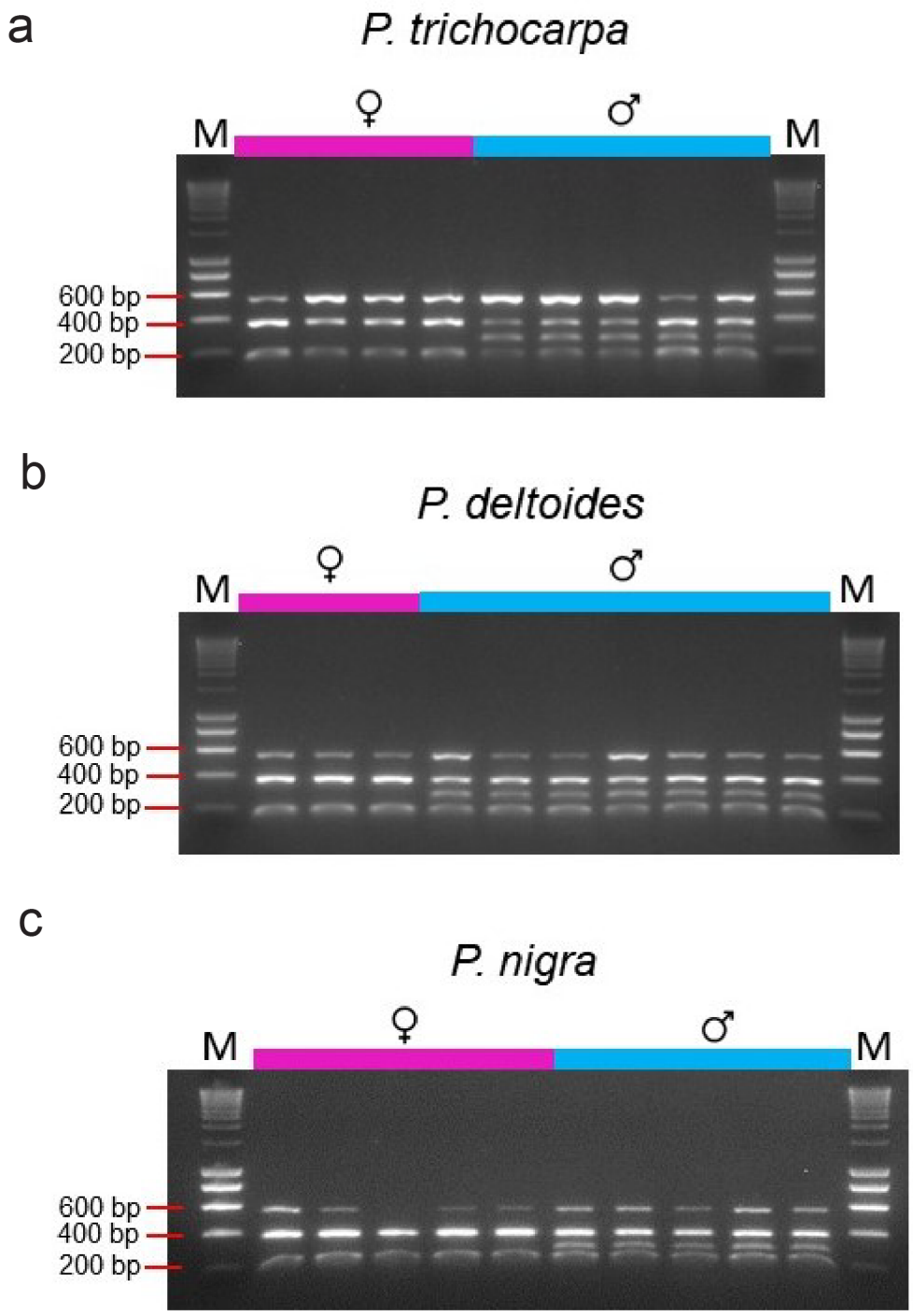


Figure 5. Pcr-rflp assays of (a) *p. Trichocarpa*, (b) *p. Deltoides* and (c) *p. Nigra*. The fragment at 600 bp represents the undigested pcr product (HEMA1_*tsp*RI). Females present pcr products at 400 bp, and 200 bp, while males present y-specific pcr products at 300bp in addition to the ones at 400 bp and 200 bp. Samples were run on a 1.5% agarose gel. M = smart ladder (eurogentech, cologne, germany). Fragment sizes of the dna ladder are given on the left side.

Discussion

Tree species can take a long time before producing flowers for the first time. In the genus *Populus*, time to flowering is often more than a decade. To assess the sex of a

plant at an earlier stage or outside the flowering season, molecular markers can be employed. Such markers have been described for different dioecious species including several poplar species (Geraldes et al., 2015; Pakull et al., 2015). Here, we used the *ARR17* gene, which functions as a sex switch in poplars, to differentiate female and male individuals in *P. alba*, a species with female heterogamety (ZW system). A nearly identical sex-specific coverage pattern of re-sequencing data of *P. alba* and *P. adenopoda* along the *ARR17* genomic region suggests that these two ZW systems of sex determination may have a shared origin. Additionally, the outgroup species of *P. alba* and *P. adenopoda*, *P. qiongdaoensis* exhibits partially shared female-specific coverage along the *ARR17* locus. While we cannot exclude three independent evolutionary origins without more detailed analyses including high-quality genome assemblies, the phylogenetic relationship of the three species is consistent with a single origin early in the evolution of the section Populus (Figure 6). The described ARR17 sex marker may thus work in *P. alba*, *P. adenopoda* and their respective hybrids and potentially also in *P. qiongdaoensis*. In line with this, sex identification using our PCR-based ARR17 marker was successful in the Chinese white poplar (*P. × tomentosa*). Phylogenetic analyses (An et al., 2020; Gao et al., 2019) support the assumption that *P. × tomentosa*, a closely related species of both *P. alba* and *P. adenopoda*, is the result of hybridization between these two species, where *P. adenopoda* is the female parent and *P. alba* the male parent.

The identification of the genetic basis and the molecular mechanisms of sex determination in several species of the genus *Populus* provides insights into sex chromosome evolution. The genus *Populus* is characterized by a diversity of different sex chromosomal systems demonstrating several sex chromosome turnover events. Sex-determining regions are located at different genomic positions including different chromosomes (Geraldes et al., 2015; Kersten et al., 2014; Müller et al., 2020; Yang et al., 2021). Additionally, *P. alba* exhibits a ZW system as opposed to the XY systems of most other species. The analyses of *P. adenopoda* and *P. qiongdaoensis* reported here indicating ZW systems in those species as well, suggest that the ZW system of *P. alba* may not have evolved specifically in the *P. alba* lineage. Shared female-specific hemizygous regions argue for a common origin and would phylogenetically place the evolution of the ZW system to an early ancestor of the section *Populus* (Figure 6) (Wang et al., 2020). While the derived nature of the XY system found in the aspens *P. tremula*

and *P. tremuloides* has been reported before (Müller et al., 2020, Zhou et al., 2020) our results indicate that it re-evolved from the ZW system found in *P. alba*, *P. adenopoda* and *P. qiongdaoensis.* Strikingly, this new XY system is mechanistically identical to the much older system found in several species of the *Populus* sections Tacamahaca, Aigeiros and Turanga, demonstrating a remarkable example of convergent evolution (Figure 6).

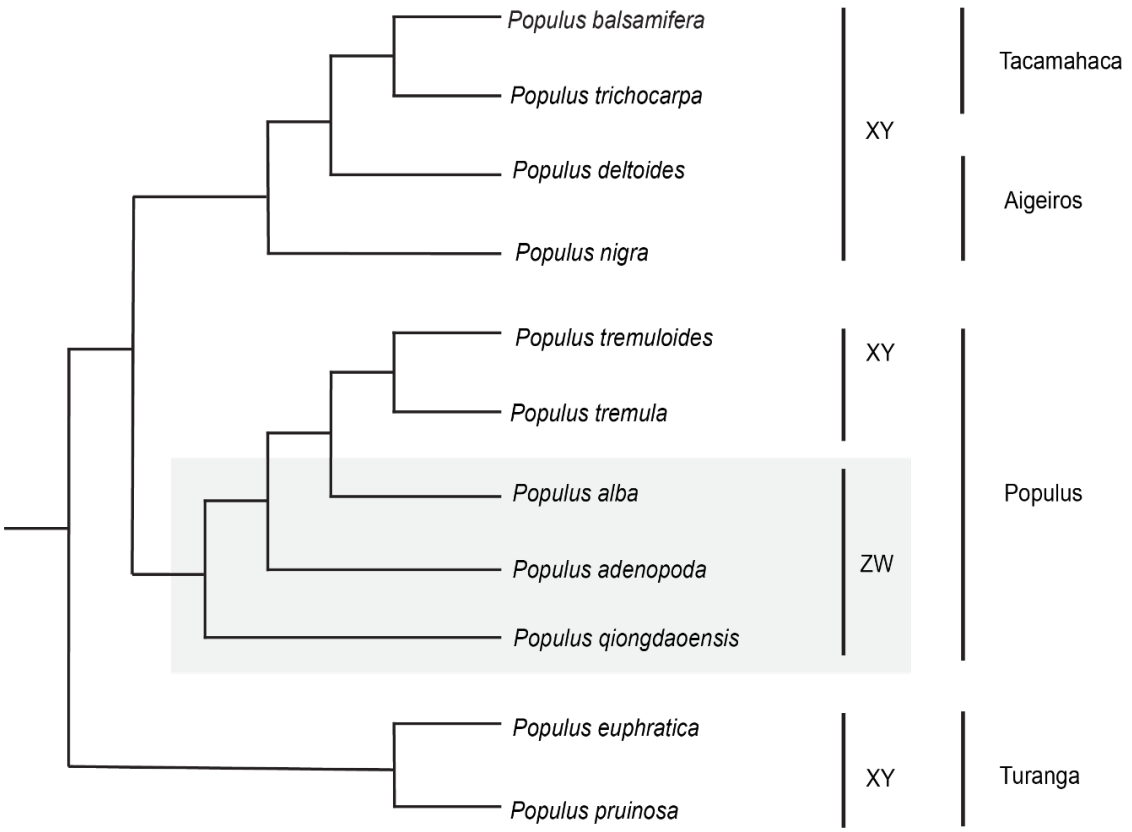


Figure 6. Phylogenetic relationships among *Populus* species based on Wang et al. (2020) with respective sections (Tacamahaca, Aigeiros, Turanga and Populus) and sex determination systems XY or ZW, in the case where *P.alba*, *P. adenopoda* and P. *qiongdaoensis* feature a ZW system.

To summarize all poplar sex markers in one paper (Figure 7), we included results from PCR amplification utilizing previously described sex markers. The male-specific aspen homologue of the *P. trichocarpa* gene Potri.019G047300 (named *TOZ19*, (Pakull et al., 2015)) is located within the SDR of aspens. The *TOZ19* gene is only present in males, while it is missing completely (*P. tremuloides*) or partially (*P. tremula*) in females (Pakull et al., 2015). The paralog Potri.013G079600 (named *TOZ13*) is not sex-linked and can be amplified in all male and female aspen and white poplar species tested. It can therefore serve as a PCR control.

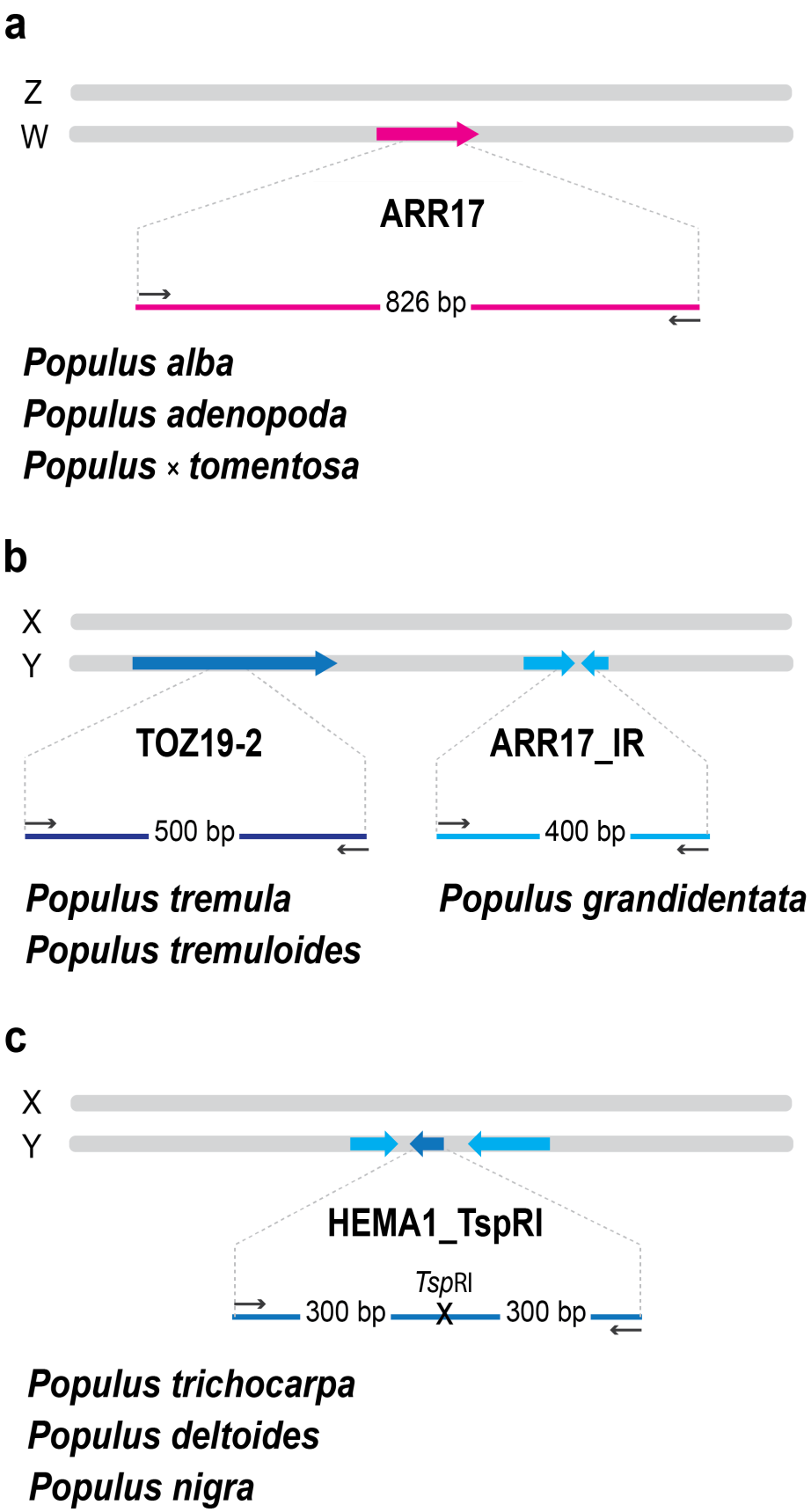


Figure 7. Summary of poplar sex markers. (a) Marker ARR17 in *P. alba*, *P. adenopoda* and *P. × tomentosa*. (b) Marker TOZ19-2 in aspens (*P. tremula* and *P. tremuloides*); ARR17_IR in *P. grandidentata* (c) Marker HEMA1_*Tsp*RI in *P. trichocarpa*, *P. deltoides* and *P. nigra*. In XY systems, ARR17 is present but not sex-linked.

There are few studies involving the bigtooth aspen (*P. grandidentata*) making it difficult to determine its exact phylogenetic relationships although it appears to be an outgroup to the other aspen species (An et al., 2020). However, it seems to behave differently from the other aspens, since PCR amplification of the *TOZ19* gene does not work in this species. As an alternative, the *ARR17* inverted repeat was used as a male-specific marker (ARR17_IR) to differentiate males from females. The results demonstrate a male heterogametic system in *P. grandidentata*. Nevertheless, both the genomic location of the SDR and the mechanism of action remain unknown. It would be interesting to assemble the genome of a male *P. grandidentata* individual to resolve the male-specific sequence of the Y chromosome (MSY) of that species.

Just as the TOZ19 marker that only works for the aspens *P. tremula* and *P. tremuloides*, the *ARR17* inverted repeat (ARR17_IR) marker only works for *P. grandidentata*. The *ARR17* inverted repeat is non-coding. While there must be some sequence conservation for the siRNAs to specifically target the *ARR17* locus for RNA-directed DNA methylation, the siRNAs are rather generic (Müller et al., 2020). They span a relatively wide genomic region of approximately 1,000 bp and mutations within the *ARR17* inverted repeat may thus not affect its overall function. Again, an assembly of the MSY of *P. grandidentata* would be interesting for a comparison between the different aspen species.

In conclusion, our work reveals new details on the evolution of sex determination in the *Populus* genus. This allowed the development of additional molecular markers to determine the sex of non-flowering poplar individuals in different species. Additional sex-chromosomal systems will be likely uncovered in the genus, which is characterized by frequent sex chromosome turnover. It will be interesting to elucidate potential biological reasons for these evolutionary dynamics in future studies.

Acknowledgements

We thank M. Mader for constant support with the Linux cluster, and K. Groppe and A. Eikhof for help in the laboratory. Funding from a grant of the German Research Foundation to N.A.M. (DFG grant no. MU 4357/1-1) and a core grant of the Thünen-

Institute by the Federal Ministry of Food and Agriculture is acknowledged. The funding body played no role in the design of the study or analysis and interpretation of the data.

References

Akagi, T., Henry, I.M., Kawai, T., Comai, L., Tao, R., 2016. Epigenetic regulation of the sex determination gene *MeGI* in polyploid persimmon. Plant Cell 28, 2905–2915. https://doi.org/10.1105/tpc.16.00532

Akagi, T., Henry, I.M., Ohtani, H., Morimoto, T., Beppu, K., Kataoka, I., Tao, R., 2018. A Y-encoded suppressor of feminization arose via lineage-specific duplication of a cytokinin response regulator in kiwifruit. Plant Cell 30, 780–795. https://doi.org/10.1105/tpc.17.00787

Akagi, T., Henry, I.M., Tao, R., Comai, L., 2014. A Y-chromosome–encoded small RNA acts as a sex determinant in persimmons. Science 346, 646–650. https://doi.org/10.1126/science.1257225

Akagi, T., Pilkington, S.M., Varkonyi-Gasic, E., Henry, I.M., Sugano, S.S., Sonoda, M., Firl, A., McNeilage, M.A., Douglas, M.J., Wang, T., Rebstock, R., Voogd, C., Datson, P., Allan, A.C., Beppu, K., Kataoka, I., Tao, R., 2019. Two Y-chromosome-encoded genes determine sex in kiwifruit. Nature Plants 5, 801–809. https://doi.org/10.1038/s41477-019-0489-6

An, X., Gao, K., Chen, Z., Li, J., Yang, Xiong, Yang, Xiaoyu, Zhou, J., Guo, T., Zhao, T., Huang, S., Miao, D., Khan, W.U., Rao, P., Ye, M., Lei, B., Liao, W., Wang, J., Ji, L., Li, Y., Guo, B., Mustafa, N.S., Li, S., Yun, Q., Keller, S.R., Mao, J., Zhang, R., Strauss, S.H., 2020. Hybrid origin of *Populus tomentosa* Carr. identified through genome sequencing and phylogenomic analysis. bioRxiv 38. https://doi.org/10.1101/2020.04.07.030692

Arora, S., Morgan, M., Carlson, M., Pagès H., 2021. GenomeInfoDb: Utilities for manipulating chromosome names, including modifying them to follow a particular naming style. R package version 1.26.7, https://bioconductor.org/packages/GenomeInfoDb

Bhat, B.K., Bindroo, B.B., 1980. Sex chromosomes in *Dioscorea deltoidea* Wall. Cytologia 45, 739–742.

Bolger, A.M., Lohse, M., Usadel, B., 2014. Trimmomatic: a flexible trimmer for Illumina sequence data. Bioinformatics 30, 2114–2120. https://doi.org/10.1093/bioinformatics/btu170

Carey, S., Yu, Q., Harkess, A., 2021. The diversity of plant sex chromosomes highlighted through advances in genome sequencing. Genes 12, 381. https://doi.org/10.3390/genes12030381

Cormier, F., Lawac, F., Maledon, E., Gravillon, M.-C., Nudol, E., Mournet, P., Vignes, H., Chaïr, H., Arnau, G., 2019. A reference high-density genetic map of greater yam (Dioscorea alata L.). Theor Appl Genet 132, 1733–1744. https://doi.org/10.1007/s00122-019-03311-6

Feng, G., Sanderson, B.J., Keefover-Ring, K., Liu, J., Ma, T., Yin, T., Smart, L.B., DiFazio, S.P., Olson, M.S., 2020. Pathways to sex determination in plants: how many roads lead to Rome? Current Opinion in Plant Biology 54, 61–68. https://doi.org/10.1016/j.pbi.2020.01.004

Gao, K., Li, J., Khan, W.U., Zhao, T., Yang, Xiong, Yang, Xiaoyu, Guo, B., An, X., 2019. Comparative genomic and phylogenetic analyses of *Populus* section Leuce using complete chloroplast genome sequences. Tree Genetics & Genomes 15, 32. https://doi.org/10.1007/s11295-019-1342-9

Geraldes, A., Hefer, C.A., Capron, A., Kolosova, N., Martinez-Nuñez, F., Soolanayakanahally, R.Y., Stanton, B., Guy, R.D., Mansfield, S.D., Douglas, C.J., Cronk, Q.C.B., 2015. Recent Y chromosome divergence despite ancient origin of dioecy in poplars (*Populus*). Molecular Ecology 24, 3243–3256. https://doi.org/10.1111/mec.13126

Harkess, A., Huang, K., van der Hulst, R., Tissen, B., Caplan, J.L., Koppula, A., Batish, M., Meyers, B.C., Leebens-Mack, J., 2020. Sex determination by two Y-linked genes in garden asparagus. Plant Cell 32, 1790–1796. https://doi.org/10.1105/tpc.19.00859

Harkess, A., Zhou, J., Xu, C., Bowers, J.E., Van der Hulst, R., Ayyampalayam, S., Mercati, F., Riccardi, P., McKain, M.R., Kakrana, A., Tang, H., Ray, J., Groenendijk, J., Arikit, S., Mathioni, S.M., Nakano, M., Shan, H., Telgmann-Rauber, A., Kanno, A., Yue, Z., Chen, Haixin, Li, W., Chen, Y., Xu, X., Zhang, Y., Luo, S., Chen, Helong, Gao, J., Mao, Z., Pires, J.C., Luo, M., Kudrna, D., Wing, R.A., Meyers, B.C., Yi, K., Kong, H., Lavrijsen, P., Sunseri, F., Falavigna, A., Ye, Y., Leebens-Mack, J.H., Chen, G., 2017. The asparagus genome sheds light on the origin and evolution of a young Y chromosome. Nature Communications 8, 1279. https://doi.org/10.1038/s41467-017-01064-8

He, L., Jia, K.-H., Zhang, R.-G., Wang, Y., Shi, T.-L., Li, Z.-C., Zeng, S.-W., Cai, X.-J., Wagner, N.D., Hörandl, E., Muyle, A., Yang, K., Charlesworth, D., Mao, J.-F., 2021. Chromosome-scale assembly of the

genome of *Salix dunnii* reveals a male-heterogametic sex determination system on chromosome 7. Molecular Ecology Resources. 21, 1966-1982. https://doi.org/10.1111/1755-0998.13362

Kersten, B., Pakull, B., Groppe, K., Lueneburg, J., Fladung, M., 2014. The sex-linked region in *Populus tremuloides* Turesson 141 corresponds to a pericentromeric region of about two million base pairs on *P. trichocarpa* chromosome 19. Plant Biology 16, 411–418. https://doi.org/10.1111/plb.12048

Langmead, B., Salzberg, S.L., 2012. Fast gapped-read alignment with Bowtie 2. Nature Methods 9, 357–359. https://doi.org/10.1038/nmeth.1923

Lawrence, M., Huber, W., Pagès, H., Aboyoun, P., Carlson, M., Gentleman, R., Morgan, M.T., Carey, V.J., 2013. Software for computing and annotating genomic ranges. PLoS Comput Biol 9, e1003118. https://doi.org/10.1371/journal.pcbi.1003118

Leite Montalvão, A.P., Kersten, B., Fladung, M., Müller, N.A., 2021. The diversity and dynamics of sex determination in dioecious plants. Front. Plant Sci. 11, 1-12. https://doi.org/10.3389/fpls.2020.580488

Martin, H., Carpentier, F., Gallina, S., Godé, C., Schmitt, E., Muyle, A., Marais, G.A.B., Touzet, P., 2019. Evolution of young sex chromosomes in two dioecious sister plant species with distinct sex determination systems. Genome Biology and Evolution 11, 350–361. https://doi.org/10.1093/gbe/evz001

McKown, A.D., Klápště, J., Guy, R.D., Soolanayakanahally, R.Y., La Mantia, J., Porth, I., Skyba, O., Unda, F., Douglas, C.J., El-Kassaby, Y.A., Hamelin, R.C., Mansfield, S.D., Cronk, Q.C.B., 2017. Sexual homomorphism in dioecious trees: extensive tests fail to detect sexual dimorphism in *Populus*. Sci Rep 7, 1831. https://doi.org/10.1038/s41598-017-01893-z

Müller, N.A., Kersten, B., Leite Montalvão, A.P., Mähler, N., Bernhardsson, C., Bräutigam, K., Carracedo Lorenzo, Z., Hoenicka, H., Kumar, V., Mader, M., Pakull, B., Robinson, K.M., Sabatti, M., Vettori, C., Ingvarsson, P.K., Cronk, Q., Street, N.R., Fladung, M., 2020. A single gene underlies the dynamic evolution of poplar sex determination. Nature Plants 6, 630–637. https://doi.org/10.1038/s41477-020-0672-9

Pakull, B., Kersten, B., Lüneburg, J., Fladung, M., 2015. A simple PCR-based marker to determine sex in aspen. Plant Biology 17, 256–261. https://doi.org/10.1111/plb.12217

Paolucci, I., Gaudet, M., Jorge, V., Beritognolo, I., Terzoli, S., Kuzminsky, E., Muleo, R., Scarascia Mugnozza, G., Sabatti, M., 2010. Genetic linkage maps of *Populus alba* L. and comparative mapping analysis of sex determination across *Populus* species. Tree Genet. Genomes 6, 863–875. https://doi.org/10.1007/s11295-010-0297-7

R Core Team, 2018. URL https://www.R-project.org (Accessed 26.04.21).

Renner, S.S., Müller, N.A., 2021. Plant sex chromosomes defy evolutionary models of expanding recombination suppression and genetic degeneration. Nature Plants 7, 392–402. https://doi.org/10.1038/s41477-021-00884-3

Robinson, K.M., Delhomme, N., Mähler, N., Schiffthaler, B., Önskog, J., Albrectsen, B.R., Ingvarsson, P.K., Hvidsten, T.R., Jansson, S., Street, N.R., 2014. *Populus tremula* (European aspen) shows no evidence of sexual dimorphism. BMC Plant Biol 14, 276. https://doi.org/10.1186/s12870-014-0276-5

Sanderson, B.J., Feng, G., Hu, N., Carlson, C.H., Smart, L.B., Keefover-Ring, K., Yin, T., Ma, T., Liu, J., DiFazio, S.P., Olson, M.S., 2021. Sex determination through X–Y heterogamety in *Salix nigra*. Heredity 126, 630–639. https://doi.org/10.1038/s41437-020-00397-3

Schiffthaler, B., Bernhardsson, C., Jenkins, J., Jansson, S., Ingvarsson, P., Schmutz, J., Street, N., 2019. An improved genome assembly of the European aspen *Populus tremula*. bioRxiv 1–10.

Shang, H., Hess, J., Pickup, M., Field, D.L., Ingvarsson, P.K., Liu, J., Lexer, C., 2020. Evolution of strong reproductive isolation in plants: broad-scale patterns and lessons from a perennial model group. Philosophical Transactions of the Royal Society B. https://doi.org/10.1098/rstb.2019.0544

Slavov, G.T., Zhelev, P., 2010. Salient biological features, systematics, and genetic variation of *Populus*, in: Jansson, S., Bhalerao, R., Groover, A. (Eds.), Genetics and Genomics of *Populus*, Plant Genetics and Genomics: Crops and Models. Springer, New York, NY, pp. 15–38. https://doi.org/10.1007/978-1-4419-1541-2_2

Tuskan, G.A., DiFazio, S., Jansson, S., Bohlmann, J., Grigoriev, I., Hellsten, U., Putnam, N., Ralph, S., Rombauts, S., Salamov, A., Schein, J., Sterck, L., Aerts, A., Bhalerao, R.R., Bhalerao, R.P., Blaudez, D., Boerjan, W., Brun, A., Brunner, A., Busov, V., Campbell, M., Carlson, J., Chalot, M., Chapman, J., Chen, G.-L., Cooper, D., Coutinho, P.M., Couturier, J., Covert, S., Cronk, Q., Cunningham, R., Davis, J., Degroeve, S., Déjardin, A., dePamphilis, C., Detter, J., Dirks, B., Dubchak, I., Duplessis, S., Ehlting, J., Ellis, B., Gendler, K., Goodstein, D., Gribskov, M., Grimwood, J., Groover, A., Gunter, L., Hamberger, B., Heinze, B., Helariutta, Y., Henrissat, B., Holligan, D., Holt, R., Huang, W., Islam-

Faridi, N., Jones, S., Jones-Rhoades, M., Jorgensen, R., Joshi, C., Kangasjärvi, J., Karlsson, J., Kelleher, C., Kirkpatrick, R., Kirst, M., Kohler, A., Kalluri, U., Larimer, F., Leebens-Mack, J., Leplé, J.-C., Locascio, P., Lou, Y., Lucas, S., Martin, F., Montanini, B., Napoli, C., Nelson, D.R., Nelson, C., Nieminen, K., Nilsson, O., Pereda, V., Peter, G., Philippe, R., Pilate, G., Poliakov, A., Razumovskaya, J., Richardson, P., Rinaldi, C., Ritland, K., Rouzé, P., Ryaboy, D., Schmutz, J., Schrader, J., Segerman, B., Shin, H., Siddiqui, A., Sterky, F., Terry, A., Tsai, C.-J., Uberbacher, E., Unneberg, P., Vahala, J., Wall, K., Wessler, S., Yang, G., Yin, T., Douglas, C., Marra, M., Sandberg, G., Peer, Y.V. de, Rokhsar, D., 2006. The genome of black cottonwood, *Populus trichocarpa* (Torr. & Gray). Science 313, 1596–1604. https://doi.org/10.1126/science.1128691

Wang, M., Zhang, L., Zhang, Z., Li, M., Wang, D., Zhang, X., Xi, Z., Keefover-Ring, K., Smart, L.B., DiFazio, S.P., Olson, M.S., Yin, T., Liu, J., Ma, T., 2020. Phylogenomics of the genus *Populus* reveals extensive interspecific gene flow and balancing selection. New Phytologist 225, 1370–1382. https://doi.org/10.1111/nph.16215

Wang, Y., Huang, J., Wang, Z., Xu, S., Zhang, Z., 2019. The complete chloroplast genome sequence of *Populus pamirica* Kom. (Salicaceae). Mitochondrial DNA Part B 4, 785–786. https://doi.org/10.1080/23802359.2019.1566792

Wickham, H., 2016. Elegant graphics for data analysis. URL https://ggplot2.tidyverse.org/ (accessed 22.04.21).

Wickham, H., Averick, M., Bryan, J., Chang, W., McGowan, L.D., François, R., Grolemund, G., Hayes, A., Henry, L., Hester, J., Kuhn, M., Pedersen, T.L., Miller, E., Bache, S.M., Müller, K., Ooms, J., Robinson, D., Seidel, D.P., Spinu, V., Takahashi, K., Vaughan, D., Wilke, C., Woo, K., Yutani, H., 2019. Welcome to the Tidyverse. Journal of Open Source Software 4, 1686. https://doi.org/10.21105/joss.01686

Yang, W., Wang, D., Li, Y., Zhang, Z., Tong, S., Li, M., Zhang, X., Zhang, L., Ren, L., Ma, X., Zhou, R., Sanderson, B.J., Keefover-Ring, K., Yin, T., Smart, L.B., Liu, J., DiFazio, S.P., Olson, M., Ma, T., 2021. A general model to explain repeated turnovers of sex determination in the Salicaceae. Molecular Biology and Evolution 38, 968–980. https://doi.org/10.1093/molbev/msaa261

Zeileis, A., Grothendieck, G., 2005. zoo: S3 Infrastructure for regular and irregular time series. Journal of Statistical Software 14, 1–27. https://doi.org/10.18637/jss.v014.i06

Zhou, R., Macaya-Sanz, D., Rodgers-Melnick, E., Carlson, C.H., Gouker, F.E., Evans, L.M., Schmutz, J., Jenkins, J.W., Yan, J., Tuskan, G.A., Smart, L.B., DiFazio, S.P., 2018. Characterization of a large sex determination region in *Salix purpurea* L. (Salicaceae). Mol Genet Genomics 293, 1437–1452. https://doi.org/10.1007/s00438-018-1473-y

Zhou, R., Macaya-Sanz, D., Schmutz, J., Jenkins, J.W., Tuskan, G.A., DiFazio, S.P., 2020. Sequencing and analysis of the sex determination region of *Populus trichocarpa*. Genes 11, 843. https://doi.org/10.3390/genes11080843

Chapter 5

ARR17 controls dioecy in *Populus* by repressing B-class MADS-box gene expression

Ana P. Leite Montalvão, Birgit Kersten, Gihwan Kim, Matthias Fladung & Niels A. Müller*

Thünen Institute of Forest Genetics, Sieker Landstraße 2, 22927 Großhansdorf, Germany
* Corresponding author

Accepted in *Philosophical Transactions of the Royal Society B* (in press)

Summary

The number of dioecious species for which the genetic basis of sex determination has been resolved is rapidly increasing. Nevertheless, the molecular mechanisms downstream of the sex determinants remain largely elusive. Here, by RNA-sequencing early-flowering isogenic aspen (*Populus tremula*) lines differing exclusively for the sex switch gene *ARR17*, we show that a narrowly defined genetic network controls differential development of female and male flowers. Although *ARR17* encodes a type-A response regulator supposedly involved in cytokinin hormone signaling, CRISPR-Cas9-mediated *arr17* knockout only affected the expression of a strikingly small number of genes indicating a specific role in the regulation of floral development rather than a generic function in hormone signaling. Notably, the *UNUSUAL FLORAL ORGANS* (*UFO*) gene, encoding an F-box protein acting as a transcriptional co-factor with *LEAFY* (*LFY*) to activate B class MADS-box gene expression, and the B class gene *PISTILLATA* (*PI*), necessary for male floral organ development, were strongly de-repressed in the *arr17* CRISPR mutants. Our data highlight a cytokinin-independent role of the poplar response regulator ARR17 and further emphasize the minimal differences between female and male individuals.

Keywords: cytokinin, dioecy, flower development, poplar, sex determination

Introduction

Poplars are dioecious trees with a genetically controlled system of sex determination (Cronk, 2005). The genomic architecture of sex determination varies between species. The sex-determining regions (SDRs) have different locations and sizes (Geraldes et al., 2015; Kersten et al., 2014; Müller et al., 2020; Yang et al., 2021), and different species even exhibit different heterogametic systems (Müller et al., 2020; Paolucci et al., 2010; Yang et al., 2021). Several studies have characterized the genetic basis of sex determination in poplars (Geraldes et al., 2015; Kersten et al., 2014; McKown et al., 2017; Müller et al., 2020; Yang et al., 2021) including an experimental validation of a single-gene sex switch, named *ARR17*, in early-flowering aspens (*Populus tremula* L.) (Müller et al., 2020). This gene likely underlies sex determination in both XY and ZW systems (Müller et al., 2020;

Yang et al., 2021). Although previous RNA-sequencing studies in different *Populus* species provided valuable insights into sexual development (Chen et al., 2018; Cronk et al., 2020; Sanderson et al., 2019), the molecular function of *ARR17* and the downstream regulatory pathways remained largely elusive. In particular, the possible involvement of cytokinin hormone signaling and the molecular pathways connecting *ARR17* and B class MADS-box gene expression represent open questions.

Cytokinin (CK) is a phytohormone that plays a crucial role in plant growth and development including sexual development, especially the gynoecium (Leuendorf and Schmülling, 2021). Cytokinin is perceived via a two-component system in which signal transduction is achieved by phosphorylation of response regulators (RRs) by histidine kinases (HKs), similar to the two-component systems employed by bacteria to respond to environmental stimuli (Kieber and Schaller, 2018; Stock et al., 2000). The response regulators are particularly interesting as they cause cytokinin-dependent transcriptional reprogramming (Brenner and Schmülling, 2015; Kieber and Schaller, 2018). Since the poplar sex switch gene *ARR17* is homologous to the *ARABIDOSPIS RESPONSE REGULATOR 17*, one of the type-A response regulators which are reported to negatively regulate the CK signaling cascade (To et al., 2004), a connection between *ARR17* and the cytokinin pathway might be expected. Differential sexual development may be controlled by differential hormone signaling.

The separation of the sexes in dioecious species with type II flowers, that is flowers that are unisexual from inception (Diggle et al., 2011), are thought to depend on proper temporal and spatial expression of floral homeotic genes (Cronk and Müller, 2020; Liu and Mara, 2010). According to the ABC model of floral development (Coen and Meyerowitz, 1991), floral organs (i.e. sepals, petals, stamens and carpels) are arranged in four distinct whorls, and within a regulatory network the whorl-specific combination of homeotic gene expression determines floral organ identity. Different genes were found for each class encoding MADS-box transcription factors in *Arabidopsis thaliana* L. (Murai, 2013). The A-class gene *APETALA 1* (*AP1*) is responsible for sepal development. The B class genes *PISTILLATA* (*PI*) and *APETALA 3* (*AP3*) specify the petals and stamens depending on whether they are expressed together with A-class or C-class genes.

Finally, the C-class gene *AGAMOUS* (*AG*) determines carpel development (Coen and Meyerowitz, 1991; Liu and Mara, 2010; Murai, 2013).

The B class genes *PI* and *AP3*, which are essential for stamen development, were highlighted as differentially expressed male-biased genes in different dioecious species, such as the persimmon *Diospyros lotus* (Yang et al., 2019) and the balsam poplar *Populus balsamifera* (Cronk et al., 2020). However, the molecular pathways connecting the sex switch genes *MeGI* in persimmon or *ARR17* in poplar with the floral MADS-box genes remained unclear. In this study, we aimed to specifically characterize the molecular function of the poplar sex switch *ARR17*. To this end, we generated transcriptomic data of isogenic early-flowering male and female aspen lines only differing for a point mutation in the *ARR17* gene. These data allowed us to investigate the molecular mechanisms downstream of *ARR17* without the confounding effects of different genetic backgrounds. We find that, in poplar, *ARR17* functions independently of cytokinin and triggers female development by repressing the *UFO-PI* cascade suggesting a direct role of *ARR17* on the specification of floral organ identity.

Material and methods

Plant material, growth, and sampling

The plant material (flower buds) was obtained from one female early-flowering line (T222-3), which expresses the *Arabidopsis thaliana FLOWERING LOCUS T* (*FT*) gene under the control of the heat-inducible promoter derived from the soybean gene *hs6871* (Schöffl et al., 1984) encoding a heat shock protein (HSP) , and three independent T222-3-based isogenic *arr17* CRISPR mutants (N500-1, N500-3 and N500-5) previously described (Müller et al., 2020). Each of these lines contains a CRISPR-induced mutation disrupting the open reading frame of *ARR17*. *In vitro*-grown plants were transferred to soil and cultivated under 16/8 h light/dark and 22/17 °C temperature cycles for 1.5 months. To induce *FT* expression and the consequential development of generative buds, a heat shock treatment was applied for 2 h at 40 °C every day for one month. The plants were randomized and watered daily. The experiments were conducted in two batches, under the same conditions. For the first batch, the flower buds were sampled every five days

after the start of the heat shock treatment until fully formed flowers were observed. Samples from days 5, 10, 15, and 20 were used for RNA-sequencing. For each time point, three biological replicates were collected for each sex (female: 3x T222-3, *arr17* CRISPR: 1x N500-1, 1x N500-3 and 1x N500-5). Each replicate consisted of flower buds pooled from three plants. For the second batch was prepared the same way, however, only samples from day 20 were used for RNA-sequencing. For *P. alba*, flower buds from a female (clone Jap1) and a male (clone Monrepos) field-grown tree (three samples per tree) were collected at a single time point on 22nd July 2020. All flower buds were snap frozen in liquid nitrogen and stored at -70 °C until RNA extraction.

RNA extraction, cDNA synthesis and qRT-PCR

The frozen flower buds were ground in a Retsch mill (Retsch GmbH, Germany) at 25 Hz for 30 seconds to a fine powder that was used for RNA extraction. Total RNA was extracted with the Spectrum Plant Total RNA kit (Sigma-Aldrich, USA) according to the manufacturer's manual, Protocol A. Following that, DNase I digestion was performed using the Turbo DNA-free kit (Invitrogen, USA). The RNA concentration and purity were assessed using a Nanodrop 1000 spectrophotometer (Peqlab Biotechnologie GmbH, Germany) and by native agarose gels. The RNA Integrity Number (RIN) was determined using the plant specific protocol of Agilent Bioanalyzer (Agilent Technologies, Inc., USA). All samples presented RIN > 7. For cDNA synthesis, 2 µg of RNA, Oligo (dT) primers and SuperScript IV reverse transcriptase (Invitrogen, USA) were used following the manufacturer's protocol, using 10 µl reactions without RNaseOUT. Reverse transcriptase quantitative PCR (qRT-PCR) was carried out in duplicates on a CFX96 Touch Real Time PCR Detection System (Bio-Rad Laboratories GmbH, USA) using the SsoAdvanced Universal SYBR Green Supermix (Bio-Rad Industries, Inc., USA) and a two-step PCR program with annealing temperature of 60 °C. Relative expression levels were calculated using the $2^{-\Delta\Delta Ct}$ method (Schmittgen and Livak, 2008). The primers are given in Supplementary Table 1.

RNA-sequencing and data analysis

Strand specific RNA-seq libraries were generated by Novogene (Novogene (UK) Company Ltd., Cambridge, UK) and sequenced using the Illumina HiSeq platform. Paired end 150 bp reads to a target depth of 30 million paired end reads per sample were produced. The filtering of sequenced reads consisted of removing reads containing adapters, reads containing undetermined bases (N >10%) and low-quality reads (*Qscore* ≤ 5). The quality of the raw reads was assessed using FASTQC. The reads were mapped to the *P. tremula* v2.2 genome (Schiffthaler et al., 2019) using STAR aligner (version 2.7.1a) with default settings and with the annotation gene file and subsequently used to calculate read counts with the R package Rsubread (Liao et al., 2019) and the command *featurecounts*. Differential expression (DE) analyses between the lines was performed in R version 4.0.4 using the DESeq2 package (version 1.30.1) (Love et al., 2014). The raw data set was filtered by removing genes for which the sum of reads for all samples was below 10. From the second experiment, an outlier sample was removed from further analysis, since no reads of *ARR17* were detected (Supplementary Figure 1). The remaining samples and genes were used for the differential expression analysis using *DESeq* function (design = ~batch + sex). The adjusted p-value (padj < 0.05) and an absolute Log2FoldChange > 1.5 were used to assess significance and identify differentially expressed genes. The variance-stabilizing *rlog* was used and the counts were normalized using DESeq2's own normalization method for exploratory analyses such as Principal Component Analysis (PCA). Batch effects were removed using the *removeBatchEffect* function from the R package limma (version 3.46.0) (Ritchie et al., 2015). The raw read counts are given in Supplementary Table 2.

Gene set enrichment analysis based on gene ontology (GO)

A GO term enrichment analysis was performed utilizing the topGO package in R (version 2.42.0) (Alexa and Rahnenfuhrer, 2020) with default settings as well as the optional function *nodeSize* = 10, which removes terms with less than 10 annotated genes. We considered 24,464 genes (out of the 29,549 expressed genes used for the DE analyses) that had a GO annotation for *P. tremula* (Schiffthaler et al., 2019). The analysis was performed with differentially expressed genes at our standard cut-off (p<0.05 &

|log2FC|>1.5) and using a relaxed significance level of p<0.1 to avoid false negatives and to control for random effects due to threshold choice (Pan et al., 2005).

Cytokinin treatment

Female and male heat-inducible early-flowering aspen lines were grown for one month in tissue culture and subsequently transferred to Magenta plant incubation boxes (Sigma-Aldrich, USA) containing woody plant medium (WPM) with and without 6-benzylaminopurine (BAP), which is a synthetic cytokinin that promotes growth and is involved in various developmental processes such as cell division, shoot formation and promotion of flowering. Different concentrations of BAP were used: 110 µM, 220 µM, 440 µM and 880 µM, in four individuals per sex, totalling 32 treated plants. Moreover, six plants of each sex were used as a control (without BAP). Following a three-week incubation in a climate room at 21 °C and constant light, the cultures were placed under a daily heat treatment (2h at 40°C) to trigger flowering as described above. Flowers were analysed after 45 days, and the numbers were recorded.

Data availability

The RNA-sequencing data were deposited in NCBI's SRA under the bioproject accession number PRJNA773612.

Results and discussion

Expression of *ARR17* in poplar is tissue-specific, occurring only in female flower buds (Müller et al., 2020; Ramírez-Carvajal et al., 2008). Thus, to identify the downstream genes controlled by *ARR17* we analyzed transcriptome datasets of female and male flower buds collected at different times of development. For reliable sample collection, we took advantage of an artificial flower induction system, which utilizes the *Arabidopsis thaliana FLOWERING LOCUS T* (*FT*) gene under the control of a soybean heat shock protein (HSP) promoter. This system enables heat-inducible early-flowering and allows poplar flower development, which in nature takes almost one year and is inconsistent between individuals and years (Boes and Strauss, 1994; Hoenicka et al.,

2016), to be compressed into one month by heat shock-mediated induction of *FT* expression (Hoenicka et al., 2016; Müller et al., 2020). During this month, we repeatedly sampled the developing flower buds of female and male early-flowering aspen (*Populus tremula*) lines. Importantly, these female and male lines (henceforth referred to as female and *arr17* CRISPR) are genetically identical, except for a CRISPR-Cas9-inducued *arr17* mutation. This single induced mutation, which disrupts the open reading frame of *ARR17*, changes females to males (Müller et al., 2020). These isogenic lines provide a unique possibility to study the genetic networks downstream of the sex-determining gene without any confounding effects from different genetic backgrounds which usually complicate comparisons between female and male individuals.

To assess differentially expressed genes downstream of *ARR17*, samples from days 5, 10, 15 and 20 after the start of flower induction were employed for RNA-sequencing, with three replicates per sex and day. Initially, we analyzed the general patterns of transcriptome variation with a principal component analysis (PCA) (Figure 1a). The PCA indicated rapid and dynamic changes in the transcriptome consistent with the transitioning from vegetative to generative development. PC1 and PC2 represented 71% and 11% of the total variance, respectively, and most of the variation in the transcriptome can be accounted for by the sampling day. Nevertheless, a marked separation between female and *arr17* CRISPR lines occurred at day 20 (Figure 1a) suggesting that this may be the earliest stage of sexual differentiation. In line with this, day 20 was the first time-point with robust *ARR17* expression in a qRT-PCR expression time-course (Figure 1b). *ARR17* expression was present in a narrow temporal window during the developmental trajectory from vegetative buds to fully developed flowers. These results highlight that *ARR17* expression is not only tissue-specific but also time-specific. Considering that *ARR17* represents a single-gene sex switch, differences between females and males could not be less.

While *ARR17* expression is female-specific in the natural sex-determining systems (Cronk et al., 2020; Müller et al., 2020), we see expression at day 20 also in our male *arr17* CRISPR lines (Figure 1b). This can, however, be explained by the fact that the *arr17* CRISPR lines present a male phenotype due to a loss of function mutation of the *ARR17* gene at the protein level and not due to transcriptional silencing or gene absence. The peak of *ARR17*

expression corresponds to an early stage of flower development (Supplementary Figure 2a). Notably, no sex-specific differences in flower bud morphologically or any other floral structures could be observed between female and *arr17* CRISPR lines at time-point 20 (Supplementary Figure 2b). A study in *P. balsamifera* demonstrated that in nature the highest expression of *ARR17* also occurs at the earliest stages of reproductive development (Cronk et al., 2020). Together these data suggest that *ARR17* may determine sex early on during flower development.

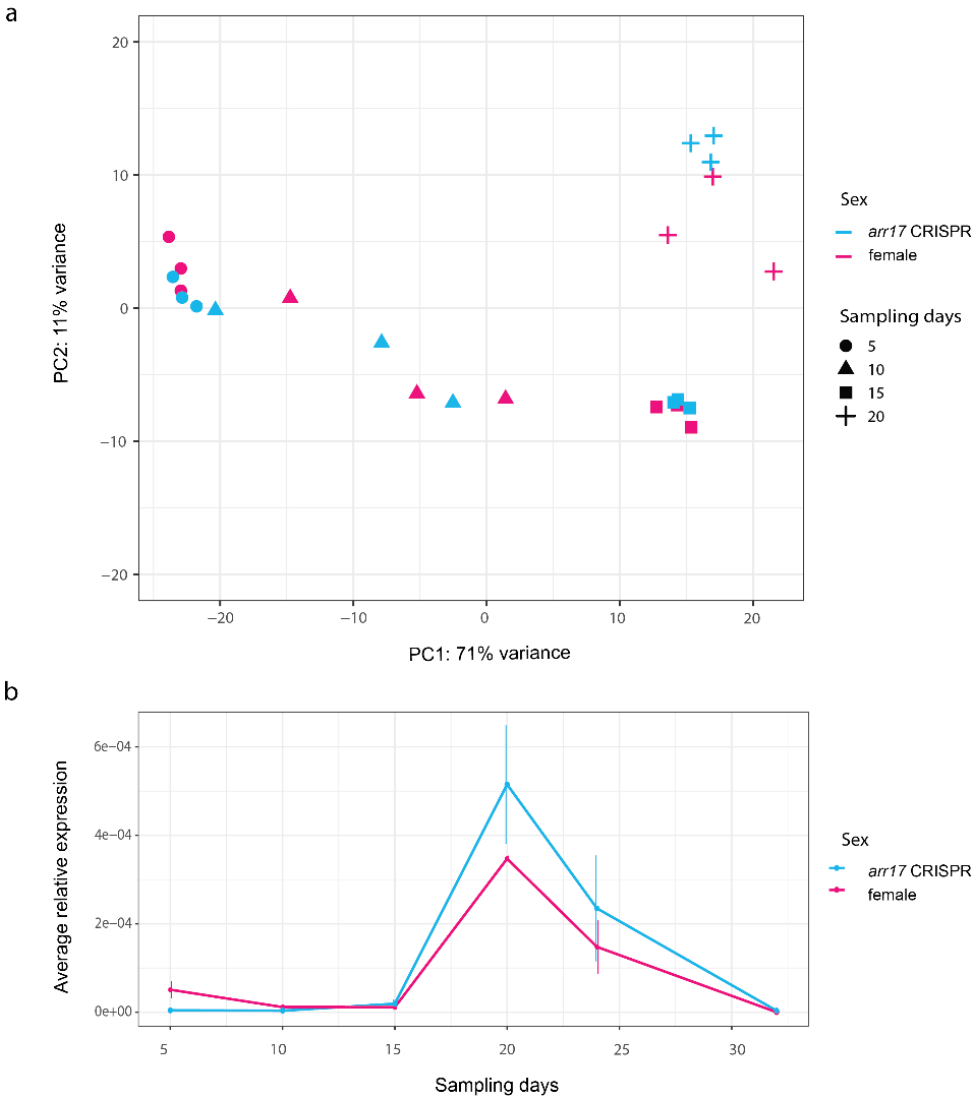


Figure 1. Transcriptome variation and *ARR17* expression in female and male flower buds of early-flowering aspen lines. (a) Principal component analysis (PCA) of transcriptome variation of early-flowering female (magenta) and male *arr17* CRISPR (blue) aspen (*P. tremula*) lines. Different symbols indicate different sampling days (i.e. days 5, 10, 15 and 20 after the start of flower induction). The first principal component PC1 explains 71% and the second principal component PC2 11% of the total variance. **(b)** *ARR17* expression occurs in a narrow temporal window during poplar flower development. Average relative expression (n=3) of *ARR17* expression determined via qRT-PCR over a developmental time-course in the same two genotypes shown in (a). *ARR17* expression peaks at day 20. Error bars indicate the standard error of the mean (SEM).

Since our developmental RNA-seq and *ARR17* qRT-PCR time-courses demonstrated that the earliest substantial sex-specific differences are expected for day 20, we focused further differential gene expression analyses on that time-point. It must be noted, however, that three biological replicates limit the statistical power to identify sex-specific differences. We only found two significantly ($p<0.05$, $|log2FC|>1.5$) differentially expressed genes (DEGs), i.e. Potra2n2c5701 (*PI*) and Potra2n4c8755 (*BGLU13*) (Supplementary Figure 3). We therefore generated a second identical RNA-seq dataset for day 20. The combination of replicates from both experiments should provide sufficient statistical power and allow the reliable identification of differentially expressed genes. A total of 29,549 expressed genes were analyzed for differential expression (Supplementary Table 3). To get a first overview of the biological processes that may be involved in sex determination, we performed a gene ontology (GO) term enrichment analysis. For this analysis we selected DEGs based on different significance thresholds (using a relaxed setting: $p<0.1$, and a more stringent setting: $p<0.05$ & $|log2FC|>1.5$), since the threshold choice can have a relevant effect on the results (Pan et al., 2005). With both settings we identified the biological processes "positive regulation of transcription by polymerase II" (GO:0045944) and "maintenance of meristem identity" (GO:0010074) among the top five categories (Supplementary Tables 4 and 5). Strikingly, almost 90% (15/17) of the DEGs ($p<0.1$) involved in positive regulation of transcription were MIKC-type MADS-box genes, which play prominent roles in the control of reproductive development (Gramzow and Theissen, 2010; Kaufmann et al., 2005). Particularly all *APETALA 3* (*AP3*) and *PISTILLATA* (*PI*) paralogs, which are essential for stamen development, were up-regulated in the *arr17* CRISPR mutants (Supplementary Table 5). The three identified meristem identity genes included both *UNUSUAL FLORAL ORGANS* (*UFO*) paralogs, which encode F-box proteins acting as transcriptional co-factor with *LEAFY* (*LFY*) to activate B class MADS-box gene expression and are reported to provide the spatial cues for the expression of *AP3* and *PI* (Honma and Goto, 2000; Ng and Yanofsky, 2001). No category related to the cytokinin signaling pathway, such as "response to cytokinin" (GO:0009735), was enriched. These results argue against a function of *ARR17* in modulating cytokinin signaling to control sex determination in poplar but rather highlight the importance of *ARR17* in repressing MIKC-type MADS-box transcription factors to specify floral organ identity.

The differential gene expression analysis of the combined dataset with a standard significance cut-off (i.e. p<0.05 & |log2FC|>1.5) resulted in a strikingly small set of 33 DEGs (Table 1, Figure 2), indicating minimal changes in the transcriptome upon *arr17* knockout. Among those 33 DEGs, 13 are upregulated in the *arr17* CRISPR mutants. Two of these genes stand out compared to all others: *PISTILLATA* (Potra2n2c5701) required for stamen development (Theißen and Saedler, 2001) and *UFO* (Potra2n1c1412), which activates B class MADS-box gene expression (Chae et al., 2008; Lee et al., 1997; Parcy et al., 1998; Souer et al., 2008). Both these genes are strongly upregulated in the *arr17* CRISPR mutants (Figure 2). It should be noted that their paralogs are also differentially expressed (Table 1, rows 6 and 7). We were wondering whether the same genes may be differentially expressed in poplar species with independently evolved systems of sex determination. For example, *Populus alba* features a ZW system of sex determination in which *ARR17* is located in the female-specific region of the W chromosome (Müller et al., 2020; Yang et al., 2021). *Populus balsamifera* exhibits an XY system similar to the one found in the aspens but with a different genomic architecture and an independent evolutionary origin (Geraldes et al., 2015; Müller et al., 2020; Zhou et al., 2020). For *P. balsamifera*, 854 DEGs in early developing female and male floral buds (July 2017) have been reported before (Cronk et al., 2020). For *P. alba*, we generated RNA-seq data to assess differential gene expression (p<0.05 & |log2FC|>1.5) in female and male samples collected at an early stage of reproductive development (July 2020) as well. These data identified a total of 1,725 DEGs (Supplementary Table 6). All three datasets, which are not expected to share any gene by chance, shared exactly two differentially expressed genes representing the two *PI* paralogs (Supplementary Figure 4, Supplementary Table 7). *UFO* was not assessed in *P. balsamifera* but was shared between *P. tremula* and *P. alba* (Supplementary Table 7). These additional data further highlight the prominent role of *PI* and *UFO* and suggest that the molecular mechanism of sex determination may be shared between species with independently evolved SDRs.

Table 1. Differentially expressed genes (DEGs) at day 20. For each gene, the Log2FoldChange (female vs. *arr17* CRISPR), the adjusted p-values, the respective *P. trichocarpa* and *A. thaliana* gene identifiers and as well as the *A. thaliana* synonym are given.

Potra_v2.2_ID	Log2FoldChange	P adjusted	Potri_v3.0_ID	*A. thaliana* ID	*A. thaliana* synonym
Potra2n2c5701	-3.56254	1.99E-24	Potri.002G079000	AT5G20240	Pistillata (PI)
Potra2n1c1412	-1.61664	1.01E-20	Potri.001G160900	AT1G30950	Unusual floral organs (UFO)
Potra2n10c20292	-2.04511	3.17E-11	Potri.010G236300	AT3G21510	Histidine-containing phosphotransmiter 1 (AHP1)
Potra2n2c4152	-2.09616	7.15E-10	Potri.002G250000	AT3G25400	dCTP pyrophosphatase
Potra2n4c8755	1.647911	7.41E-07	Potri.004G040700	AT5G44640	Beta glucosidase 13 (BLU13)
Potra2n3c7869	-2.35103	8.11E-06	Potri.003G074100	AT1G30950	Unusual floral organs (UFO)
Potra2n5c11227	-2.11149	3.66E-05	Potri.005G182200	AT5G20240	Pistillata (PI)
Potra2n16c29771	1.947375	4.83E-05	Potri.016G058500	AT4G38180	Far1-related sequence (FRS5)
Potra2n2c5611	-1.77295	6.07E-05	Potri.002G088200	AT1G37140	MEI2- C terminal RRM only like 1
Potra2n18c32797	2.520892	8.14E-05	Potri.018G053600	AT5G56860	GATA transcription factor 21 (GATA21/GNC)
Potra2n9c19634	-1.79274	0.000513	Potri.009G055700	AT5G13790	Agamous like 15 (AGL15)
Potra2n10c21177	2.17995	0.000764	Potri.010G141000	AT5G49330	MYB domain protein 111 (MYB111)
Potra2n299s35250	-1.8207	0.002193	Potri.008G131100	AT1G70890	Major latex protein like 43 (MLP43)
Potra2n1c1097	1.68173	0.002836	Potri.003G106800	AT5G51330	Switch 1 (SWI1)
Potra2n6c13866	1.523646	0.003592	Potri.006G165900	AT4G30190	Plasma membrane protein ATPase 2 (PMA2)
Potra2n15c28326	1.720965	0.006144	Potri.005G036600	AT1G54820	Protein Kinase superfamily protein
Potra2n5c12526	1.589281	0.006144	Potri.015G095900	AT5G50400	Purple acid phosphatase 27 (PAP27)
Potra2n2c5221	1.687623	0.007187	Potri.014G038500	-	
Potra2n11c23459	1.933203	0.010079	Potri.011G031800	AT3G25820	Terpene synthase-related protein (TPS-CIN)
Potra2n5c12753	1.945979	0.011358	Potri.005G014900	AT4G21390	S-locus lectin protein kinase family protein (B120)
Potra2n6c13588	1.872412	0.014159	Potri.006G199300	AT1G68450	Pigment defective 337 (PDE337)
Potra2n5c12584	1.531046	0.015329	Potri.005G028200	AT3G26040	HXXXD-type acyl transferase
Potra2n2c4059	-1.71657	0.016211	Potri.014G195800	AT5G44070	Phytochelatin synthase 1 (PCS1)
Potra2n6c14378	1.534229	0.016528	Potri.006G107700	AT2G30400	Ovate family protein 2 (OFP2)
Potra2n10c20471	1.605425	0.016988	Potri.008G043900	AT1G07900	LOB domain-containing protein 1
Potra2n12c23975	-1.86887	0.020274	Potri.012G032300	AT5G15290	Domain of unknown function (DUF588)
Potra2n13c25563	-1.95196	0.021018	Potri.013G084400	AT3G26120	Terminal ear1-like (TEL1)
Potra2n432s35661	1.757162	0.021045	Potri.001G015400	AT3G45140	Lipoxygenase 2 (LOX2)
Potra2n1c775	1.583641	0.021279	Potri.003G138400	AT5G42800	Dihydroflavonol 4 reductase (DFR4)
Potra2n6c15208	1.986666	0.026446	Potri.006G019800	-	
Potra2n3c7698	2.028289	0.029643	Potri.003G091200	AT4G17810	Zinc finger protein 1 (ZP1)
Potra2n18c32253	1.629004	0.04113	Potri.018G113300	AT4G02050	Sugar transporter protein 7 (STP7)
Potra2n14c27869	-1.54097	0.041274	Potri.014G179400	AT1G32450	Nitrate transporter 1.5 (NRT1.5)

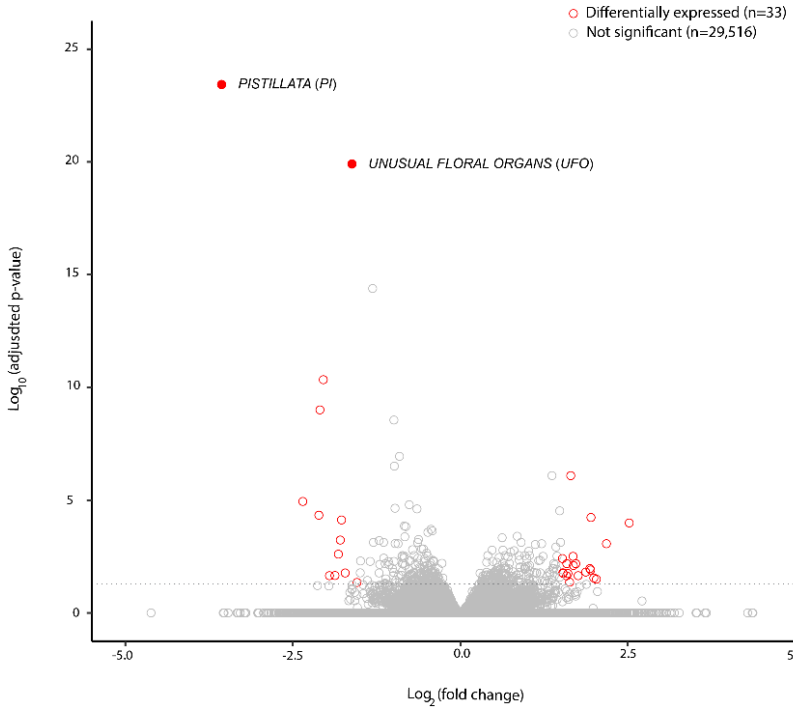


Figure 2. *PISTILLATA* (*PI*) and *UNUSUAL FLORAL ORGANS* (*UFO*) are strongly upregulated in *arr17* CRISPR mutants on day 20. Volcano plot showing 29,549 expressed genes. Significantly differentially expressed genes (female vs. *arr17* CRISPR lines) at day 20 (padj < 0.05 and abs(Log2FoldChange > 1.5)) are depicted in red. *PI* and *UFO* are highlighted by filled symbols. The dashed grey line indicates the p-value significance threshold.

Interestingly, key genes from the cytokinin signaling pathway such as the type-B response regulators *ARR1*, *ARR10*, and *ARR12* are not differentially expressed indicating that *ARR17* may not be involved in cytokinin signaling (Supplementary Figure 5). This is in line with the GO term enrichment analyses, which also failed to detect any connection of *arr17* mutation with cytokinin signaling. Cytokinin-independent roles of type-A response regulators have been described before. For instance, in *A. thaliana* the type-A response regulators *ARR3* and *ARR4* play cytokinin-independent roles in the circadian clock (Salomé et al., 2006). To further explore the potential role of cytokinin signaling in poplar sex determination, we adapted our early-flowering system to generate poplar plants flowering *in vitro* with a height of only 10 cm. This system allowed us to test the effect of exogenous application of synthetic cytokinin into the growth medium on flower development. In particular, we wanted to assess whether treatment with 6-benzylaminopurine (BAP) may have any effect on sexual development. While we observed an increase in the total number of flowers (Supplementary Figure 6), there was no effect on flower sex. This is in contrast with other dioecious species where an exogenous application of cytokinin in male flowers stimulates the induction of carpel development converting them to hermaphrodites (Akagi et al., 2018; Wang et al., 2013; Yonemori et al., 1993).

In summary, our results suggest a specific function of the poplar sex switch gene *ARR17* on floral organ identity rather than a generic function in the cytokinin signaling pathway. The poplar *ARR17* gene is orthologous to the *A. thaliana* gene pair *ARR16*/*ARR17*. There is no one-to-one orthology. In *A. thaliana*, overexpression of the *ARR16* and *ARR17* genes slightly affects flowering time but does not appear to change floral organ identity (Ren et al., 2009). Additionally, the *ARR16*/*ARR17* gene pair appears to be specifically involved in regulating cell divisions of the stomatal lineage (Vatén et al., 2018). In contrast, our results in poplar highlight *UFO* and *PI* as major downstream factors and thus the regulation of floral organ identity as the key function of the sex determinant *ARR17*. In the absence of *ARR17* activity, expression of *UFO* is ensured and as a transcriptional co-factor with *LEAFY* (*LFY*) it activates B class MADS-box genes (Chae et al., 2008; Zhao et al., 2001). On the other hand, in females *ARR17* prevents the expression of *UFO* and therefore represses male development (Figure 3).

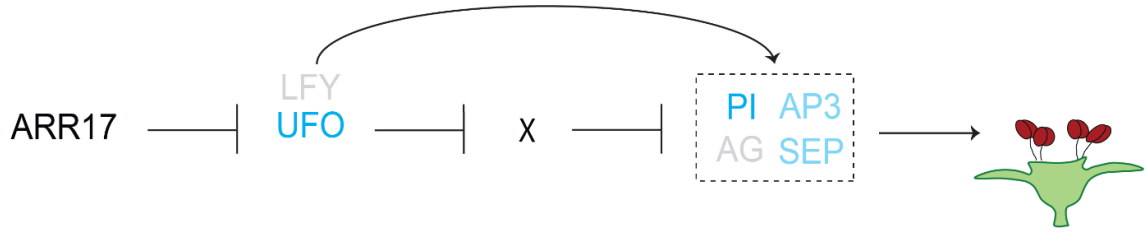


Figure 3. Potential downstream pathway of the sex switch ARR17. For the development of male flower organs (i.e. stamens), the B class MADS-box genes *PISTILLATA* and *APETALA 3* are essential as they form a heterodimer (dashed square) with C-class and E-class genes *AGAMOUS* and *SEPALLATA*, respectively. These genes are regulated transcriptionally by genes such as *LEAFY* and the co-factor *UFO* (as part of the SCF complex), which potentially represses a factor (depicted as "X") via degradation (Zhao et al., 2001) that would repress B class gene expression.

Remaining open questions concern the mechanism by which *ARR17* prevents *UFO* expression in females, the linearity of the pathway and the presence of possible additional factors on the male Y chromosome that might contribute to differential sexual development in nature. *ARR17* is a single-domain response regulator because it contains only a receiver domain (Jenal and Galperin, 2009). The C-terminal extension in *ARR17* of *P. tremula* is short (only 9 amino acids in Potra000483g02981.1; Supplementary Figure 7) similar to *ARR16* and *ARR17* in *A. thaliana* (D'Agostino et al., 2000). The absence of any kind of effector domain at the C-terminus in *ARR17* argues against direct transcriptional control of *UFO* by *ARR17*. Single-domain response regulators may rely on protein-protein interactions to exert their downstream biological effects, after phosphorylation by a histidine kinase and conformational change of the receiver domain (Sarkar et al., 2010). Thus, one mode of action could be that *ARR17* interacts at the protein level with a transcriptional regulator of *UFO*. Regarding the linearity of the pathway, *ARR17* could be involved in the repression of *UFO* and *PI* only, or it could fulfil additional essential functions to determine the sex of poplars. This question should be addressed in future experiments by knocking out the poplar genes *UFO* and *PI*. In case of a linear pathway, *ufo* and *pi* mutations should convert males to females.

Conclusions

RNA-sequencing of developing flower buds of early-flowering isogenic female and male aspen lines only differing for a CRISPR-induced mutation in the sex determinant *ARR17* identified differentially expressed genes likely involved in poplar sex determination.

During poplar development, *ARR17* is only expressed in floral buds and only in a narrow temporal window during flower bud development. The difference between females and males could hardly be smaller, which is in line with sexual homomorphism reported in different poplar species (McKown et al., 2017; Renner and Müller, 2021; Robinson et al., 2014). Despite being a type-A response regulator, *ARR17* does not appear to control cytokinin signaling. Instead, *UFO* and the B class MADS-box gene *PI* were highlighted by several analyses as key components of the gene network downstream of *ARR17* (Figure 3) indicating a highly targeted role of *ARR17* in specifying floral organ identity. It will be exciting to further explore the proposed pathway and generate *ufo* and *pi* knockouts to test whether the modulation of additional signaling cascades is essential or whether the repression of *UFO* and *PI* alone are sufficient to specify differential sex expression.

Acknowledgments

We would like to thank Annika Eikhof for technical assistance in the laboratory and Malte Mader for technical support with the Linux cluster. We would like to acknowledge a grant from the Deutsche Forschungsgemeinschaft (DFG grant number: MU 4357/1-1) to Niels A. Müller. Finally, we want to thank Susanne S. Renner and the Royal Society editorial office for making this theme issue possible.

Supplementary information

The Supplementary Tables 1 to 7 in excel format for this article can be found in a temporary link (until publication):

https://drive.google.com/drive/folders/1v5xFO5PXC4viOtTf1qz6g4JGbEXFW7tb?usp=sharing

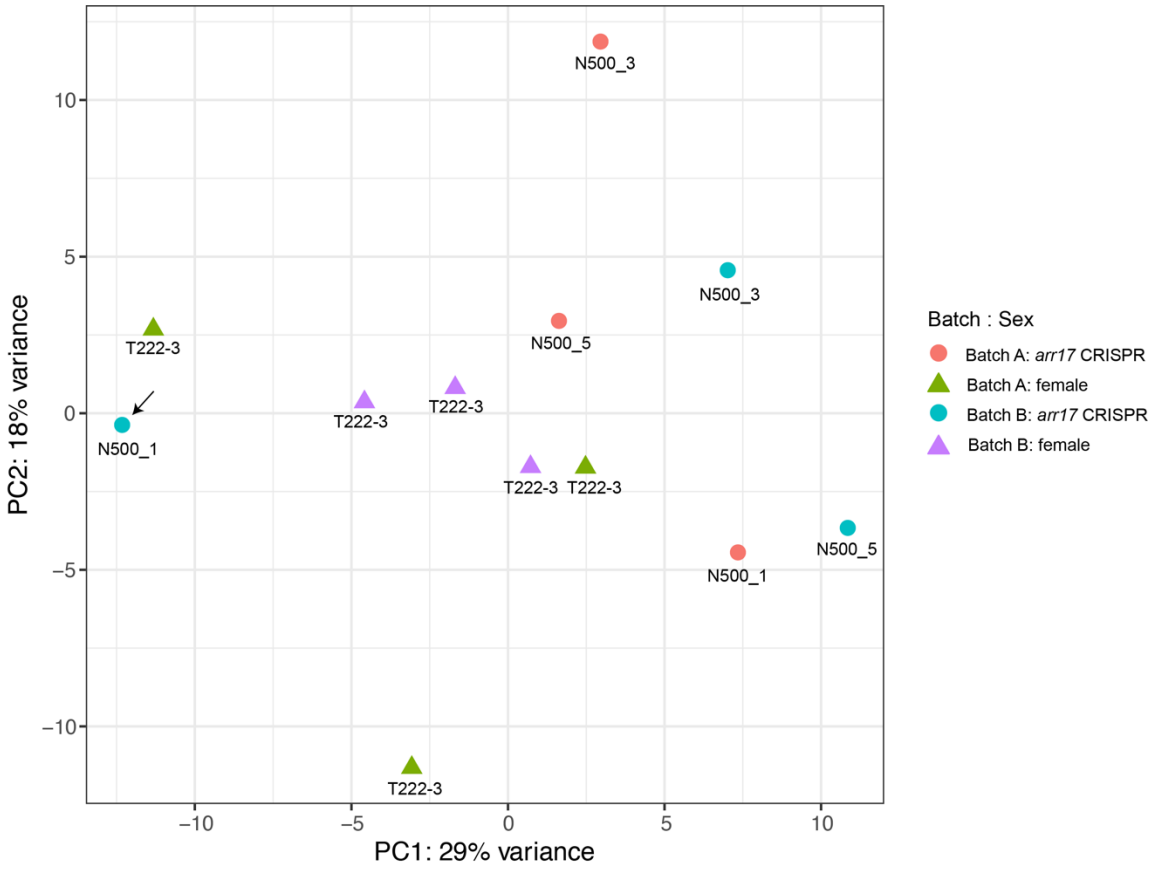


Figure S1. Principal Component analysis of day 20 with RNA-seq data from two batches (i.e. Batch A and Batch B). The black arrow marks the outlier sample which was excluded from the differential expression analysis because it did not present any reads of *ARR17*, and therefore was in a different developmental stage compared to all other samples.

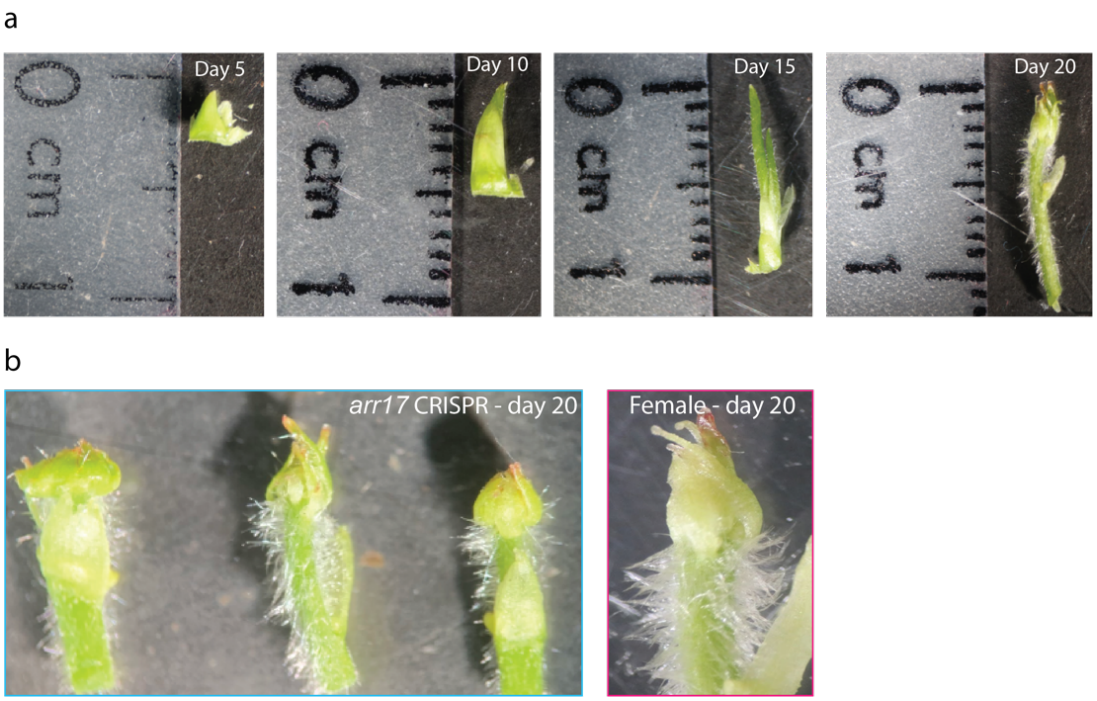


Figure S2. (a) Morphological observation of flower buds collected for RNA-seq for days 5, 10, 15 and 20. **(b)** No morphological differences at day 20 between flowers from *arr17* CRISPR mutants (blue contour) and a female flower (pink contour).

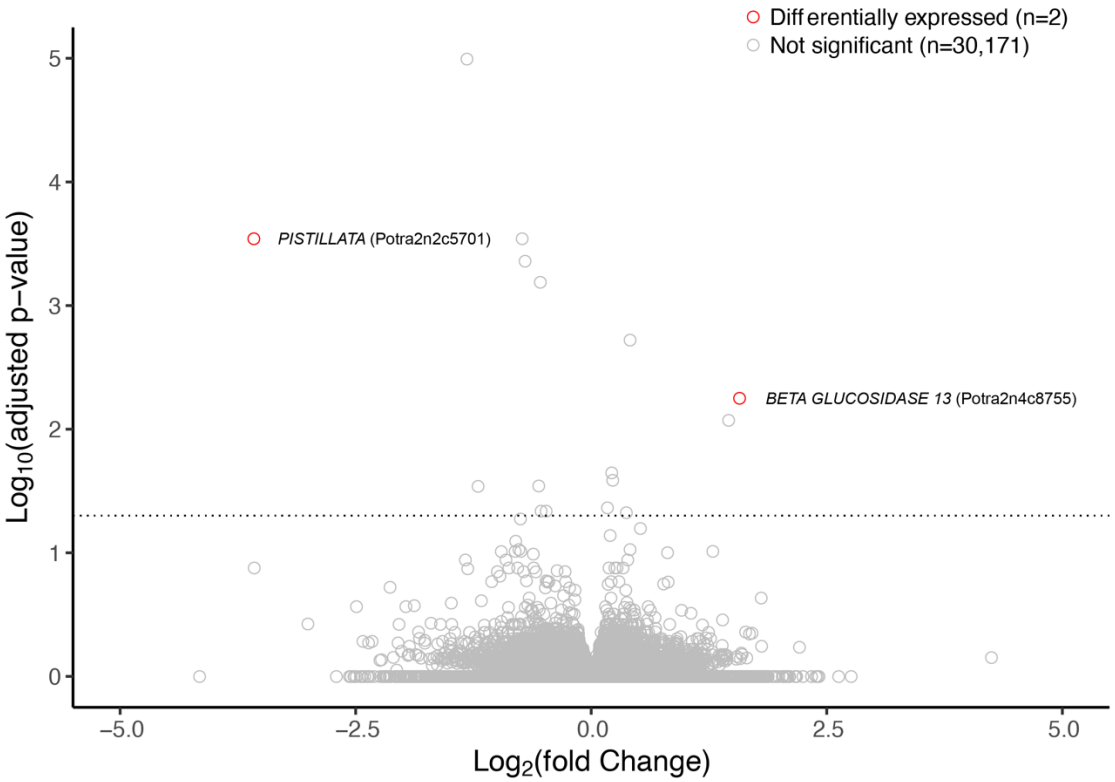


Figure S3. Volcano plot for the time series (Batch A). Only two genes – *PISTILLATA* (Potra2n2c5701) and *BETA GLUCOSIDASE 13* (Potra2n4c8755) – appear as differentially expressed (p adjusted < 0.05 and absolute Log2 Fold Change > 1.5).

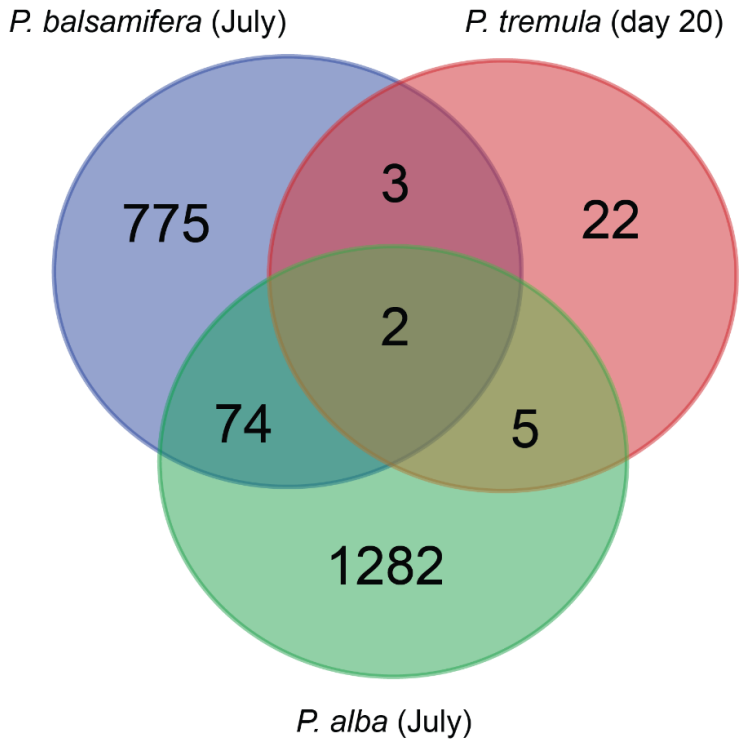


Figure S4. Overlapping sex-differential expressed genes between our early-flowering isogenic aspen (*P. tremula*) lines at day 20, our *P. alba* collected in July 2020 and *P. balsamifera* collected in July from Cronk et al. (2020).

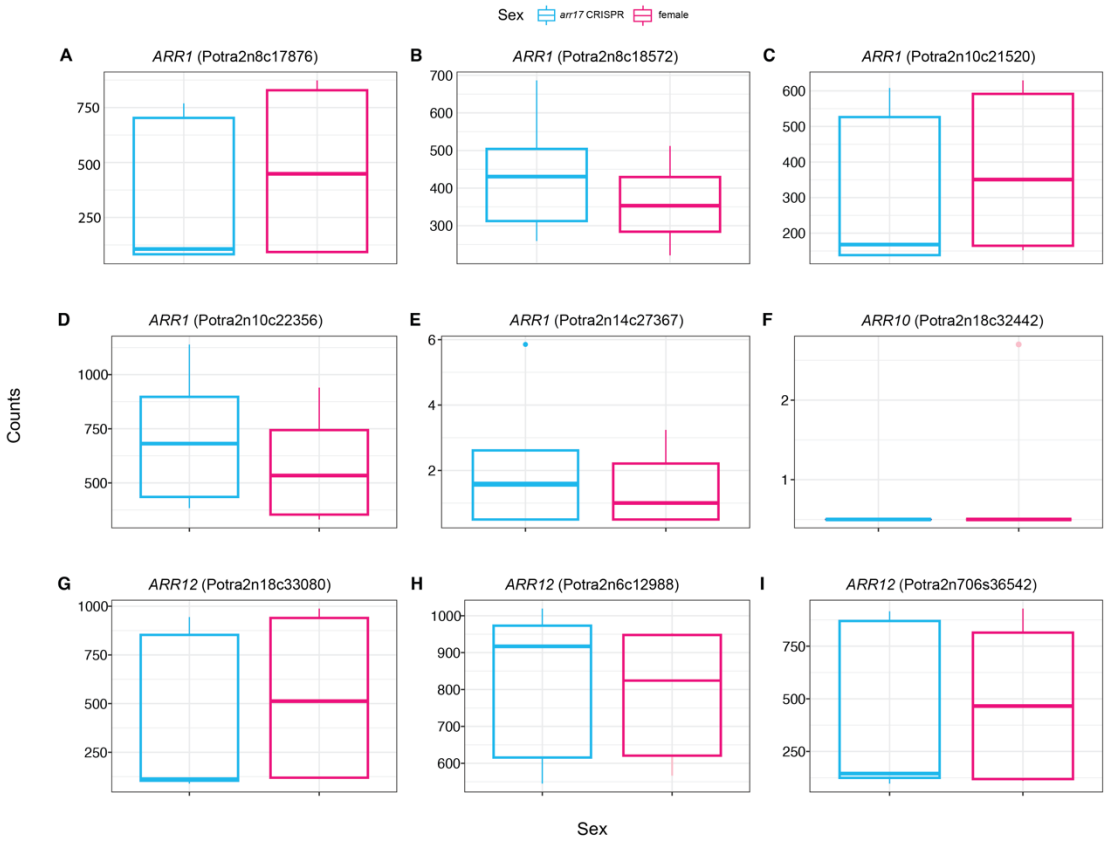


Figure S5. Read counts for *arr17* CRISPR mutants and female highlights no differential gene expression for the type-A response regulators *ARR1*, *ARR10* and *ARR12* that are key components of the cytokinin signaling pathway.

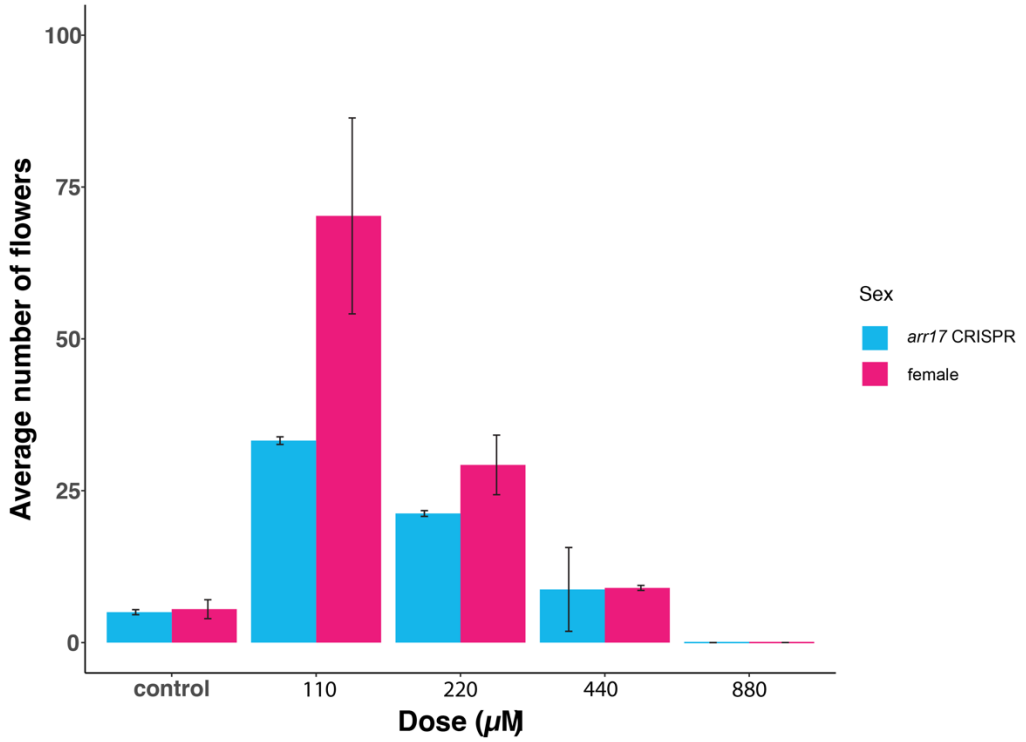


Figure S6. Total number of flowers observed after cytokinin overexpression under different concentrations in male and female flowers of early flowering aspen lines.

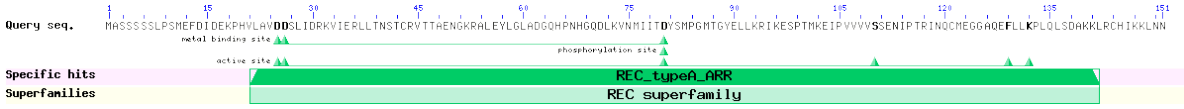


Figure S7. The type-A response regulator *ARR17* can be considered as a single-domain response regulator.

References

Akagi, T., Henry, I.M., Ohtani, H., Morimoto, T., Beppu, K., Kataoka, I., Tao, R., 2018. A Y-encoded suppressor of feminization arose via lineage-specific duplication of a cytokinin response regulator in kiwifruit. Plant Cell 30, 780–795. https://doi.org/10.1105/tpc.17.00787

Alexa, A., Rahnenfuhrer, J., 2020. Gene set enrichment analysis for gene ontology with topGO. R package 26.

Boes, T.K., Strauss, S.H., 1994. Floral phenology and morphology of Black Cottonwood, *Populus trichocarpa* (Salicaceae). American Journal of Botany 81, 562. https://doi.org/10.2307/2445730

Brenner, W.G., Schmülling, T., 2015. Summarizing and exploring data of a decade of cytokinin-related transcriptomics. Frontiers in Plant Science 6, 1–13. https://doi.org/10.3389/fpls.2015.00029

Chae, E., Tan, Q.K.-G., Hill, T.A., Irish, V.F., 2008. An Arabidopsis F-box protein acts as a transcriptional co-factor to regulate floral development. Development 135, 1235–1245. https://doi.org/10.1242/dev.015842

Chen, Z., Rao, P., Yang, Xiaoyu, Su, X., Zhao, T., Gao, K., Yang, Xiong, An, X., 2018. A global view of transcriptome dynamics during male floral bud development in *Populus tomentosa*. Sci Rep 8, 722. https://doi.org/10.1038/s41598-017-18084-5

Coen, E.S., Meyerowitz, E.M., 1991. The war of the whorls: genetic interactions controlling flower development. Nature 353, 31–37. https://doi.org/10.1038/353031a0

Cronk, Q., Müller, N.A., 2020. Default sex and single gene sex determination in dioecious plants. Front. Plant Sci. 11, 1–5. https://doi.org/10.3389/fpls.2020.01162

Cronk, Q., Soolanayakanahally, R., Bräutigam, K., 2020. Gene expression trajectories during male and female reproductive development in balsam poplar (*Populus balsamifera* L.). Sci Rep 10, 8413. https://doi.org/10.1038/s41598-020-64938-w

Cronk, Q.C.B., 2005. Plant eco-devo: the potential of poplar as a model organism. New Phytologist 166, 39–48. https://doi.org/10.1111/j.1469-8137.2005.01369.x

D'Agostino, I.B., Deruère, J., Kieber, J.J., 2000. Characterization of the response of the Arabidopsis response regulator gene family to cytokinin. Plant Physiol 124, 1706–1717. https://doi.org/10.1104/pp.124.4.1706

Diggle, P.K., Stilio, V.S.D., Gschwend, A.R., Golenberg, E.M., Moore, R.C., Russell, J.R.W., Sinclair, J.P., 2011. Multiple developmental processes underlie sex differentiation in angiosperms. Trends in Genetics 27, 368–376. https://doi.org/10.1016/j.tig.2011.05.003

Geraldes, A., Hefer, C.A., Capron, A., Kolosova, N., Martinez-Nuñez, F., Soolanayakanahally, R.Y., Stanton, B., Guy, R.D., Mansfield, S.D., Douglas, C.J., Cronk, Q.C.B., 2015. Recent Y chromosome divergence despite ancient origin of dioecy in poplars (*Populus*). Molecular Ecology 24, 3243–3256. https://doi.org/10.1111/mec.13126

Gramzow, L., Theissen, G., 2010. A hitchhiker's guide to the MADS world of plants. Genome Biology 11, 214. https://doi.org/10.1186/gb-2010-11-6-214

Hoenicka, H., Lehnhardt, D., Briones, V., Nilsson, O., Fladung, M., 2016. Low temperatures are required to induce the development of fertile flowers in transgenic male and female early flowering poplar (*Populus tremula* L.). Tree Physiol 36, 667–677. https://doi.org/10.1093/treephys/tpw015

Honma, T., Goto, K., 2000. The Arabidopsis floral homeotic gene *PISTILLATA* is regulated by discrete cis-elements responsive to induction and maintenance signals. Development 127, 2021–2030.

Jenal, U., Galperin, M.Y., 2009. Single-domain response regulators: molecular switches with emerging roles in cell organization and dynamics. Curr Opin Microbiol 12, 152–160. https://doi.org/10.1016/j.mib.2009.01.010

Kaufmann, K., Melzer, R., Theißen, G., 2005. MIKC-type MADS-domain proteins: structural modularity, protein interactions and network evolution in land plants. Gene 347, 183–198. https://doi.org/10.1016/j.gene.2004.12.014

Kersten, B., Pakull, B., Groppe, K., Lueneburg, J., Fladung, M., 2014. The sex-linked region in *Populus tremuloides* Turesson 141 corresponds to a pericentromeric region of about two million base pairs on *P. trichocarpa* chromosome 19. Plant Biology 16, 411–418. https://doi.org/10.1111/plb.12048

Kieber, J., J., Schaller, G.E., 2018. Cytokinin signaling in plant development. Development 145. https://doi.org/10.1242/dev.149344

Lee, I., Wolfe, D.S., Nilsson, O., Weigel, D., 1997. A *LEAFY* co-regulator encoded by *UNUSUAL FLORAL ORGANS*. Current Biology 7, 95–104. https://doi.org/10.1016/S0960-9822(06)00053-4

Leuendorf, J.E., Schmülling, T., 2021. Meeting at the DNA: Specifying cytokinin responses through transcription factor complex formation. Plants 10, 1458. https://doi.org/10.3390/plants10071458

Liao, Y., Smyth, G.K., Shi, W., 2019. The R package Rsubread is easier, faster, cheaper and better for alignment and quantification of RNA sequencing reads. Nucleic Acids Research 47, 1–9. https://doi.org/10.1093/nar/gkz114

Liu, Z., Mara, C., 2010. Regulatory mechanisms for floral homeotic gene expression. Seminars in Cell & Developmental Biology, Tumor-Stroma Interactions 21, 80–86. https://doi.org/10.1016/j.semcdb.2009.11.012

Love, M.I., Huber, W., Anders, S., 2014. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biology 15, 550. https://doi.org/10.1186/s13059-014-0550-8

McKown, A.D., Klápště, J., Guy, R.D., Soolanayakanahally, R.Y., La Mantia, J., Porth, I., Skyba, O., Unda, F., Douglas, C.J., El-Kassaby, Y.A., Hamelin, R.C., Mansfield, S.D., Cronk, Q.C.B., 2017. Sexual homomorphism in dioecious trees: extensive tests fail to detect sexual dimorphism in *Populus*. Sci Rep 7, 1831. https://doi.org/10.1038/s41598-017-01893-z

Müller, N.A., Kersten, B., Leite Montalvão, A.P., Mähler, N., Bernhardsson, C., Bräutigam, K., Carracedo Lorenzo, Z., Hoenicka, H., Kumar, V., Mader, M., Pakull, B., Robinson, K.M., Sabatti, M., Vettori, C., Ingvarsson, P.K., Cronk, Q., Street, N.R., Fladung, M., 2020. A single gene underlies the dynamic evolution of poplar sex determination. Nature Plants 6, 630–637. https://doi.org/10.1038/s41477-020-0672-9

Murai, K., 2013. Homeotic genes and the ABCDE model for floral organ formation in wheat. Plants (Basel) 2, 379–395. https://doi.org/10.3390/plants2030379

Ng, M., Yanofsky, M.F., 2001. Activation of the Arabidopsis B class homeotic genes by *APETALA1*. Plant Cell 13, 739–754.

Pan, K.-H., Lih, C.-J., Cohen, S.N., 2005. Effects of threshold choice on biological conclusions reached during analysis of gene expression by DNA microarrays. PNAS 102, 8961–8965. https://doi.org/10.1073/pnas.0502674102

Paolucci, I., Gaudet, M., Jorge, V., Beritognolo, I., Terzoli, S., Kuzminsky, E., Muleo, R., Scarascia Mugnozza, G., Sabatti, M., 2010. Genetic linkage maps of *Populus alba* L. and comparative mapping analysis of sex determination across *Populus* species. Tree Genet. Genomes 6, 863–875. https://doi.org/10.1007/s11295-010-0297-7

Parcy, F., Nilsson, O., Busch, M.A., Lee, I., Weigel, D., 1998. A genetic framework for floral patterning. Nature 395, 561–566. https://doi.org/10.1038/26903

Ramírez-Carvajal, G.A., Morse, A.M., Davis, J.M., 2008. Transcript profiles of the cytokinin response regulator gene family in *Populus* imply diverse roles in plant development. New Phytologist 177, 77–89. https://doi.org/10.1111/j.1469-8137.2007.02240.x

Ren, B., Liang, Y., Deng, Y., Chen, Q., Zhang, J., Yang, X., Zuo, J., 2009. Genome-wide comparative analysis of type-A Arabidopsis response regulator genes by overexpression studies reveals their diverse roles and regulatory mechanisms in cytokinin signaling. Cell Res 19, 1178–1190. https://doi.org/10.1038/cr.2009.88

Renner, S.S., Müller, N.A., 2021. Plant sex chromosomes defy evolutionary models of expanding recombination suppression and genetic degeneration. Nature Plants 7, 392–402. https://doi.org/10.1038/s41477-021-00884-3

Ritchie, M.E., Phipson, B., Wu, D., Hu, Y., Law, C.W., Shi, W., Smyth, G.K., 2015. *limma* powers differential expression analyses for RNA-sequencing and microarray studies. Nucleic Acids Research 43, 1–13. https://doi.org/10.1093/nar/gkv007

Robinson, K.M., Delhomme, N., Mähler, N., Schiffthaler, B., Önskog, J., Albrectsen, B.R., Ingvarsson, P.K., Hvidsten, T.R., Jansson, S., Street, N.R., 2014. *Populus tremula* (European aspen) shows no evidence of sexual dimorphism. BMC Plant Biol 14, 276. https://doi.org/10.1186/s12870-014-0276-5

Salomé, P.A., To, J.P.C., Kieber, J.J., McClung, C.R., 2006. Arabidopsis response regulators *ARR3* and *ARR4* play cytokinin-independent roles in the control of circadian period. Plant Cell 18, 55–69. https://doi.org/10.1105/tpc.105.037994

Sanderson, B.J., Wang, L., Tiffin, P., Wu, Z., Olson, M.S., 2019. Sex-biased gene expression in flowers, but not leaves, reveals secondary sexual dimorphism in *Populus balsamifera*. New Phytol. 221, 527–539. https://doi.org/10.1111/nph.15421

Sarkar, M.K., Paul, K., Blair, D., 2010. Chemotaxis signaling protein CheY binds to the rotor protein FliN to control the direction of flagellar rotation in *Escherichia coli*. Proc Natl Acad Sci USA 107, 9370–9375. https://doi.org/10.1073/pnas.1000935107

Schiffthaler, B., Bernhardsson, C., Jenkins, J., Jansson, S., Ingvarsson, P., Schmutz, J., Street, N., 2019. An improved genome assembly of the European Aspen *Populus tremula*. bioRxiv 1–10.

Schmittgen, T.D., Livak, K.J., 2008. Analyzing real-time PCR data by the comparative CT method. Nat Protoc 3, 1101–1108. https://doi.org/10.1038/nprot.2008.73

Schöffl, F., Raschke, E., Nagao, R.T., 1984. The DNA sequence analysis of soybean heat-shock genes and identification of possible regulatory promoter elements. EMBO J 3, 2491–2497.

Souer, E., Rebocho, A.B., Bliek, M., Kusters, E., de Bruin, R.A.M., Koes, R., 2008. Patterning of inflorescences and flowers by the F-Box protein *DOUBLE TOP* and the *LEAFY* homolog *ABERRANT LEAF AND FLOWER* of petunia. Plant Cell 20, 2033–2048. https://doi.org/10.1105/tpc.108.060871

Stock, A.M., Robinson, V.L., Goudreau, P.N., 2000. Two-component signal transduction. Annu. Rev. Biochem. 69, 183–215. https://doi.org/10.1146/annurev.biochem.69.1.183

Theißen, G., Saedler, H., 2001. Floral quartets. Nature 409, 469–471. https://doi.org/10.1038/35054172

To, J.P.C., Haberer, G., Ferreira, F.J., Deruère, J., Mason, M.G., Schaller, G.E., Alonso, J.M., Ecker, J.R., Kieber, J.J., 2004. Type-A Arabidopsis response regulators are partially redundant negative regulators of cytokinin signaling. Plant Cell 16, 658–671. https://doi.org/10.1105/tpc.018978

Vatén, A., Soyars, C.L., Tarr, P.T., Nimchuk, Z.L., Bergmann, D.C., 2018. Modulation of asymmetric division diversity through cytokinin and *SPEECHLESS* regulatory interactions in the *Arabidopsis* stomatal lineage. Developmental Cell 47, 53-66.e5. https://doi.org/10.1016/j.devcel.2018.08.007

Wang, Z., Jiao, Z., Xu, P., Chen, L., Ai, J., Liu, X., Yang, Y., 2013. Bisexual flower ontogeny after chemical induction and berry characteristics evaluation in male *Vitis amurensis* Rupr. Scientia Horticulturae 162, 11–19. https://doi.org/10.1016/j.scienta.2013.07.038

Yang, H.-W., Akagi, T., Kawakatsu, T., Tao, R., 2019. Gene networks orchestrated by *MeGI*: a single-factor mechanism underlying sex determination in persimmon. The Plant Journal 98, 97–111. https://doi.org/10.1111/tpj.14202

Yang, W., Wang, D., Li, Y., Zhang, Z., Tong, S., Li, M., Zhang, X., Zhang, L., Ren, L., Ma, X., Zhou, R., Sanderson, B.J., Keefover-Ring, K., Yin, T., Smart, L.B., Liu, J., DiFazio, S.P., Olson, M., Ma, T., 2021. A general model to explain repeated turnovers of sex determination in the Salicaceae. Molecular Biology and Evolution 38, 968–980. https://doi.org/10.1093/molbev/msaa261

Yonemori, K., Sugiura, A., Tanaka, K., Kameda, K., 1993. Floral ontogeny and sex determination in monoecious-type persimmons. Journal of the American Society for Horticultural Science 118, 293–297. https://doi.org/10.21273/JASHS.118.2.293

Zhao, D., Yu, Q., Chen, M., Ma, H., 2001. The *ASK1* gene regulates B function gene expression in cooperation with *UFO* and *LEAFY* in Arabidopsis. Development 128, 2735–2746.

Zhou, R., Macaya-Sanz, D., Schmutz, J., Jenkins, J.W., Tuskan, G.A., DiFazio, S.P., 2020. Sequencing and analysis of the sex determination region of *Populus trichocarpa*. Genes (Basel) 11, 843. https://doi.org/10.3390/genes11080843

General discussion

During this PhD project, I took advantage of the natural variation in the genus *Populus* to resolve the genetic basis and regulatory mechanisms involved in dioecy. This has been possible after i) studying the theoretical background of sex determination in flowering plants, ii) investigating *Populus* sex chromosomes utilizing publicly available genomes as well as newly generated data, including long-read sequencing, iii) experimentally testing whether sex determination in poplars is controlled by a single gene acting as a sex switch, iv) profiling gene expression to describe the molecular function of *ARR17*, v) using experimental populations to assess recombination frequencies along the SDR in white poplar and vi) exploring the system turnovers in *Populus* and testing the practical uses of *ARR17* as a sex marker. I now discuss and connect the results from each chapter to highlight the implications of the genetic basis and molecular mechanism of dioecy in poplars controlled by *ARR17*.

Sex determination in dioecious plants

Sex determination of monoecious and dioecious plant species has been investigated for over a century and despite progress, there are still several open questions surrounding the evolutionary dynamics and the genetic basis underlying unisexuality in plants. As an increasing number of plant systems are being studied, the emerging experimental data provide new insights and alternatives to the classical theoretical frameworks reinforcing the view that the mechanisms of sex determination in dioecious plants are diverse (Diggle et al., 2011; Golenberg and West, 2013; **Leite Montalvão et al., 2021**; Ming et al., 2011; Renner, 2016).

Although most studies shown in the **Chapter 1** were targeted at identifying a region of suppressed recombination comprising two sex-determining genes involved in female and male sterility, such as *SOFF* and a*spTDF1* in asparagus (Harkess et al., 2020) or *FrBy* and *SyGI* in kiwifruit (Akagi et al., 2019) the artificial generation of dioecy from monoecy in maize (Jones, 1934) and melon (Boualem et al., 2015) demonstrated a possible alternative pathway for the evolution of dioecy, and outlined an important contrast between two evolutionary pathways to dioecy (**Chapter 1: Figure 3**). There can be either

physical linkage by having two sex-determining factors within a non-recombining SDR, which was long considered the most likely scenario, or alternatively genetic linkage (**Leite Montalvão et al., 2021**). In the latter case only a single sex-determinant is needed in the SDR since the other factor is connected via an epistatic genetic interaction (Boualem et al., 2015; Golenberg and West, 2013; **Leite Montalvão et al., 2021**), establishing the feasibility of a single gene controlling the development of unisexual flowers. Among dioecious angiosperm species, *MeGI* in persimmon (Akagi et al., 2014) and *ARR17* in poplar (**Müller et al., 2020**) (**Chapter 2**) represents the currently known naturally occurring single-gene systems, functioning as master regulators of sex determination, similar to sex-determining region of the Y (*SRY*) in placental and marsupial mammals (Gubbay et al., 1990; Kashimada and Koopman, 2010).

Sex chromosomes, SDRs and suppression of recombination

A major theoretical hypothesis for sex chromosome evolution is centered on the two-gene model and states that selection against recombination favors the coinheritance of the two sex-linked antagonistic genes (Bachtrog et al., 2014; Charlesworth, 2021; Charlesworth et al., 2005; Ming et al., 2011). Still, sex chromosomes and SDRs can evolve in single-gene systems in which there is no need to select for suppressed recombination since a second genetically linked factor is not required (Renner and Müller, 2021).

Different regions of the sex chromosomes have been nonrecombining over different amounts of time (Charlesworth et al., 2005) resulting in sex chromosomes with SDRs of varied differentiation (Hobza et al., 2018). While some SDRs evolve and stay conserved across all species in some genera such as in *Vitis* (Zou et al., 2021), others such as in *Populus* present a strikingly dynamic genetic architecture characterized by frequent sex chromosome turnover. The accumulating data describe various genomic locations for the SDRs and distinct sizes in different *Populus* species. Besides the two SDRs initially reported (in balsam poplar and aspens), two more distinct SDRs (in white poplar and desert poplar) were revealed and at least two turnover events (XY ↔ ZW) occurred (**Figure 6**) (Geraldes et al., 2015; Kersten et al., 2014; **Kim et al., 2021**; **Müller et al., 2020**; Yang et al., 2021), indicating an independent evolution of the SDRs across the genus.

This is comparatively different from the jumping SDR cassette that occurs in strawberries (Tennessen et al., 2018).

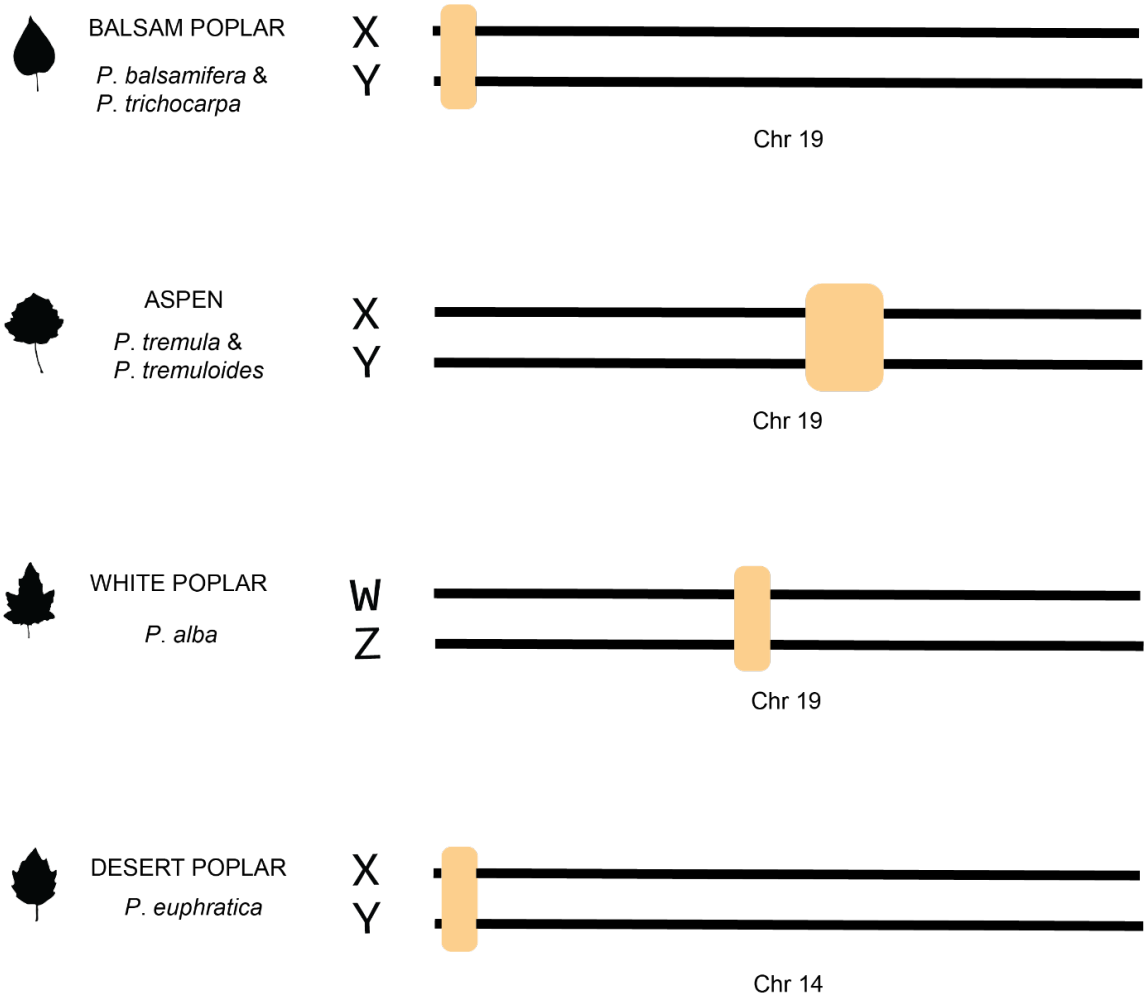


FIGURE 6. DIFFERENT GENOMIC ARCHITECTURE OF SEX DETERMINATION ACROSS THE GENUS *POPULUS*. A SMALL SDR IN BALSAM POPLAR IS LOCATED AT THE BEGINNING OF CHROMOSOME 19, WHILE IN ASPEN THE SDR IS LARGER AND LOCATED IN A PERICENTROMERIC REGION OF CHROMOSOME 19 (BOTH WITH XY SYSTEM). DESPITE THE SMALL SDR IN WHITE POPLAR ALSO BEING LOCATED IN THE MIDDLE OF CHROMOSOME 19, IT PRESENTS A FEMALE HETEROGAMETIC SYSTEM (ZW). IN THE DESERT POPLAR, THE SMALL SDR IS LOCATED AT THE BEGINNING OF CHROMOSOME 14 IN A XY SYSTEM. THE SDRS ARE REPRESENTED BY THE ORANGE RECTANGLE.

Aspens and white poplar belong to the same section (sect. Populus) within the genus, although they exhibit different heterogametic systems (XY and ZW, respectively). Turnover events between XY and ZW systems have been well documented in animals, especially snakes, fishes and amphibians (Furman et al., 2020), but also in other plant genera such as *Dioscorea*, *Salix* and *Silene* (Carey et al., 2021; **Leite Montalvão et al., 2021**; Ming et al., 2011). In *Populus*, these transitions suggest a convergent evolution of sex

determination within the genus (**Chapter 4: Figure 6**). The XY system from aspens seems to have re-evolved from the ZW system present in *P. alba* (and possibly *P. adenopoda* Maxim. and *P. qiongdaoensis* T. Hong & P. Luo), and it is mechanistically highly similar to the much older XY system present in balsam poplars (**Kim et al., 2021**) (**Chapter 4**).

All poplar species evaluated to date present homomorphic sex chromosomes. Interestingly, aspens present a centrally located larger SDR (~1 Mb) (Kersten et al., 2014; **Müller et al., 2020**) compared to evolutionarily older species such as *P. euphratica* Oliv., *P. balsamifera* L., *P. trichocarpa* and *P. alba* in which the SDR ranges between 100 and 300 kb (Geraldes et al., 2015; **Kim et al., 2021**; **Müller et al., 2020**; Yang et al., 2021). Comparing the patterns of SDR evolution in closely related species can provide insights regarding sex chromosome evolution in the genus *Populus*.

Thus, to test the hypothesis that the SDR of white poplar shows suppressed recombination and it is linked to a segregation distorter, the recombination frequencies and segregation patterns of a white poplar F1 mapping population were investigated in **Chapter 3**. While it was not possible to draw conclusions whether recombination is suppressed between the paternal alleles Z_1 and Z_2, Z_3 and W did not recombine at all (**Chapter 3: Figure 3**). From the multiple possible scenarios in which sex-determining genes and SDRs can arise (Charlesworth, 2021), the turnover event changing heterogamety in white poplars seem to have a direct effect on recombination suppression enabling the evolution of a small hemizygous SDR (~200 kb) based on the single-gene sex-determining system. Hemizygosity of the W-specific poplar sex-determining gene *ARR17* may directly cause the recombination suppression between Z and W (**Chapter 3**). A relatively small SDR (~1.3 Mb) was also identified in persimmon, which is another species with a single-gene system of sex determination (Akagi et al., 2014). In the persimmon SDR there is no antagonistic sex-linked gene, thus a large SDR is not likely to evolve (Akagi and Charlesworth, 2019). Physically small SDRs in homomorphic sex chromosomes are not an exclusive characteristic of single-gene systems. Species consistent with the two-gene model such as kiwifruit, asparagus and grapes also present relatively small SDRs (Akagi et al., 2019; Carey et al., 2021; Fechter et al., 2012; Harkess et al., 2017; Picq et al., 2014).

The expansion of the non-recombining region and heteromorphy have been often considered as hallmarks of sex chromosome evolution (Ming et al., 2011). Large SDRs on heteromorphic sex chromosomes have been identified in a few species such as *Cannabis sativa* L. (Prentout et al., 2020), *Coccinia grandis* (L.) Voigt (Sousa et al., 2013) and *Silene latifolia* Poir. (Papadopulos et al., 2015), although most plant species have retained homomorphic sex chromosomes (Carey et al., 2021). Those with small SDRs have often been interpreted to be evolutionary young (Fechter et al., 2012; Geraldes et al., 2015; Pucholt et al., 2017; Tennessen et al., 2018). Nevertheless, the size of the SDRs and their age do not seem to be correlated (Renner and Müller, 2021). In fact, the existence of homomorphic sex chromosomes with small SDRs in many dioecious plant species may be a result of reduced selection for linked sexually antagonistic alleles (in genes involved in sexually antagonistic traits) in non-recombining regions, which in turn, may favor keeping sex determination precise and flexible (Renner and Müller, 2021) challenging the established idea that expansion of SDRs is an inevitable result of sex chromosome evolution (Charlesworth, 1991).

Sex-determining genes may also evolve within an already physically larger non-recombining region (Charlesworth, 2021). For instance, in aspens the SDR evolved within the pericentromeric region (Kersten et al., 2014), considered to be a non-recombining region (Stapley et al., 2017). Considering the widespread occurrence of non-recombining haplotype blocks (Todesco et al., 2020), the evolution of an SDR in an already recombination-suppressed region does not appear unlikely. Further research on the evolution of the aspen SDR is necessary to confirm this hypothesis.

More generally, the identification of sex-linked non-recombining regions is still a crucial first step in narrowing down putative sex-determining genes for in-depth analysis of their sex-related functions.

Sex-determining genes and their functional validation

The progress made in investigating the genetic basis of sex determination in several dioecious plants in the past years exposed the many different sex-determining genes recruited in the different species (Carey et al., 2021; Henry et al., 2018; **Leite Montalvão et al., 2021**; Ming et al., 2011; Renner and Müller, 2021). Many different genes and

pathways can be potential targets of the sex determinants. Their expression can affect pollen viability, ovule formation, embryo sac, tapetum cells, or other key tissues involved in gamete formation and function (Renner and Müller, 2021).

Sex determination in species consistent with the two-gene model (e.g. kiwifruit, asparagus) involve independent DNA polymorphisms affecting male and female sterility functions (Charlesworth and Charlesworth, 1978). The functional validation of sex-determining genes can be limited, since not all plant species can be genetically transformed, and it thus sometimes performed in model plants such as *A. thaliana* or *Nicotiana tabacum* L. The characterization of the biological function of the sex-determining genes consistent with the two-gene model has been demonstrated in these model plants either suppressing carpel development (i.e. *SyGI* in kiwifruit, *SOFF* in asparagus and possibly *LOG*-like in date palm and *VviYABBY3* in grapes) or promoting pollen development (i.e. *FrBy* in kiwifruit, *aspTDF1* in asparagus and *VviINP1* in grapes) (**Chapter 1**).

In the single-gene system in persimmon, the function of *MeGI* has been validated in the model system *Nicotiana benthamiana* Domin, *N. tabacum*, and *A. thaliana* (Akagi et al., 2014). In contrast, in **Chapter 2** the early flowering aspen lines were instrumental to experimentally validate the sex-determining function of *ARR17* directly in poplars (i.e. *P. tremula*).

The evolution of dioecy in single-gene systems may involve whole-genome duplication events giving rise to the sex-determining gene (Akagi et al., 2020; Carey et al., 2021; Zhou et al., 2018; Hou et al., 2019). Comparative genomics revealed that numerous gene duplication events occurred during plant speciation, and thousands of paralogous genes were formed, e.g. in *Populus* (Tuskan et al., 2006).

The single-gene system in *Populus*, which is conserved despite millions of years of independent evolution, offers a high level of flexibility. This flexibility is a consequence of genome duplication events, which in turn allows functional differentiation between paralogs (i.e. neofunctionalization of the sex-determining genes), and establishment of a new sexual system (Akagi et al., 2020; Henry et al., 2018). Similarly, the persimmon SDR evolved due to a duplication event of a pre-existing gene causing suppression of

recombination in the male-specific region (Akagi et al., 2014). The persimmon *MeGI* acquired a new function allowing control of sex determination, whereas its paralog, sister of *MeGI* (*SiMeGI*), did not evolve any sex-related functions. Instead, a second segmental-duplication gave rise to the Y-specific gene, *OGI* (Akagi et al., 2020).

Notwithstanding the evolutionary pathway, the accumulating data indicate that most of these sex-determining genes (and candidate genes in other plant species) may be involved in similar signaling pathways (e.g. cytokinin pathway) or have expression patterns in specific tissues (e.g. tapetum cells for pollen development) causing similar changes of the underlying genetic networks (**Chapter 1: Figure 4**) (Carey et al., 2021; Feng et al., 2020; **Leite Montalvão et al., 2021**).

The poplar sex switch ARR17

In line with the hypothesis that *ARR17* is a feminizing factor in aspen, the poplar *ARR17* is a high-level sex switch, and its biological function is to "make" female flowers. Essentially, by default male development occurs, unless *ARR17* is expressed, which initiates female development instead (Cronk and Müller, 2020; **Müller et al., 2020**) regardless the sexual system – XY or ZW (**Chapter 2**). The underlying mechanism of sex determination in *Populus* seems to have been preserved, supporting the hypothesis that *ARR17* is a master sex switch shared across the genus *Populus*, and it is regulated simply through different expression patterns of the same high-level factor – the poplar *ARR17* gene, which may or may not be located at the SDR.

In poplars with XY system, the *ARR17* gene is not at the SDR. Instead, there is a dominant negative regulator – the Y-specific partial *ARR17* duplicates – responsible for sex determination in aspen and balsam poplars by targeting and repressing the *ARR17* gene at the end of the chromosome (Geraldes et al., 2015; **Müller et al., 2020**) (**Chapter 2**). These Y-specific duplications form an RNA hairpin structure that is processed into small RNAs (particularly 24-nt), which in turn down-regulate *ARR17* expression, thereby allowing male development (Akagi et al., 2014; **Müller et al., 2020**) (**Chapter 2**). The silencing mechanism by trans silencing via non-coding RNAs (e.g. sRNA, lncRNAs) triggers DNA methylation of the target sequence via RNA-directed DNA Methylation (RdDM) (Erdmann and Picard, 2020; Muskens et al., 2000; Wambui Mbichi et al., 2020).

Male-specific DNA methylation in the *ARR17* locus was reported in *P. balsamifera* (Bräutigam et al., 2017) and *P. tremula* (Müller et al., 2020). Collectively, these factors support the hypothesis of trans silencing of *ARR17* via RdDM. On the other hand, sex determination in white poplar is based on a ZW system with a Z-specific deletion of *ARR17* whereas three copies of the W-specific *ARR17* are located within the SDR (**Müller et al., 2020**; Yang et al., 2021). The location of the sex switch i.e. at the end or the middle of the chromosome exerts the same control over sex determination regardless the position (**Chapter 1: Figure 2**).

Sex-determining genes as molecular sex markers

Phenotypic differences (i.e. sexual dimorphism) in dioecious plants are more subtle than in animals. A few dioecious plant species, such as *Mercurialis annua* L. (Cossard et al., 2019) show differences in the plant architecture earlier during development, allowing to discriminate among the sexes. Most dioecious plant, however, show minimal sexual differences before flowering (Renner and Müller, 2021).

Genetic mapping of sex determination and the identification of sex-specific markers is not only biologically important, but also has great practical relevance in guiding plant breeding and management strategies (Charlesworth and Mank, 2010). Female individuals are often valued in agriculture due to the production of seeds and fruits, while male individuals might be desired due to biomass production, fiber or chemical content (Charlesworth and Mank, 2010; Milewicz and Sawicki, 2013). Moreover, the obligatory outcrossing nature of dioecious species results in hybrid vigor of the offspring, which is an advantageous consequence of dioecy (Henry et al., 2018) and pursued characteristic in several breeding programs. Thus, molecular sexing of juveniles can maximize gains and can be performed in several economically important crops such as date palm (Ali et al., 2018), yam (Tamiru et al., 2017), spinach (Qian et al., 2017), kiwifruit (Zhang et al., 2015) and tree species, such as *Populus* (Gaudet et al., 2007; Geraldes et al., 2015; **Kim et al., 2021**; Pakull et al., 2015) (**Chapter 4**).

Essentially, a male-specific molecular marker indicates an XY system, while the presence of a female-specific marker indicates a ZW system (Gamble and Zarkower, 2014). In the genus *Populus*, besides the previously established Y-specific sex marker *TOZ19* in aspens

(Pakull et al., 2015), genetic sex markers for sex determination in *Populus* can be developed based on W-specific *ARR17* in ZW systems and on Y-specific partial *ARR17* duplicates in XY systems (**Chapter 4**). The white poplar *P. alba* is to date the only species in the genus *Populus* with confirmed female heterogamety (ZW system) (**Müller et al., 2020**; Paolucci et al., 2010; Yang et al., 2021). Since the female SDR is hemizygous and contains basically only the sex-determining gene *ARR17* in three copies, this minimal sexual difference allows using this gene as a molecular marker to differentiate between males and females (robust PCR-based marker) (**Kim et al., 2021**). Moreover, the presence/absence pattern of *ARR17* in *P. alba* (**Müller et al., 2020**; Yang et al., 2021) not only explains the turnover event and change in heterogamety, but also provides preliminary evidence for other ZW systems within the genus. For instance, the DNA-seq coverage from publicly available data indicates that female-specific hemizygous regions (including the *ARR17* gene) are shared in species closely related to *P. alba* (**Chapter 4: Figure 1**). Thus, the *ARR17* gene can be also employed as a sex marker (**Chapter 4: Figure 2**) in the closely related species *P. adenopoda* and the hybrid between the two i.e. *P.* × *tomentosa* (**Kim et al., 2021**).

On the other hand, the Y-specific partial *ARR17* duplicates may be employed to develop PCR-based genetic sex markers in *Populus* species with XY systems, as has been shown so far for *P. grandidentata* Michx. (**Chapter 4: Figure 3**). Although, a final proof for an XY system is still pending in this species, the results - applying the developed PCR-based marker - suggest a male heterogametic system in this species (**Kim et al., 2021**). In other *Populus* species with XY systems, genes that are linked to the *ARR17* partial duplicates but are not involved in sex determination, have been used as genetic sex markers, such as *TOZ19* in aspens (Pakull et al., 2015) and *HEMA1* in *P. trichocarpa*, *P. deltoides* Marshall and *P. nigra* L. (Geraldes et al., 2015).

Floral organ identity

The origin of separate sexes in plants is often associated with the evolution of varying levels of sexual dimorphism which is pronounced within the flowers. Flowering is a complex process, normally connected with the regulation of multiple genetic and environmental factors. Genetically, the transition between vegetative to reproductive

development is initiated based on the expression of *FLOWERING LOCUS T* (*FT*), *FLOWERING LOCUS D* (*FD*), and *SUPRESSOR OF OVEREXPRESSION OF CO 1* (*SOC1*) genes. Once activated, these genes initiate expression of floral meristem identity genes (Blázquez et al., 2001; Melzer et al., 2008). To date, five meristem identity genes in *Arabidospsis* have been identified i.e. *LEAFY* (*LFY*), *APETALA 1* (*AP1*), *CAULIFLOWER* (*CAL*), *FRUITFULL* (*FUL*) and *UNUSUAL FLORAL ORGANS* (*UFO*) controlling the initiation of floral primordia (Pidkowich et al., 1999; Weigel et al., 1992; Ferrándiz et al., 2000). Subsequently, the floral meristem genes activate homeotic genes e.g. *APETALA 3* (*AP3*), *PISTILLATA* (*PI*), *AGAMOUS* (*AG*) and *SEPALLATA* (*SEP*), which are based on the general ABC model (**Figure 7**), and define floral organ identity (Coen and Meyerowitz, 1991; Murai, 2013; Parcy et al., 1998).

The genes from the ABC model, later expanded to the ABCE model (Ditta et al., 2004; Honma and Goto, 2000; Pelaz et al., 2001), were originally identified in *A. thaliana* and *Antirrhinum majus* L. based on mutant phenotypes (Krizek and Fletcher, 2005; Krizek and Meyerowitz, 1996). Several of these genes encode MADS-domain transcription factors, which are essential for correct organ identity in a similar manner to the *HOX* genes in vertebrate body-plan formation (Gramzow and Theissen, 2010; Kaufmann et al., 2005; Krizek and Fletcher, 2005; Mallo et al., 2010). MADS-box family transcription factors contain a number of conserved domains. Those involved in floral organ identification are the Type II MADS proteins which share the MIKC domain structure. The N-terminal MADS domain (M) which is highly conserved and is required for protein-DNA interactions, is followed by the Intervening (I), Keratin-like (K), and C-terminal (C) domains (Becker and Ehlers, 2016).

The theory behind the ABCE model of flower identity assumes that different whorl-specific classes of genes, when combined, control the formation of the structures that comprise a typical hermaphrodite flower, such as sepals, petals, stamens and carpels (Coen and Meyerowitz, 1991). Class A+E genes are responsible for sepal development; A+B+E specify petals; B+C+E account for stamen development and C+E for carpels. The class E genes encode proteins redundantly required to specify the other classes of genes (Ditta et al., 2004; Zahn et al., 2005). Moreover, while the B class functions independently, the functions of A and C classes are mutually exclusive, thus the absence

of one result in the expansion of the other. The variation in the expression of the ABC genes (e.g. loss of function) account for different mutant phenotypes (**Figure 7**).

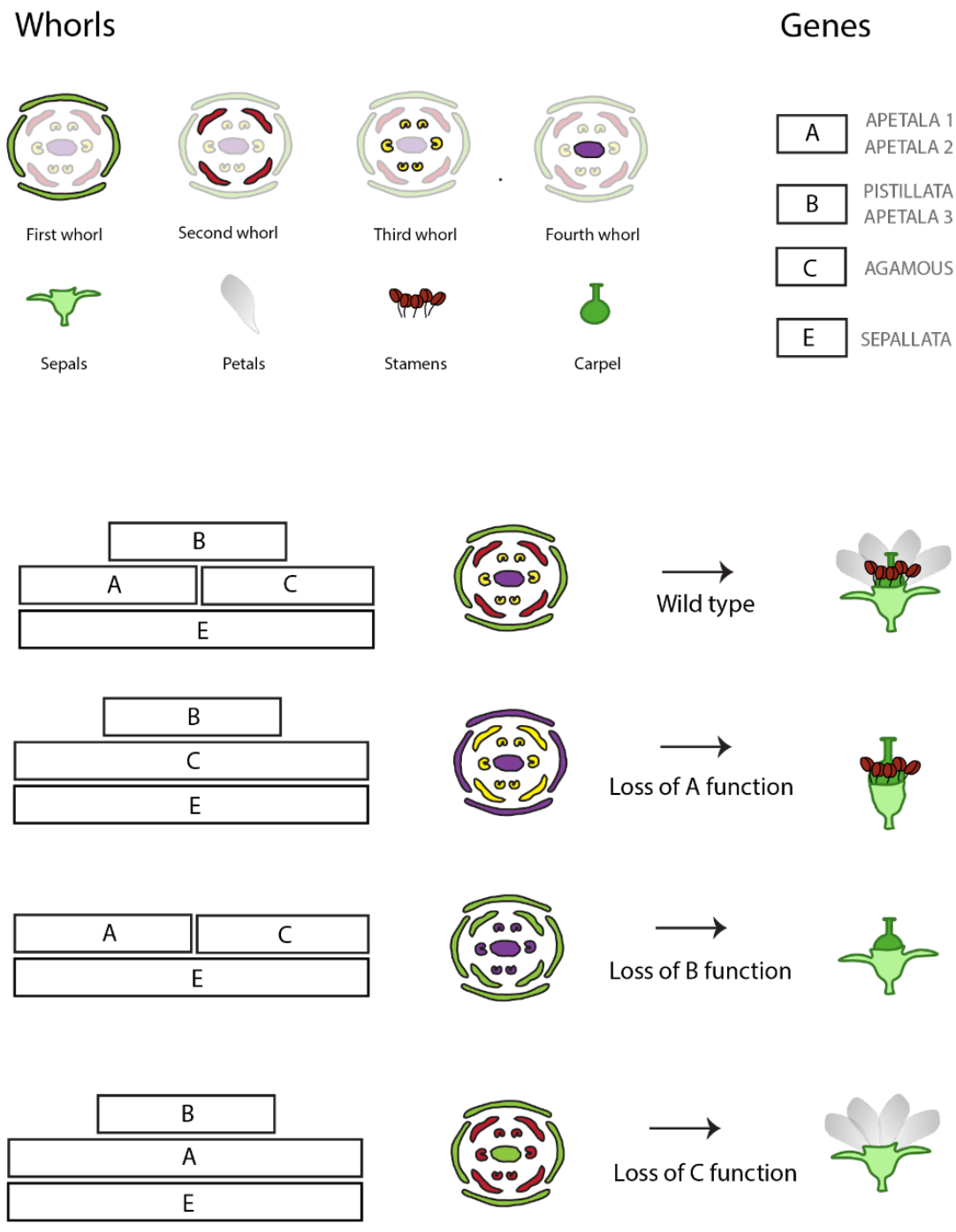


FIGURE 7. ABCE MODEL OF FLOWER ORGAN IDENTITY. THE COMBINATION OF EACH WHORL-SPECIFIC CLASS OF HOMEOTIC GENES DETERMINES THE FLORAL ORGAN IDENTITY. DIFFERENT GENES WERE FOUND FOR EACH CLASS ENCODING MADS-BOX TRANSCRIPTION FACTORS IN *ARABIDOPSIS THALIANA* (A CLASS: *APETALA 1*, *APETALA 2*; B CLASS: *PISTILLATA*, *APETALA 3*; C CLASS: *AGAMOUS*; E CLASS: *SEPALLATA*). THE WILD TYPE CONSISTS OF FOUR WHORLS OF ORGANS; THE LOSS OF A FUNCTION RESULTS IN THE EXPANSION OF CLASS C TO THE FIRST WHORL AND FORMATION OF STAMENS IN THE SECOND AND THIRD WHORLS; THE LOSS OF B FUNCTION PRODUCES FEMALE FLOWERS, WITH CARPELS IN THE THIRD AND FOURTH WHORLS AND SEPALS IN THE FIRST AND SECOND; THE LOSS OF C FUNCTION RESULTS IN THE EXPANSION OF CLASS A, AND ONLY PETALS AND SEPALS ARE FORMED.

In two-gene systems, the connection between the sex-determining genes and the floral phenotype allows the development of floral organs to be activated or repressed separately. While a mutation in the female suppressor converts males into hermaphrodites, a mutation in the male activator converts males into neuters (Renner and Müller, 2021). In contrast, in single-gene systems, such as in poplar, there is only one factor to switch between male and female development (**Müller et al., 2020**). However, the knowledge of the molecular function of such high-level regulators and how they are associated with the floral phenotype is limited (Cronk and Müller, 2020; Yang et al., 2019). While the formation of neuters in males might occur downstream of the flower organ identity by interrupting the genetic cascade responsible for pollen formation, the development of hermaphrodites remains unclear.

The molecular machinery of flower organ identity is associated with the correct expression of MADS-box genes, according to the ABC model (Coen and Meyerowitz, 1991). Unisexual flowers may present an early stage of selective suppression (e.g. *Asparagus officinalis* L., *Actinidia deliciosa* (A. Chev.) C.F Liang & A.R.Ferguson, *Vitis vinifera* L.) leaving rudimentary and non-functional floral organs. Appropriate expression of MADS-box genes is expected to occur so that both male and female sexual structures initiate development. Nevertheless, these genes do not seem to be instrumental for the process of sex determination, since further blocking of the correct development of one of the structures occurs downstream of the ABC genes (Cronk and Müller, 2020). Instead, ABC genes are more likely to be involved in the differentiation of structures at a very early stage, even before the development of stamens and carpels (e.g. *Populus* spp. or *Spinacia oleracea* L.) allowing only one sexual structure to be initiated (Boes and Strauss, 1994; Cronk and Müller, 2020; Diggle et al., 2011; Mao et al., 2017; Sather et al., 2010). The loss of function of specific MADS-box genes (i.e. B class) allows a clean deletion of one of the sexual structures (**Figure 7**) (Cronk and Müller, 2020).

In the case of *Populus*, while the genetic basis of sex determination mediated by *ARR17* has been resolved in **Chapter 2** the molecular function of *ARR17* and its connection to flower phenotype remained unexplored. It has been previously proposed that the

expression of the B class gene pair *PI* and *AP3* is expected in males, as first described in *Arabidopsis* (Krizek and Meyerowitz, 1996). Since males appear to be the default sex, the presence of *ARR17* would downregulate these B class genes, allowing female development (Cronk and Müller, 2020). However, the molecular pathway connecting the poplar *ARR17* gene and the MADS-box genes was not clarified. To test the hypothesis that sex-specific expression networks can provide clues on the molecular connection of *ARR17* to floral phenotypes, **Chapter 5** explored the downstream genetic networks that are specifically controlled by *ARR17* to elucidate its molecular mechanism as well as potential connections between *ARR17* and sex expression (i.e. flower organ development). A major advantage of this study was the use of isogenic early flowering lines, which is comparable to using a monoecious species (Yang et al., 2019). This strategy allows exploring gene expression without associated background effects that typically separate female and male genotypes.

Interestingly, during poplar development, the *ARR17* gene is highly regulated on the transcriptional level. It is only expressed in a certain tissue (female flower buds) in a specific narrow time window during poplar flower development, in line with the minimal sexual dimorphism observed in poplars (McKown et al., 2017; Renner and Müller, 2021; Robinson et al., 2014) (**Chapter 5: Figure 1**). Additionally, *ARR17* does not seem to follow its predicted function in the cytokinin signaling pathway based on its annotation as a type-A response regulator in *A. thaliana.* At least in aspens, the differential gene expression analysis (**Chapter 5: Figure S5**) and an experimental assay utilizing different concentrations of cytokinin did not reveal any connection between *ARR17* and cytokinin signaling (**Chapter5: Figure S6**). The poplar *ARR17* gene appears to have a more specific function connected to flower organ identity (**Chapter 5**). To activate female development and repress male formation simultaneously, the poplar *ARR17* likely controls the expression of *UFO*. Among the different functions of *UFO* is the regulation of B class MADS-box genes *AP3* and *PI* (Krizek and Meyerowitz, 1996). The regulation of B class gene expression was also suggested to control the evolution of dioecy in spinach *S. oleracea* (Sather et al., 2010). The genes *PI* and *AP3* are expressed in early stages of development in male flowers, but not in female ones (Pfent et al., 2005). Additionally, in the mechanism for sex determination in spinach the *GIBBERELIC ACID INSENSITIVE* (*SpGAI*) is critical for feminization while its inhibition results in maleness implying an

epistatic connection between female and male functions, similar to poplar. Moreover, the *SpGAI* acts upstream of the B class genes. Altogether, this suggests *SpGAI* as an important regulator in a single-gene system of sex determination in spinach (West and Golenberg, 2018). The regulation of B class genes was also identified in the monoecious persimmon *Diospyros kaki* L.f., in which one of the pathways regulated by the sex-determining gene *MeGI* involves a direct control of the expression of *SHORT VEGETATIVE PHASE* (*SVP*) gene, which in turn represses *PI* expression (Yang et al., 2019).

As mentioned above, the molecular function of *ARR17* appears to be a dominant suppressor of *UFO* to establish dioecy, and perhaps even monoecy (**Chapter 5: Figure 3**). It could be that *ARR17* already neofunctionalized to act specifically in the flower organ formation in a monoecious ancestor, perhaps to establish gynomonoecy or monoecy (if it was differentially expressed in separate parts of the plant), pondering the question when exactly *ARR17* evolved the function of a sex switch. The unanswered question concerns the mechanism by which *ARR17* represses *UFO*. The small set of differentially expressed genes in aspens do not offer promising candidates for mediating a transcriptional regulation (**Chapter 5: Table 1**), the more so *ARR17* is a single-domain response regulator containing only a receiver domain and missing any kind of effector domain (**Chapter 5: Figure S7**). Therefore, a direct transcriptional regulation of *UFO* by *ARR17* appears unlikely (**Chapter 5**). A future screen for protein-protein interaction e.g. via yeast 2 hybrid assays using *ARR17* as bait could help in narrowing down, e.g. a potential positive regulator (or a repressor) of *UFO* that is inactivated (or activated) by *ARR17* via protein-protein interaction.

Another relevant question concerns whether repressing *UFO* is the only function that *ARR17* has or if a secondary pathway may be triggered by *ARR17*. Further knockout experiments of *PI* and *UFO* will help answering this question (**Chapter 5**). In *D. kaki* for instance, *MeGI* regulates two independent pathways, one for androecium and another for gynoecium development (Yang et al., 2019).

Transitions between sexual systems

New sex determination loci may be formed due to transitions in the sex chromosomes (Charlesworth, 2021). The evolutionary pathways are described in **Chapter 1** (**Leite Montalvão et al., 2021**).

The domestication process of some plant species also contributed to favor sexual transitions, including reverse transitions away from dioecy. These transitions have been studied in a few species, such as grapevine (Zou et al., 2021), persimmon (Yang et al., 2019) and papaya (Chae et al., 2021). Environmental factors such as temperature and hormones have also been reported to favor these sexual transitions (Chae et al., 2021; Furman et al., 2020; Golenberg and West, 2013). In cultivated cucurbits, for instance, the most common sexual system is monoecy, although variations in temperature triggers an epigenetic regulation of sex expression, contributing to different phenotypes, including hermaphroditism (Lai et al., 2018; Li et al., 2019).

The conventional theory asserts that hermaphrodites evolve genetically determined dioecy via a gynodioecious intermediate (Charlesworth and Charlesworth, 1978). However, the alternative evolutionary trajectory from monoecy to dioecy seems to be most common (Renner and Müller, 2021; Renner and Ricklefs, 1995). Would it be possible that the genetic networks differentially regulating female and male flower development are shared between related monoecious and dioecious species? In theory, the separation of unisexual flowers already occurred in the monoecious ancestor (Cronk and Müller, 2020; Renner and Müller, 2021). Thus, the evolution towards dioecy lies on separating the regulatory mechanism to different individuals. This may be accomplished by a single master regulator gene. Thus far, only the Oriental Persimmon (*D. kaki*) provides some evidence for such shared molecular basis of monoecy and dioecy (Akagi et al., 2016).

Sex determination in poplar and evidence for a path to dioecy via monoecy

Pretty much like a toggle switch in an electrical circuit, single-gene systems for sex determination have a simple design: they either fully turn on or off. To determine sex, the switch ensures continuity of a signaling pathway for the development of one sex

while blocking the other. An important practical difference compared to the two-gene systems is that a sex switch will only shift between female and male development. As demonstrated in **Chapters 2 and 5** the poplar sex switch *ARR17* activates female development and at the same time represses male formation by regulating the expression of B class genes. Nonetheless, there have been reports of sex lability in poplars (Fladung et al., 2019; Lester, 1963; Schlenker, 1953). Additionally, after screening one hundred catkins from 44 different female *P. tremula* clones in the arboretum at the Thünen Institute, three catkins presented a single hermaphroditic flower with the presence of a stamen and a carpel within a flower at the base of the catkin (**Figure 8**).

FIGURE 8. HERMAPHRODITIC *P. TREMULA*. IN A *P. TREMULA* CATKIN FROM A CLONE TAKEN FROM THE ARBORETUM AT THE THÜNEN INSTITUTE IN GROßHANSDORF, A STAMEN (MALE SEXUAL ORGAN) CAN BE IDENTIFIED IN A SINGLE FLOWER AT THE BASE OF THE OTHERWISE FEMALE CATKIN.

Considering that *ARR17* is a sex switch and can only be "on" or "off", how can a hermaphrodite possibly happen? Tissue/cell-specific gene expression seems to have major importance. If *UFO* and the B class genes are not completely repressed by *ARR17* but instead are activated in some cells of the floral meristem, the development of flowers with both carpel and stamen may be possible. In contrast, the differential regulation of the expression of the sex switch in different parts of the plant, e.g. by spatiotemporal signals would lead to monoecy. Considering that the stamens were

found at the base of the female catkin, could this distribution represent a manifestation of a monoecious developmental trajectory suggesting potential conservation of the underlying molecular networks and evidence of a pathway to dioecy via monoecy (Renner and Müller 2021)?

As previously mentioned, in single-gene systems the male and female functions are connected via an epistatic genetic interaction, like in monecious species, suggesting that single-gene systems of dioecy might have evolved from monoecy (Renner, 2016; **Leite Montalvão et al., 2021**). In monoecious plants, hormones such as jasmonic acid or ethylene play a major role to control unisexual flower development in different parts of the individual (Boualem et al., 2008; Acosta et al., 2009). Spatially variable levels of hormones within a single plant can thus explain the evolution of monoecy (Golenberg & West 2013). In the monoecious persimmon, gene network analysis suggested that some cytokinin, auxin, and gibberellin signaling genes function cooperatively during gynoecium differentiation (Yang et al., 2019). Even though a direct connection in *Populus* between *ARR17* and cytokinin tested in **Chapter 5** was not established, the dioecious *Populus* species could have originated from a monoecious ancestor which may have had a base-to-apex cytokinin gradient (Renner and Müller 2021), activating *ARR17* and linked female development only in apical parts of the plant. Thus, investigating hermaphroditic and monoecious poplar relatives could shed light on whether the function of *ARR17* in the aspens could turn out to be conserved in a monoecious relative.

Transitions between sexual systems in the Salicaceae

The *ARR17* gene is the only sex-associated homologous sequence found in the sex chromosomes in several reported species in the Salicaceae family (e.g. *P. euphratica*, *P. alba*, *P. trichocarpa*, *P. deltoides*, *P. tremula*, and *Salix purpurea* L.) (Almeida et al., 2020; Geraldes et al., 2015; **Müller et al., 2020**; Xue et al., 2020; Yang et al., 2021; Zhou et al., 2020). Although in *Populus* the ZW system seems to be infrequent, it is, on the other hand, the predominant system found in the species of its sister genera *Salix* studied so far. Similarly to the situation in *P. alba*, several copies of *ARR17* were found in the *S. purpurea* genome (Zhou et al., 2020). Although there are no functional studies yet, recent work found a female-specific region harboring *ARR17* copies in the W

chromosome in *Salix viminalis* L. (Almeida et al., 2020) and *S. purpurea* (Zhou et al., 2018). This raises the intriguing possibility that not only *Populus* but other genera from the Salicaceae family might employ the same evolutionary solution for sex determination.

In line with that, based on preliminary analysis of a draft assembly, one copy of the gene *ARR17* has also been identified in *Poliothyrsis sinensis* Oliv. (data not published), the only monoecious genus closely related to *Populus* (Cronk et al., 2015). This again supports *ARR17* as likely candidate gene for sex determination in the Salicaceae family. In this case, instead of allowing *ARR17* activity only in female individuals, *ARR17* activity would be controlled along the inflorescence axis, perhaps with the additional layer of a hormonal gradient mediated by cytokinin.

Future research focusing on the molecular evolution of plant sexual systems by exploring the transitions between hermaphroditism, monoecy and dioecy in the Salicaceae will verify whether the function of *ARR17* has remained conserved across these different sexual systems. This will also help to understand when *ARR17* may have acquired its feminizing function (i.e. before the evolution to dioecy).

Conclusions and perspectives

The evolution of the diversity in sexual systems in angiosperms is related to the evolution of genetic regulatory networks. The few dioecious systems in which sex determination is well-documented already provides important insights in the diversity of sex determination systems, even though only a small number of pathways are known so far. Although there are indications that similar sex-determining genes and pathways are being used, further studies in more systems are necessary. Additionally, the patterns of expression of sex-determining genes and the phenotype of the homeotic mutations in flower meristem seem to be key to understand the way these genes direct sex expression.

Combined, the results from the five chapters of this thesis explain and highlight how the poplar *ARR17* gene acts as a feminizing factor and no additional antagonistic factor is necessary for sex determination in poplars. Besides, transcriptomic analysis demonstrated a highly targeted function of *ARR17* on specifying floral organ identity,

controlling sex determination in a switch-like manner. Future studies are needed to clarify the open questions regarding the poplar ARR17. The timing of the evolution of *ARR17* function as a switch and the molecular mechanism by which *ARR17* controls dioecy by repressing *UFO* are of particular interest. Additionally, since *Populus* and *Salix* seem to share the underlying mechanism of sex determination in both XY and ZW systems, mediated by *ARR17,* future research in functional genomics is needed to further support the possibility that *ARR17* might be conserved across the Salicaceae family, in monoecious and dioecious species alike.

To fill these knowledge gaps and considering that several dioecious species are among crop species with economic value, the path moving forward is to continue working on more systems, including non-dioecious ones, to elucidate the biological reasons for the different evolutionary dynamics. As the genomics technologies improve, high-quality genome assemblies, pangenomes, gene co-expression networks and gene editing methods are rapidly becoming available. It is likely that in the next few years the sex determination systems in several species will be resolved and geneticists, botanists and breeders will all have a better understanding of the molecular control leading to the evolution of dioecy.

References

Akagi, T., Charlesworth, D., 2019. Pleiotropic effects of sex-determining genes in the evolution of dioecy in two plant species. Proc. R. Soc. B Biol. Sci. 286, 20191805. https://doi.org/10.1098/rspb.2019.1805

Akagi, T., Henry, I.M., Kawai, T., Comai, L., Tao, R., 2016. Epigenetic regulation of the sex determination gene *MeGI* in polyploid persimmon. Plant Cell 28, 2905–2915. https://doi.org/10.1105/tpc.16.00532

Akagi, T., Henry, I.M., Tao, R., Comai, L., 2014. A Y-chromosome–encoded small RNA acts as a sex determinant in persimmons. Science 346, 646–650. https://doi.org/10.1126/science.1257225

Akagi, T., Pilkington, S.M., Varkonyi-Gasic, E., Henry, I.M., Sugano, S.S., Sonoda, M., Firl, A., McNeilage, M.A., Douglas, M.J., Wang, T., Rebstock, R., Voogd, C., Datson, P., Allan, A.C., Beppu, K., Kataoka, I., Tao, R., 2019. Two Y-chromosome-encoded genes determine sex in kiwifruit. Nat. Plants 5, 801–809. https://doi.org/10.1038/s41477-019-0489-6

Akagi, T., Shirasawa, K., Nagasaki, H., Hirakawa, H., Tao, R., Comai, L., Henry, I.M., 2020. The persimmon genome reveals clues to the evolution of a lineage-specific sex determination system in plants. PLOS Genet. 16, e1008566. https://doi.org/10.1371/journal.pgen.1008566

Ali, H.B.M., Abubakari, A., Wiehle, M., Krutovsky, K.V., 2018. Gene-specific sex-linked genetic markers in date palm (*Phoenix dactylifera* L.). Genet. Resour. Crop Evol. 65, 1–10. https://doi.org/10.1007/s10722-017-0564-7

Allen, C.E., 1917. A chromosome difference correlated with sex differences in Sphærocarpos. Science. https://doi.org/10.1126/science.46.1193.466

Almeida, P., Proux-Wera, E., Churcher, A., Soler, L., Dainat, J., Pucholt, P., Nordlund, J., Martin, T., Rönnberg-Wästljung, A.-C., Nystedt, B., Berlin, S., Mank, J.E., 2020. Genome assembly of the basket willow, *Salix viminalis*, reveals earliest stages of sex chromosome expansion. BMC Biol. 18. https://doi.org/10.1186/s12915-020-00808-1

Amarasinghe, S.L., Su, S., Dong, X., Zappia, L., Ritchie, M.E., Gouil, Q., 2020. Opportunities and challenges in long-read sequencing data analysis. Genome Biol. 21, 30. https://doi.org/10.1186/s13059-020-1935-5

Ashman, T.-L., Bachtrog, D., Blackmon, H., Goldberg, E.E., Hahn, M.W., Kirkpatrick, M., Kitano, J., Mank, J.E., Mayrose, I., Ming, R., Otto, S.P., Peichel, C.L., Pennell, M.W., Perrin, N., Ross, L., Valenzuela, N., Vamosi, J.C., The Tree of Sex Consortium, 2014. Tree of Sex: A database of sexual systems. Sci. Data 1, 140015. https://doi.org/10.1038/sdata.2014.15

Bachtrog, D., Mank, J.E., Peichel, C.L., Kirkpatrick, M., Otto, S.P., Ashman, T.-L., Hahn, M.W., Kitano, J., Mayrose, I., Ming, R., Perrin, N., Ross, L., Valenzuela, N., Vamosi, J.C., Consortium, T.T. of S., 2014. Sex Determination: Why so many ways of doing It? PLOS Biol. 12, e1001899. https://doi.org/10.1371/journal.pbio.1001899

Becker, A., Ehlers, K., 2016. *Arabidopsis* flower development of protein complexes, targets, and transport. Protoplasma 253, 219–230. https://doi.org/10.1007/s00709-015-0812-7

Bergero, R., Qiu, S., Charlesworth, D., 2015. Gene loss from a plant sex chromosome system. Curr. Biol. 25, 1234–1240. https://doi.org/10.1016/j.cub.2015.03.015

Blázquez, M., Koornneef, M., Putterill, J., 2001. Flowering on time: genes that regulate the floral transition. EMBO Rep. 2, 1078–1082. https://doi.org/10.1093/embo-reports/kve254

Boes, T.K., Strauss, S.H., 1994. Floral phenology and morphology of Black Cottonwood, *Populus trichocarpa* (Salicaceae). Am. J. Bot. 81, 562. https://doi.org/10.2307/2445730

Boualem, A., Troadec, C., Camps, C., Lemhemdi, A., Morin, H., Sari, M.-A., Fraenkel-Zagouri, R., Kovalski, I., Dogimont, C., Perl-Treves, R., Bendahmane, A., 2015. A cucurbit androecy gene reveals how unisexual flowers develop and dioecy emerges. Science 350, 688–691. https://doi.org/10.1126/science.aac8370

Bradshaw, H.D., Ceulemans, R., Davis, J., Stettler, R., 2000. Emerging model systems in plant biology: Poplar (*Populus*) as a model forest tree. J. Plant Growth Regul. 19, 306–313. https://doi.org/10.1007/s003440000030

Bull, J., J., 1983. Evolution of sex determining mechanisms. The Benjamin/Cummings Publishing Company, Inc.

Carey, S., Yu, Q., Harkess, A., 2021. The diversity of plant sex chromosomes highlighted through advances in genome sequencing. Genes 12, 381. https://doi.org/10.3390/genes12030381

Chae, T., Harkess, A., Moore, R.C., 2021. Sex-linked gene expression and the emergence of hermaphrodites in *Carica papaya*. Am. J. Bot. 108, 1029–1041. https://doi.org/10.1002/ajb2.1689

Charlesworth, B., Charlesworth, D., 1978. A model for the evolution of dioecy and gynodioecy. Am. Nat. 112, 975–997. https://doi.org/10.1086/283342
Charlesworth, B., Harvey, P.H., Charlesworth, Brian, Charlesworth, D., 2000. The degeneration of Y chromosomes. Philos. Trans. R. Soc. Lond. B. Biol. Sci. 355, 1563–1572. https://doi.org/10.1098/rstb.2000.0717
Charlesworth, D., 2021. When and how do sex-linked regions become sex chromosomes? Evolution 75, 569–581. https://doi.org/10.1111/evo.14196
Charlesworth, D., 2016. Plant sex chromosomes. Annu. Rev. Plant Biol. 67, 397–420. https://doi.org/10.1146/annurev-arplant-043015-111911
Charlesworth, D., 2013. Plant sex chromosome evolution. J. Exp. Bot. 64, 405–420. https://doi.org/10.1093/jxb/ers322
Charlesworth, B., 1991. The evolution of sex chromosomes. Science. 251, 1030-1033. https://doi.org/10.1126/science.1998119
Charlesworth, D., Charlesworth, B., Marais, G., 2005. Steps in the evolution of heteromorphic sex chromosomes. Heredity 95, 118–128. https://doi.org/10.1038/sj.hdy.6800697
Charlesworth, D., Mank, J.E., 2010. 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. https://doi.org/10.1534/genetics.110.117697
Coen, E.S., Meyerowitz, E.M., 1991. The war of the whorls: genetic interactions controlling flower development. Nature 353, 31–37. https://doi.org/10.1038/353031a0
Cossard, G.G., Gerchen, J.F., Li, X., Cuenot, Y., Pannell, J.R., 2019. Rapid dissolution of dioecy by experimental evolution. Curr. Biol. 31, 1277-1283. https://doi.org/10.1101/712414
Cronk, Q., Müller, N.A., 2020. Default sex and single gene sex determination in dioecious plants. Front. Plant Sci. 11, 1162. https://doi.org/10.3389/fpls.2020.01162
Cronk, Q.C.B., 2005. Plant eco-devo: the potential of poplar as a model organism. New Phytol. 166, 39–48. https://doi.org/10.1111/j.1469-8137.2005.01369.x
Cronk, Q.C.B., Needham, I., Rudall, P.J., 2015. Evolution of catkins: inflorescence morphology of selected Salicaceae in an evolutionary and developmental context. Front. Plant Sci. 6, 1030. https://doi.org/10.3389/fpls.2015.01030
Darwin, C., 1877. The different forms of flowers on plants of the same species. John Murray, London (UK).
Diggle, P.K., Stilio, V.S.D., Gschwend, A.R., Golenberg, E.M., Moore, R.C., Russell, J.R.W., Sinclair, J.P., 2011. Multiple developmental processes underlie sex differentiation in angiosperms. Trends Genet. 27, 368–376. https://doi.org/10.1016/j.tig.2011.05.003
Ditta, G., Pinyopich, A., Robles, P., Pelaz, S., Yanofsky, M.F., 2004. The *SEP4* gene of *Arabidopsis thaliana* functions in floral organ and meristem identity. Curr. Biol. 14, 1935–1940. https://doi.org/10.1016/j.cub.2004.10.028
Fechter, I., Hausmann, L., Daum, M., Rosleff Sörensen, T., Viehöver, P., Weisshaar, B., Töpfer, R., 2012. Candidate genes within a 143 kb region of the flower sex locus in *Vitis*. Mol. Genet. Genomics 287, 247–259. https://doi.org/10.1007/s00438-012-0674-z
Feng, G., Sanderson, B.J., Keefover-Ring, K., Liu, J., Ma, T., Yin, T., Smart, L.B., DiFazio, S.P., Olson, M.S., 2020. Pathways to sex determination in plants: how many roads lead to Rome? Curr. Opin. Plant Biol., Genome studies and molecular genetics 54, 61–68. https://doi.org/10.1016/j.pbi.2020.01.004
Ferrándiz, C., Gu, Q., Martienssen, R., Yanofsky, M.F., 2000. Redundant regulation of meristem identity and plant architecture by *FRUITFULL*, APETALA1 and *CAULIFLOWER*. Dev. Camb. Engl. 127, 725–734. https://doi.org/10.1242/dev.127.4.725
Filatov, D.A., 2015. Homomorphic plant sex chromosomes are coming of age. Mol. Ecol. 24, 3217–3219. https://doi.org/10.1111/mec.13268
Fladung, M., Kumar, S., Raj Ahuja, M., 1997. Genetic transformation of *Populus* genotypes with different chimaeric gene constructs: transformation efficiency and molecular analysis. Transgenic Res. 6, 111–121. https://doi.org/10.1023/A:1018421620040
Fladung, M., Schildbach, M., Hoenicka, H., Kersten, B., Müller, N.A., 2019. Selfing of a single monoecious *Populus tremula* tree produces viable males, females and "supermales." Trees 33, 803–816. https://doi.org/10.1007/s00468-019-01817-6
Furman, B.L.S., Metzger, D.C.H., Darolti, I., Wright, A.E., Sandkam, B.A., Almeida, P., Shu, J.J., Mank, J.E., 2020. Sex chromosome evolution: so many exceptions to the rules. Genome Biol. Evol. 12, 750–763. https://doi.org/10.1093/gbe/evaa081

Gamble, T., Zarkower, D., 2014. Identification of sex-specific molecular markers using restriction site-associated DNA sequencing. Mol. Ecol. Resour. 14, 902–913. https://doi.org/10.1111/1755-0998.12237

Gaudet, M., Jorge, V., Paolucci, I., Beritognolo, I., Mugnozza, G.S., Sabatti, M., 2007. Genetic linkage maps of *Populus nigra* L. including AFLPs, SSRs, SNPs, and sex trait. Tree Genet. Genomes 4, 25–36. https://doi.org/10.1007/s11295-007-0085-1

Geraldes, A., Hefer, C.A., Capron, A., Kolosova, N., Martinez-Nuñez, F., Soolanayakanahally, R.Y., Stanton, B., Guy, R.D., Mansfield, S.D., Douglas, C.J., Cronk, Q.C.B., 2015. Recent Y chromosome divergence despite ancient origin of dioecy in poplars (*Populus*). Mol. Ecol. 24, 3243–3256. https://doi.org/10.1111/mec.13126

Golenberg, E.M., West, N.W., 2013. Hormonal interactions and gene regulation can link monoecy and environmental plasticity to the evolution of dioecy in plants. Am. J. Bot. 100, 1022–1037. https://doi.org/10.3732/ajb.1200544

Gramzow, L., Theissen, G., 2010. A hitchhiker's guide to the MADS world of plants. Genome Biol. 11, 214. https://doi.org/10.1186/gb-2010-11-6-214

Gschwend, A.R., Yu, Q., Tong, E.J., Zeng, F., Han, J., VanBuren, R., Aryal, R., Charlesworth, D., Moore, P.H., Paterson, A.H., Ming, R., 2012. Rapid divergence and expansion of the X chromosome in papaya. Proc. Natl. Acad. Sci. 109, 13716–13721. https://doi.org/10.1073/pnas.1121096109

Gubbay, J., Collignon, J., Koopman, P., Capel, B., Economou, A., Münsterberg, A., Vivian, N., Goodfellow, P., Lovell-Badge, R., 1990. A gene mapping to the sex-determining region of the mouse Y chromosome is a member of a novel family of embryonically expressed genes. Nature 346, 245–250. https://doi.org/10.1038/346245a0

Harkess, A., Huang, K., van der Hulst, R., Tissen, B., Caplan, J.L., Koppula, A., Batish, M., Meyers, B.C., Leebens-Mack, J., 2020. Sex determination by two Y-linked genes in garden asparagus. Plant Cell 32, 1790–1796. https://doi.org/10.1105/tpc.19.00859

Harkess, A., Zhou, J., Xu, C., Bowers, J.E., Van der Hulst, R., Ayyampalayam, S., Mercati, F., Riccardi, P., McKain, M.R., Kakrana, A., Tang, H., Ray, J., Groenendijk, J., Arikit, S., Mathioni, S.M., Nakano, M., Shan, H., Telgmann-Rauber, A., Kanno, A., Yue, Z., Chen, Haixin, Li, W., Chen, Y., Xu, X., Zhang, Y., Luo, S., Chen, Helong, Gao, J., Mao, Z., Pires, J.C., Luo, M., Kudrna, D., Wing, R.A., Meyers, B.C., Yi, K., Kong, H., Lavrijsen, P., Sunseri, F., Falavigna, A., Ye, Y., Leebens-Mack, J.H., Chen, G., 2017. The asparagus genome sheds light on the origin and evolution of a young Y chromosome. Nat. Commun. 8, 1279. https://doi.org/10.1038/s41467-017-01064-8

Henry, I.M., Akagi, T., Tao, R., Comai, L., 2018. One hundred ways to invent the sexes: theoretical and observed paths to dioecy in plants. Annu. Rev. Plant Biol. 69, 553–575. https://doi.org/10.1146/annurev-arplant-042817-040615

Hobza, R., Hudzieczek, V., Kubat, Z., Cegan, R., Vyskot, B., Kejnovsky, E., Janousek, B., 2018. Sex and the flower – developmental aspects of sex chromosome evolution. Ann. Bot. 122, 1085–1101. https://doi.org/10.1093/aob/mcy130

Hoenicka, H., Lehnhardt, D., Briones, V., Nilsson, O., Fladung, M., 2016. Low temperatures are required to induce the development of fertile flowers in transgenic male and female early flowering poplar (*Populus tremula* L.). Tree Physiol. 36, 667–677. https://doi.org/10.1093/treephys/tpw015

Hoenicka, H., Lehnhardt, D., Polak, O., Fladung, M., 2012. Early flowering and genetic containment studies in transgenic poplar. IForest - Biogeosciences For. 5, 138. https://doi.org/10.3832/ifor0621-005

Honma, T., Goto, K., 2000. The *Arabidopsis* floral homeotic gene *PISTILLATA* is regulated by discrete cis-elements responsive to induction and maintenance signals. Dev. Camb. Engl. 127, 2021–2030. https://doi.org/10.1242/dev.127.10.2021

Hou, J., Wei, S., Pan, H., Zhuge, Q., Yin, T., 2019. Uneven selection pressure accelerating divergence of *Populus* and *Salix*. Hortic. Res. 6, 37. https://doi.org/10.1038/s41438-019-0121-y

Jiao, W.-B., Schneeberger, K., 2017. The impact of third generation genomic technologies on plant genome assembly. Curr. Opin. Plant Biol. 36, 64–70. https://doi.org/10.1016/j.pbi.2017.02.002

Jones, D.F., 1934. Unisexual maize plants and their bearing on sex differentiation in other plants and animals. Genetics 19, 552–567. https://doi.org/10.1093/genetics/19.6.552

Käfer, J., Marais, G.A.B., Pannell, J.R., 2017. On the rarity of dioecy in flowering plants. Mol. Ecol. 26, 1225–1241. https://doi.org/10.1111/mec.14020

Kardailsky, I., Shukla, V.K., Ahn, J.H., Dagenais, N., Christensen, S.K., Nguyen, J.T., Chory, J., Harrison, M.J., Weigel, D., 1999. Activation tagging of the floral inducer *FT*. Science 286, 1962–1965. https://doi.org/10.1126/science.286.5446.1962

Kashimada, K., Koopman, P., 2010. SRY: the master switch in mammalian sex determination. Dev. Camb. Engl. 137, 3921–3930. https://doi.org/10.1242/dev.048983
Kaufmann, K., Melzer, R., Theißen, G., 2005. MIKC-type MADS-domain proteins: structural modularity, protein interactions and network evolution in land plants. Gene 347, 183–198. https://doi.org/10.1016/j.gene.2004.12.014
Kersten, B., Leite Montalvão, A.P., Hoenicka, H., Vettori, C., Paffetti, D., Fladung, M., 2020. Sequencing of two transgenic early-flowering poplar lines confirmed vector-free single-locus T-DNA integration. Transgenic Res. 29, 321–337. https://doi.org/10.1007/s11248-020-00203-0
Kersten, B., Pakull, B., Fladung, M., 2017. Genomics of sex determination in dioecious trees and woody plants. Trees 31, 1113–1125. https://doi.org/10.1007/s00468-017-1525-7
Kersten, B., Pakull, B., Groppe, K., Lueneburg, J., Fladung, M., 2014. The sex-linked region in *Populus tremuloides* Turesson 141 corresponds to a pericentromeric region of about two million base pairs on *P. trichocarpa* chromosome 19. Plant Biol. 16, 411–418. https://doi.org/10.1111/plb.12048
Kieber, J., J., Schaller, G.E., 2018. Cytokinin signaling in plant development. Development 145. https://doi.org/10.1242/dev.149344
Kim, G., Leite Montalvão, A.P., Kersten, B., Fladung, M., Müller, N.A., 2021. The genetic basis of sex determination in *Populus* provides molecular markers across the genus and indicates convergent evolution. Silvae Genet. 70, 145–155. https://doi.org/10.2478/sg-2021-0012
Koornneef, M., Hanhart, C.J., Van der Veen, J.H., 1991. A genetic and physiological analysis of late flowering mutants in *Arabidopsis thaliana*. Molec. Gen. Genet. 229, 57–66 https://doi.org/10.1007/BF00264213
Krizek, B.A., Fletcher, J.C., 2005. Molecular mechanisms of flower development: an armchair guide. Nat. Rev. Genet. 6, 688–698. https://doi.org/10.1038/nrg1675
Krizek, B.A., Meyerowitz, E.M., 1996. The Arabidopsis homeotic genes *APETALA3* and *PISTILLATA* are sufficient to provide the B class organ identity function. Dev. Camb. Engl. 122, 11–22. PMID: 8565821.
Kumar, S., Fladung, M., 2001. Gene stability in transgenic aspen (*Populus*). II. Molecular characterization of variable expression of transgene in wild and hybrid aspen. Planta 213, 731–740. https://doi.org/10.1007/s004250100535
Lai, Y.-S., Shen, D., Zhang, W., Zhang, X., Qiu, Y., Wang, H., Dou, X., Li, S., Wu, Y., Song, J., Ji, G., Li, X., 2018. Temperature and photoperiod changes affect cucumber sex expression by different epigenetic regulations. BMC Plant Biol. 18, 268. https://doi.org/10.1186/s12870-018-1490-3
Leite Montalvão, A.P., Kersten, B., Fladung, M., Müller, N.A., 2021. The diversity and dynamics of sex determination in dioecious plants. Front. Plant Sci. 11, 1-12. https://doi.org/10.3389/fpls.2020.580488
Lester, D.T., 1963. Variation in sex expression in *Populus tremuloides* Michx. Silvae Genet.12, 141-151.
Li, D., Sheng, Y., Niu, H., Li, Z., 2019. Gene interactions regulating sex determination in cucurbits. Front. Plant Sci. 10, 1231. https://doi.org/10.3389/fpls.2019.01231
Linnaenus, C. von, 1758. *Systema naturae per regna tria naturae: secundum classes, ordines, genera, species, cum characteribus, differentiis, synonymis, locis*, 10th ed. Laurentius Salvius, Stockholm.
Mallo, M., Wellik, D.M., Deschamps, J., 2010. Hox genes and regional patterning of the vertebrate body plan. Dev. Biol. 344, 7–15. https://doi.org/10.1016/j.ydbio.2010.04.024
Mao, Y., Liu, W., Chen, X., Xu, Y., Lu, W., Hou, J., Ni, J., Wang, Y., Wu, L., 2017. Flower development and sex determination between male and female flowers in *Vernicia fordii*. Front. Plant Sci. 8, 1291. https://doi.org/10.3389/fpls.2017.01291
McKown, A.D., Klápště, J., Guy, R.D., Soolanayakanahally, R.Y., La Mantia, J., Porth, I., Skyba, O., Unda, F., Douglas, C.J., El-Kassaby, Y.A., Hamelin, R.C., Mansfield, S.D., Cronk, Q.C.B., 2017. Sexual homomorphism in dioecious trees: extensive tests fail to detect sexual dimorphism in *Populus*. Sci. Rep. 7, 1831. https://doi.org/10.1038/s41598-017-01893-z
Melzer, S., Lens, F., Gennen, J., Vanneste, S., Rohde, A., Beeckman, T., 2008. Flowering-time genes modulate meristem determinacy and growth form in *Arabidopsis thaliana*. Nat. Genet. 40, 1489–1492. https://doi.org/10.1038/ng.253
Milewicz, M., Sawicki, J., 2013. Sex-linked markers in dioecious plants. Plant Omics J. 144–149.
Ming, R., Bendahmane, A., Renner, S.S., 2011. Sex Chromosomes in Land Plants. Annu. Rev. Plant Biol. 62, 485–514. https://doi.org/10.1146/annurev-arplant-042110-103914
Müller, N.A., Kersten, B., Leite Montalvão, A.P., Mähler, N., Bernhardsson, C., Bräutigam, K., Carracedo Lorenzo, Z., Hoenicka, H., Kumar, V., Mader, M., Pakull, B., Robinson, K.M., Sabatti, M., Vettori, C., Ingvarsson, P.K., Cronk, Q., Street, N.R., Fladung, M., 2020. A single gene underlies the

dynamic evolution of poplar sex determination. Nat. Plants 6, 630–637. https://doi.org/10.1038/s41477-020-0672-9
Murai, K., 2013. Homeotic genes and the ABCDE model for floral organ formation in wheat. Plants 2, 379–395. https://doi.org/10.3390/plants2030379
Muskens, M.W., Vissers, A.P., Mol, J.N., Kooter, J.M., 2000. Role of inverted DNA repeats in transcriptional and post-transcriptional gene silencing. Plant Mol. Biol. 43, 243–260. https://doi.org/10.1023/a:1006491613768
Nurk, S., Walenz, B.P., Rhie, A., Vollger, M.R., Logsdon, G.A., Grothe, R., Miga, K.H., Eichler, E.E., Phillippy, A.M., Koren, S., 2020. HiCanu: accurate assembly of segmental duplications, satellites, and allelic variants from high-fidelity long reads. Genome Res. 30, 1291–1305. https://doi.org/10.1101/gr.263566.120
Pakull, B., Groppe, K., Mecucci, F., Gaudet, M., Sabatti, M., Fladung, M., 2011. Genetic mapping of linkage group XIX and identification of sex-linked SSR markers in a *Populus tremula* x *Populus tremuloides*, cross. 41, 2. https://doi.org/10.1139/X10-206
Pakull, B., Groppe, K., Meyer, M., Markussen, T., Fladung, M., 2009. Genetic linkage mapping in aspen (*Populus tremula* L. and *Populus tremuloides* Michx.). Tree Genet. Genomes 5, 505–515. https://doi.org/10.1007/s11295-009-0204-2
Pakull, B., Kersten, B., Lüneburg, J., Fladung, M., 2015. A simple PCR-based marker to determine sex in aspen. Plant Biol. 17, 256–261. https://doi.org/10.1111/plb.12217
Paolucci, I., Gaudet, M., Jorge, V., Beritognolo, I., Terzoli, S., Kuzminsky, E., Muleo, R., Scarascia Mugnozza, G., Sabatti, M., 2010. Genetic linkage maps of *Populus alba* L. and comparative mapping analysis of sex determination across *Populus* species. Tree Genet. Genomes 6, 863–875. https://doi.org/10.1007/s11295-010-0297-7
Papadopulos, A.S.T., Chester, M., Ridout, K., Filatov, D.A., 2015. Rapid Y degeneration and dosage compensation in plant sex chromosomes. Proc. Natl. Acad. Sci. 112, 13021–13026. https://doi.org/10.1073/pnas.1508454112
Parcy, F., Nilsson, O., Busch, M.A., Lee, I., Weigel, D., 1998. A genetic framework for floral patterning. Nature 395, 561–566. https://doi.org/10.1038/26903
Pelaz, S., Tapia-López, R., Alvarez-Buylla, E.R., Yanofsky, M.F., 2001. Conversion of leaves into petals in Arabidopsis. Curr. Biol. CB 11, 182–184. https://doi.org/10.1016/s0960-9822(01)00024-0
Pfent, C., Pobursky, K.J., Sather, D.N., Golenberg, E.M., 2005. Characterization of *SpAPETALA3* and *SpPISTILLATA*, B class floral identity genes in *Spinacia oleracea*, and their relationship to sexual dimorphism. Dev. Genes Evol. 215, 132–142. https://doi.org/10.1007/s00427-004-0459-4
Picq, S., Santoni, S., Lacombe, T., Latreille, M., Weber, A., Ardisson, M., Ivorra, S., Maghradze, D., Arroyo-Garcia, R., Chatelet, P., This, P., Terral, J.-F., Bacilieri, R., 2014. A small XY chromosomal region explains sex determination in wild dioecious *V. vinifera* and the reversal to hermaphroditism in domesticated grapevines. BMC Plant Biol. 14, 229. https://doi.org/10.1186/s12870-014-0229-z
Pidkowich, M.S., Klenz, J.E., Haughn, G.W., 1999. The making of a flower: control of floral meristem identity in *Arabidopsis*. Trends Plant Sci. 4, 64–70. https://doi.org/10.1016/S1360-1385(98)01369-7
Pin, P.A., Nilsson, O., 2012. The multifaceted roles of *FLOWERING LOCUS T* in plant development. Plant Cell Environ. 35, 1742–1755. https://doi.org/10.1111/j.1365-3040.2012.02558.x
Prentout, D., Razumova, O., Rhoné, B., Badouin, H., Henri, H., Feng, C., Käfer, J., Karlov, G., Marais, G.A.B., 2020. An efficient RNA-seq-based segregation analysis identifies the sex chromosomes of *Cannabis sativa*. Genome Res. 30, 164–172. https://doi.org/10.1101/gr.251207.119
Pucholt, P., Wright, A.E., Conze, L.L., Mank, J.E., Berlin, S., 2017. Recent sex chromosome divergence despite ancient dioecy in the willow *Salix viminalis*. Mol. Biol. Evol. 34, 1991–2001. https://doi.org/10.1093/molbev/msx144
Qian, W., Fan, G., Liu, D., Zhang, H., Wang, X., Wu, J., Xu, Z., 2017. Construction of a high-density genetic map and the X/Y sex-determining gene mapping in spinach based on large-scale markers developed by specific-locus amplified fragment sequencing (SLAF-seq). BMC Genomics 18, 276. https://doi.org/10.1186/s12864-017-3659-9
Renner, S.S., 2016. Pathways for making unisexual flowers and unisexual plants: moving beyond the "two mutations linked on one chromosome" model. Am. J. Bot. 103, 587–589. https://doi.org/10.3732/ajb.1600029
Renner, S.S., 2014. The relative and absolute frequencies of angiosperm sexual systems: Dioecy, monoecy, gynodioecy, and an updated online database. Am. J. Bot. 101, 1588–1596. https://doi.org/10.3732/ajb.1400196

Renner, S.S., Müller, N.A., 2021. Plant sex chromosomes defy evolutionary models of expanding recombination suppression and genetic degeneration. Nat. Plants 7, 392–402. https://doi.org/10.1038/s41477-021-00884-3

Renner, S.S., Ricklefs, R.E., 1995. Dioecy and its correlates in the flowering plants. Am. J. Bot. 82, 596–606. https://doi.org/10.1002/j.1537-2197.1995.tb11504.x

Robinson, K.M., Delhomme, N., Mähler, N., Schiffthaler, B., Önskog, J., Albrectsen, B.R., Ingvarsson, P.K., Hvidsten, T.R., Jansson, S., Street, N.R., 2014. *Populus tremula* (European aspen) shows no evidence of sexual dimorphism. BMC Plant Biol. 14, 276. https://doi.org/10.1186/s12870-014-0276-5

Sather, D.N., Jovanovic, M., Golenberg, E.M., 2010. Functional analysis of B and C class floral organ genes in spinach demonstrates their role in sexual dimorphism. BMC Plant Biol. 10, 46. https://doi.org/10.1186/1471-2229-10-46

Schlenker, G., 1953. Beobachtungen über die Geschlechtsverhältnisse bei jungen Graupappeln und Aspen. Zeitschr Forstgenet. 2, 102–110.

Slavov, G.T., Zhelev, P., 2010. Salient biological features, systematics, and genetic variation of *Populus*, in: Jansson, S., Bhalerao, R., Groover, A. (Eds.), Genetics and Genomics of *Populus*, Plant Genetics and Genomics: Crops and Models. Springer, New York, NY, pp. 15–38. https://doi.org/10.1007/978-1-4419-1541-2_2

Sousa, A., Fuchs, J., Renner, S.S., 2013. Molecular cytogenetics (FISH, GISH) of *Coccinia grandis*: A ca. 3 myr-old species of Cucurbitaceae with the largest Y/autosome divergence in flowering plants. Cytogenet. Genome Res. 139, 107–118. https://doi.org/10.1159/000345370

Stapley, J., Feulner, P.G.D., Johnston, S.E., Santure, A.W., Smadja, C.M., 2017. Variation in recombination frequency and distribution across eukaryotes: patterns and processes. Philos. Trans. R. Soc. Lond. B. Biol. Sci. 372, 20160455. https://doi.org/10.1098/rstb.2016.0455

Stern, D.L., 2013. The genetic causes of convergent evolution. Nat. Rev. Genet. 14, 751–764. https://doi.org/10.1038/nrg3483

Stevens, N.M., 1905. Studies in spermatogenesis. Carnegie Institution of Washington.

Strauss, S.H., Martin, F.M., 2004. Poplar genomics comes of age. New Phytol. 164, 1–4. https://doi.org/10.1111/j.1469-8137.2004.01179.x

Tamiru, M., Natsume, S., Takagi, Hiroki, White, B., Yaegashi, H., Shimizu, M., Yoshida, K., Uemura, A., Oikawa, K., Abe, A., Urasaki, N., Matsumura, H., Babil, P., Yamanaka, S., Matsumoto, R., Muranaka, S., Girma, G., Lopez-Montes, A., Gedil, M., Bhattacharjee, R., Abberton, M., Kumar, P.L., Rabbi, I., Tsujimura, M., Terachi, T., Haerty, W., Corpas, M., Kamoun, S., Kahl, G., Takagi, Hiroko, Asiedu, R., Terauchi, R., 2017. Genome sequencing of the staple food crop white Guinea yam enables the development of a molecular marker for sex determination. BMC Biol. 15, 86. https://doi.org/10.1186/s12915-017-0419-x

Tennessen, J.A., Wei, N., Straub, S.C.K., Govindarajulu, R., Liston, A., Ashman, T.L. (2018). Repeated translocation of a gene cassette drives sex-chromosome turnover in strawberries. *PLoS Biology*. 16 (8): e2006062. https://doi.org/10.1371/journal.pbio.2006062

To, J.P.C., Haberer, G., Ferreira, F.J., Deruère, J., Mason, M.G., Schaller, G.E., Alonso, J.M., Ecker, J.R., Kieber, J.J., 2004. Type-A Arabidopsis response regulators are partially redundant negative regulators of cytokinin signaling. Plant Cell 16, 658–671. https://doi.org/10.1105/tpc.018978

Todesco, M., Owens, G.L., Bercovich, N., Légaré, J.-S., Soudi, S., Burge, D.O., Huang, K., Ostevik, K.L., Drummond, E.B.M., Imerovski, I., Lande, K., Pascual-Robles, M.A., Nanavati, M., Jahani, M., Cheung, W., Staton, S.E., Muños, S., Nielsen, R., Donovan, L.A., Burke, J.M., Yeaman, S., Rieseberg, L.H., 2020. Massive haplotypes underlie ecotypic differentiation in sunflowers. Nature 584, 602–607. https://doi.org/10.1038/s41586-020-2467-6

Tuskan, G.A., DiFazio, S., Jansson, S., Bohlmann, J., Grigoriev, I., Hellsten, U., Putnam, N., Ralph, S., Rombauts, S., Salamov, A., Schein, J., Sterck, L., Aerts, A., Bhalerao, R.R., Bhalerao, R.P., Blaudez, D., Boerjan, W., Brun, A., Brunner, A., Busov, V., Campbell, M., Carlson, J., Chalot, M., Chapman, J., Chen, G.-L., Cooper, D., Coutinho, P.M., Couturier, J., Covert, S., Cronk, Q., Cunningham, R., Davis, J., Degroeve, S., Déjardin, A., dePamphilis, C., Detter, J., Dirks, B., Dubchak, I., Duplessis, S., Ehlting, J., Ellis, B., Gendler, K., Goodstein, D., Gribskov, M., Grimwood, J., Groover, A., Gunter, L., Hamberger, B., Heinze, B., Helariutta, Y., Henrissat, B., Holligan, D., Holt, R., Huang, W., Islam-Faridi, N., Jones, S., Jones-Rhoades, M., Jorgensen, R., Joshi, C., Kangasjärvi, J., Karlsson, J., Kelleher, C., Kirkpatrick, R., Kirst, M., Kohler, A., Kalluri, U., Larimer, F., Leebens-Mack, J., Leplé, J.-C., Locascio, P., Lou, Y., Lucas, S., Martin, F., Montanini, B., Napoli, C., Nelson, D.R., Nelson, C., Nieminen, K., Nilsson, O., Pereda, V., Peter, G., Philippe, R., Pilate, G., Poliakov, A., Razumovskaya,

J., Richardson, P., Rinaldi, C., Ritland, K., Rouzé, P., Ryaboy, D., Schmutz, J., Schrader, J., Segerman, B., Shin, H., Siddiqui, A., Sterky, F., Terry, A., Tsai, C.-J., Uberbacher, E., Unneberg, P., Vahala, J., Wall, K., Wessler, S., Yang, G., Yin, T., Douglas, C., Marra, M., Sandberg, G., Peer, Y.V. de, Rokhsar, D., 2006. The genome of black cottonwood, *Populus trichocarpa* (Torr. & Gray). Science 313, 1596–1604. https://doi.org/10.1126/science.1128691

Vekemans, X., Castric, V., Hipperson, H., Müller, N.A., Westerdahl, H., Cronk, Q., 2021. Whole-genome sequencing and genome regions of special interest: Lessons from major histocompatibility complex, sex determination, and plant self-incompatibility. Mol. Ecol. n/a, 1–15. https://doi.org/10.1111/mec.16020

Wambui Mbichi, R., Wang, Q.-F., Wan, T., 2020. RNA directed DNA methylation and seed plant genome evolution. Plant Cell Rep. 39, 983–996. https://doi.org/10.1007/s00299-020-02558-4

Weigel, D., Alvarez, J., Smyth, D.R., Yanofsky, M.F., Meyerowitz, E.M., 1992. *LEAFY* controls floral meristem identity in *Arabidopsis*. Cell 69, 843–859. https://doi.org/10.1016/0092-8674(92)90295-n

West, N.W., Golenberg, E.M., 2018. Gender-specific expression of *GIBBERELLIC ACID INSENSITIVE* is critical for unisexual organ initiation in dioecious *Spinacia oleracea*. New Phytol. 217, 1322–1334. https://doi.org/10.1111/nph.14919

Westergaard, M., 1958. The mechanism of sex determination in dioecious flowering plants, in: Demerec, M. (Ed.), Advances in Genetics. Academic Press, pp. 217–281. https://doi.org/10.1016/S0065-2660(08)60163-7

WFO (2022). World Flora Online. Published on the Internet. http://www.worldfloraonline.org. Accessed on: 05 Jan 2022

Wilson, E.B., 1909. Recent researches on the determination and heredity of sex. Science. https://doi.org/10.1126/science.29.732.53

Wilson, E.B., 1905. Studies on chromosomes. I. The behavior of the idiochromosomes in hemiptera. J. Exp. Zool. 2, 371–405. https://doi.org/10.1002/jez.1400020305

Wright, A.E., Dean, R., Zimmer, F., Mank, J.E., 2016. How to make a sex chromosome. Nat. Commun. 7, 12087. https://doi.org/10.1038/ncomms12087

Xue, L., Wu, H., Chen, Y., Li, X., Hou, J., Lu, J., Wei, S., Dai, X., Olson, M.S., Liu, J., Wang, M., Charlesworth, D., Yin, T., 2020. Evidences for a role of two Y-specific genes in sex determination in *Populus deltoides*. Nat. Commun. 11, 5893. https://doi.org/10.1038/s41467-020-19559-2

Yang, H.-W., Akagi, T., Kawakatsu, T., Tao, R., 2019. Gene networks orchestrated by *MeGI*: a single-factor mechanism underlying sex determination in persimmon. Plant J. 98, 97–111. https://doi.org/10.1111/tpj.14202

Yang, W., Wang, D., Li, Y., Zhang, Z., Tong, S., Li, M., Zhang, X., Zhang, L., Ren, L., Ma, X., Zhou, R., Sanderson, B.J., Keefover-Ring, K., Yin, T., Smart, L.B., Liu, J., DiFazio, S.P., Olson, M., Ma, T., 2021. A general model to explain repeated turnovers of sex determination in the Salicaceae. Mol. Biol. Evol. 38, 968–980. https://doi.org/10.1093/molbev/msaa261

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., Tuskan, G.A., 2008. Genome structure and emerging evidence of an incipient sex chromosome in *Populus*. Genome Res. 18, 422–430. https://doi.org/10.1101/gr.7076308

Zahn, L.M., Kong, H., Leebens-Mack, J.H., Kim, S., Soltis, P.S., Landherr, L.L., Soltis, D.E., dePamphilis, C.W., Ma, H., 2005. The evolution of the *SEPALLATA* subfamily of MADS-box genes. Genetics 169, 2209–2223. https://doi.org/10.1534/genetics.104.037770

Zhang, Q., Liu, C., Liu, Y., VanBuren, R., Yao, X., Zhong, C., Huang, H., 2015. High-density interspecific genetic maps of kiwifruit and the identification of sex-specific markers. DNA Res. Int. J. Rapid Publ. Rep. Genes Genomes 22, 367–375. https://doi.org/10.1093/dnares/dsv019

Zhou, R., Macaya-Sanz, D., Carlson, C.H., Schmutz, J., Jenkins, J.W., Kudrna, D., Sharma, A., Sandor, L., Shu, S., Barry, K., Tuskan, G.A., Ma, T., Liu, J., Olson, M., Smart, L.B., DiFazio, S.P., 2020. A willow sex chromosome reveals convergent evolution of complex palindromic repeats. Genome Biol. 21, 38. https://doi.org/10.1186/s13059-020-1952-4

Zhou, R., Macaya-Sanz, D., Rodgers-Melnick, E., Carlson, C.H., Gouker, F.E., Evans, L.M., Schmutz, J., Jenkins, J.W., Yan, J., Tuskan, G.A., Smart, L.B., DiFazio, S.P., 2018. Characterization of a large sex determination region in *Salix purpurea* L. (Salicaceae). Mol. Genet. Genomics MGG 293, 1437–1452. https://doi.org/10.1007/s00438-018-1473-y

Zou, C., Massonnet, M., Minio, A., Patel, S., Llaca, V., Karn, A., Gouker, F., Cadle-Davidson, L., Reisch, B., Fennell, A., Cantu, D., Sun, Q., Londo, J.P., 2021. Multiple independent recombinations led to hermaphroditism in grapevine. Proc. Natl. Acad. Sci. 118. https://doi.org/10.1073/pnas.2023548118

Appendix

Acknowledgements

The first time I arrived at Thünen Institute was special. The gate, the path surrounded by trees and a charming old building at the end. A PhD offer to uncover the genetic basis of dioecy in poplars. A choice that changed my life. A bit more than three years have passed now. So much learning, thinking, experimenting, failing, trying again, succeeding. A pandemic in the middle to spice things up. But now this life chapter is coming to an end. This is the perfect time for me to acknowledge the people that have supported me through my PhD journey.

First, I want to thank my supervisors. **Niels** you are a brilliant scientist and the perfect mentor. You have inspired me in so many ways. I admire your multi-tasking and leadership skills. Thank you so much for constantly motivate and teach me. You gave me the confidence and freedom to discover and develop my skills as a researcher. And I will be forever grateful for that. **Birgit** your outstanding ability to always ask the right questions are of huge importance to any student. I consider myself incredibly lucky to have you as my supervisor. Even though we didn't interact daily, it is amazing how much value you have added to my personal and academic development. Thank you. **Matthias** you have always been there. Be it in your office with open doors or on a morning tour around the lab building. Always available to give feedback. Always asking the smart questions, giving new ideas and keeping the good mood in any room.

I would like to thank other important people involved in this journey. **Annika** you have overseen all technical aspects of the project and executed everything with perfection. Thank you so much for all the organization and suggestions during the experiments. Without you this project would have been completely different. **Katrin** and **Vivi**, my forever office mates. Thank you so much for welcoming me with so much care. You both are incredible and were always available to help with anything. You showed me everything I need to know when I first arrived at the Institute. **Kiran** you are a special person. I am happy to have shared so many stories of success and (sometimes struggles)

of our experiments. **Gihwan** you came to the project right when the pandemic started and even though there were so many difficulties you never gave up from an idea. I am very grateful to have worked alongside you. **Melina** you and I have so much in common and even working on different projects we connected so well, despite the lockdowns. I wish we had spent more time in the office and the lab together though. And to everyone else I crossed paths, talked, shared stories, helped and asked for help. Thank you so much!

Finally, I want to thank my unconditional supporters. My mom **Nadia**, who is the best mom in the world. Thank you for always being my base. Thank you for your love and for supporting the decisions I make for my life even if it means I am not there near you. You are the fuel to keep me going. My sister **Maite** who is literally my own duplicate, and who has my back for anything. Thank you for your incredible ideas and experience and for teaching me to give my best every day. My Liebling **Max**. You are my best friend and my partner, and I thank you for your support, understanding, love and for being with me during this journey.

Declaration of contribution

Chapter 1 is a review paper where we examined the literature, both classical and recent, to explore the parallels and differences in sex determination among different dioecious species. My contributions were the literature search, summarization and writing of the original draft. Also, the preparation of most figures and the tables.

Chapter 2 consists in an original article where the genetic basis of sex determination in *Populus* was resolved. My contributions included development of methodology, execution of experiments both in the laboratory and growth chamber, collection and analysis of data, and involvement in the revision of the manuscript.

Chapter 3 is an investigation on the recombination frequency and segregation patterns along the SDR in white poplar. My contributions to this chapter were conceptualization and design of the experiment, use bioinformatics approach to identify and select the SNPs flanking the SDRs and the design of the specific markers. I also did the interpretation and discussion of the results and wiring of the original draft.

Chapter 4 provides a practical use of *ARR17* as a sex marker and suggest other ZW systems in the genus *Populus* based on DNA-seq coverage. My contributions to this chapter were utilizing bioinformatics approaches to analyze publicly available data from *Populus alba*, *P. adenopoda*, and *P. qiongdaoenis.* I also analyzed the marker results and wrote the original draft.

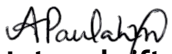
Chapter 5 explored differential gene expression based on transcriptomic data to reveal the molecular function of *ARR17*. My contributions to this article were conceptualization and design of the experiments, data collection, curation, and analysis, interpretation and discussion of the results, figure preparation and writing of the original draft.

Eidesstattliche Versicherung

Declaration on oath

Hiermit erkläre ich an Eides statt, dass ich die vorliegende Dissertationsschrift selbst verfasst und keine anderen als die angegebenen Quellen und Hilfsmittel benutzt habe.

I hereby declare, on oath, that I have written the present dissertation by my own and have not used other than the acknowledged resources and aids.



Hamburg, den 14 January 2022 **Unterschrift**

THÜNEN

The doctoral thesis "Elucidating the genetic basis and molecular mechanisms of dioecy in Populus" by the PhD. candidate Ana Paula Leite Montalvao is is written in clear and high quality English language. As an English native speaker, a translator (English/German) for the Thünen Institue in Braunschweig and a graduate of Barnard College, Columbia University in New York City, I am qualified to judge the authenticity and quality of Ms. Leite Montalvao's English language use and deem it more than sufficient.

Sincerely yours,

Dina Führmann