The Evolution of a Sex Chromosome in Asparagus

by

Alex E. Harkess

(Under the Direction of James H. Leebens-Mack)

Abstract

The overwhelming majority of flowering plants reproduce through the production of hermaphroditic flowers. A small percentage of angiosperm species instead are dioecious, producing either male or female flowers on individual plants. Dioecy can be mediated at the molecular level by a sex chromosome that genetically differentiates males and females. Sex chromosomes evolve from autosomes, and this conversion is hypothesized to involve mutations in one or more linked genes that determine sex. Given the complexities of anther and ovule development, the full suite of sex determination genes has not been described for any dioecious plant. Here we explore the conversion from autosome to an XY sex chromosome using garden asparagus (*Asparagus officinalis*), an ideal model system for studying the earliest events in sex chromosome evolution given that it recently evolved a sex chromosome pair. Focusing first on broad trends, genomic characterization of several hermaphroditic and dioecious species across the *Asparagus* genus revealed an increase in retrotransposon content coincident with the evolution of dioecy. To identify putative sex determination genes on this Y chromosome, we then explore the timing of male and female sterility events in garden asparagus, hypothesizing that anther sterility in females likely occurs before pollen microsporogenesis. Finally, we assemble and annotate a high quality reference genome for

garden asparagus, and perform a suite of mutant analyses to identify two genes in a non-recombining region of the Y that are ultimately responsible for sex determination.

INDEX WORDS: sex chromosomes, dioecy, polyploidy

The Evolution of a Sex Chromosome in Asparagus

by

Alex E. Harkess

B.S., Miami University, 2011

A Dissertation Submitted to the Graduate Faculty
of The University of Georgia in Partial Fulfillment
of the
Requirements for the Degree

Doctor of Philosophy

Athens, Georgia

2016

The Evolution of a Sex Chromosome in Asparagus

by

Alex E. Harkess

Major Professor:	James H. Leebens-Mack
Committee:	Kelly Dawe
	Katrien Devos
	Xiaoyu Zhang
	Kelly Dyer

Electronic Version Approved:

Suzanne Barbour
Dean of the Graduate School
The University of Georgia
May 2016

Acknowledgments

I dedicate this dissertation to my mentors, past and present, who have led me to be infinitely curious about plants. Because of you, I aspire to teach with as much passion as I was taught.

Contents

List of Figures

List of Tables

Chapter 1

A century of sex determination in plants

Harkess A, Leebens-Mack J. Submitted to *Heredity*. March 28, 2016.

1.1 Abstract

Plants have evolved a diverse array of strategies for sexual reproduction, particularly through modification of male and female gender functions. The immense variation in sexual systems across the land plants provides a unique opportunity to study the genetic, epigenetic, phylogenetic, and ecological underpinnings of sex determination. Here we reflect on the last century of research on plant sex determination including some of the earliest genetic and cytogenetic experiments focused on gender differences and recent genomic investigations. For more than a century botanical researchers have leveraged the diversity of plant mating systems to advance understanding of gender determination and comparative genomic analyses are now allowing us to address longstanding questions about the origin and early evolution of sex chromosomes.

1.2 Observations drive hypotheses

Plants have evolved a remarkable array of mating systems, including self-compatibility in hermaphrodites, obligate outcrossing in dioecious species, and myriad intermediate sexual systems. The Mendelian basis of sex inheritance in dioecious plants has been long studied but there have been exciting new breakthroughs in understanding the molecular basis for gender determination in dioecious species (e.g. Akagi et al. 2014, Boualem et al. 2015). As we work to place these recent findings into a coherent framework for understanding plant sex chromosome evolution and implications for sex chromosome research more broadly, it is an opportune time to consider the last century of research into gender differences and sex chromosomes in plants. In doing so, we find the same ecological and genetic questions that motivate our current research equally puzzled botanists more than a century ago: why do so many sex forms exist across the plant phylogeny; why do some dioecious plant species exhibit heteromorphic sex chromosomes while others have indistinguishable sex chromosomes; and

what are the sex-determining factors that differentiate males and females at the genic level?

Early observations of dioecious and subdioecious populations implicated separate factors influencing whether plants exhibited male, female, or mixed gender function. Geoffrey Smith (1904) was particularly influential in the earliest research of gender conversions, documenting how parasitic castration of male *Inachus mauritanicus* crabs led to the development of hermaphroditic characteristics. Given that parasitized females never developed any obvious male characteristics, Smith concluded that the male developmental pathway includes suppression of female function through changes in metabolism. Whereas this work provided insights into the nature of gender differences, the inheritance of these differences remained obscure.

Correns (1907) used dioecious *Bryonia dioica* (Cucurbitaceae) to infer that distinct pollen types produce male and females offspring in a Mendelian fashion. Correns surmised that a sex-determining factor must be homozygous in one sex and heterozygous in the other. This was a pivotal inference in the early development of sex determination research. Upon discovering and selfing reversible hermaphroditic mutants of the typically dioecious *Lychnis dioica* (now *Silene dioica*, a close relative of *Silene latifolia*), George Shull confirmed Correns Mendelian model of sex determination and inferred that mutant hermaphrodites were derived from modification of male genotypes (Shull 1911). This finding closed the circle between Smith's work on parasite-induced castration of *Inachus* with Correns' work on gender inheritance.

Even earlier, Wehrli (1892) elaborately described sexual transitions and phenomena in more than 80 plant species ranging from yearly sex changes in nutmeg, male to hermaphrodite flower conversions, and the formation of male flowers on female plants. Haring (1894) described similar plasticity in willows (*Salix caprea* and *Salix cinerea*). The plasticity of observed sex forms, especially the intergradation between discrete male and female forms, was the foundation for future curiosities about the cytogenetic differentiation of male and female

forms across the land plants.

1.3 The cytological foundation of sex differences

With a growing appreciation of the Mendelian nature of sex inheritance, the morphological diversity of plant sex chromosomes quickly became obvious through cytogenetic analyses. Following quickly on the heels of animal research (e.g. Morgan 1914), hetermorophic plant sex chromosomes were first discovered in the liverworts by Charles Allen at the University of Wisconsin, reporting a large chromosome only present in female *Sphaerocarpus donnellii* (Allen, 1917). Blackburn (1923, 1924) and Winge (1923) soon afterwards independently described *Lychnis dioica* as containing homomorphic XX chromosomes in females and heteromorphic XY sex chromosomes in males, with the Y being much larger than the X.

Botanists doing cytological research first hypothesized that the difference in chromatin content between males and females with heteromorphic sex chromosomes was a key determinant of sex. This hypothesis was soon discounted given the observation of indistinguishable homomorphic chromosomes in males and females of many animal systems (Winge 1923, Jones 1933). Researchers returned their attention to the identification of sex determination genes that may reside on either homomorphic or heteromorphic sex chromosomes in males and females. Cytological investigation of hermaphrodite mutants was exceptionally important in *Silene latifolia*. Westergaard (1940, 1946, 1958) identified hermaphrodite mutants with Y chromosomes destroyed at the distal ends, suggesting that at least part of the Y chromosome encodes a region that dominantly suppresses female carpel development in males while simultaneously promoting the proper development of anthers. But as Winge (1937) wrote, The views as to where the genes are located and whether a single gene or several genes be present in the respective [sex] chromosome suggest that the matter is more complicated than was at first supposed.

1.4 Identifying sex determination genes before genomics

Understanding the developmental genetics of male and female function in hermaphroditic species is critical for elucidation of sex determination in dioecious species. Given that we now better understand the complexity of the anther and ovule development pathways (Chang et al., 2011; Wilson et al., 2011), it is possible that mutations in myriad genes in those pathways can lead to the same terminal phenotypes of male and female sterility. Consequently in hermaphroditic or monoecious flowers, every necessary floral gene can be considered a sex determination gene. Given the relative conservation of many aspects of both the male and female floral development pathways across the angiosperms, our understanding of a sex determination gene in one species can be informed through analysis of another tractable system.

Maize, a monoecious species with separate male (tassels) and female (ears) flowers on the same plant, played a critical role in the early understanding of the genes involved in male and female function and development. In particular, the identification of various floral mutant elucidated several floral development pathways and sex determination genes. Emerson (1920) first described the seminal *tassleseed-1* and *tassleseed-2* mutants which, among disruptions in branching of inflorescences, ears, and tassels, led to the near-complete feminization of the male tassels. Phipps (1928) found a related mutant *tassleseed-4* with a similar feminization phenotype. Nearly a century later, it is now known that the tassleseed-4 mutation disrupts the miR172 microRNA that normally targets APETALA2 (AP2) floral homeotic transcription factors for proper function, whereas *tassleseed-1* eliminates expression of a plastid-encoded lipoxygenase that severely reduces jasmonic acid production (Chuck et al., 2007; Acosta et al., 2009).

Through the discovery of mutants that modified male or female function, it was further realized that sex differences could not be solely attributed to genes on a sex chromosome.

Strasburger (1900) described a fungal-induced sex change where *Ustilago violaceae* anther smut fungus (now *Microbotrytum violaceum*) can infect *Melandrium rubrum* (now *Silene dioica* previously known as *Lychnis dioica*, or red campion) females to produce rudimentary, sterile anthers inside of which the fungus sporogenously propagates. A nearly identical pattern of infection and sterile anther growth in females was seen in closely related *Lynchis dioica* by Doncaster (1912). In both experiments, male flowers were phenotypically and functionally unaffected by infection. Even X:Y chromosome dosage ratios were tested independently (and by coincidence, concurrently) by Warmke and Blakeslee (1939) and Westergaard (1940) in synthetic autotetraploid *Melandrium dioicum* and *Melandrium album* individuals, showing elegantly that regardless of X chromosome number in XY, XXY, or XXXY genotypes, a single copy of the Y was necessary for male determination. It became clear that Y chromosomes must contain male-specific sequences or alleles that contribute to the active determination of maleness, but the identification and characterization of these loci has continually proved difficult.

1.5 The modern genomics era

With the emergence of genomic tools, there has been an expanding number of model plant systems for investigating the evolution of sex chromosomes and sex determination genes. Plants, and particularly flowering plants, are unique in the sense that they have repeatedly transitioned between sexual systems across the phylogeny (Bachtrog et al., 2014). Roughly 5-10% of angiosperm species have been characterized as dioecious (Renner et al., 2012). We now understand better that male and female floral organogenesis in plants is regulated by an assortment of genes, transcription factors, small RNAs, hormones, environmental factors, and epigenetic marks that interact for proper androecium and gynoecium development (Chuck et al., 2007; Niculita-Hirzel et al., 2008; Martin et al., 2009; Nag and Jack,

2010; Hartwig et al., 2011; Adam et al., 2011; Song et al., 2013). Consequently, the particular male and female sterility mutations on a sex chromosome pair that initially drive a hermaphroditic species to dioecy are likely different across all independently evolved sex chromosomes (Ainsworth, 2000; Diggle et al., 2011). As a result, every novel sex chromosome system poses unique challenges when identifying likely candidate sex determination genes.

While there is little reason to expect independently derived plant sex chromosomes to harbor the same sex determination genes, there is reason to expect common evolutionary themes in the evolution of sex chromosome structure (Ming et al., 2011; Charlesworth, 2015) A hallmark of a typical plant nuclear sex chromosome is a non-recombining region flanked by recombining pseudo-autosomal region(s) (PAR). The physical size of the non-recombining region relative to the PAR is highly variable across sex chromosome systems. Hemizygous sex chromosomes (Y or W) are hypothesized to grow as a consequence of irreversible transposon, satellite repeat, and organellar DNA accumulation in the non-recombining region as well as possible translocation of genes contributing to male or female fitness in the case of Y and W chromosomes, respectively. Consequently, the size of the non-recombining region on the Y/W is expected to expand over time relative to the size of homologous regions on X/Z chromosomes (Ming et al., 2011). This prediction does not hold for some bird, snake, and frog lineages though, which have maintained small non-recombining regions on homomorphic sex chromosome pairs for long periods of time (Bergero and Charlesworth, 2009; Guerrero et al., 2012; Bewick et al., 2013). As it currently stands, identifying global trends across plant sex chromosomes is difficult given the small sample size of sequenced dioecious plant genomes. Comparative analysis of independently derived sex chromosome systems is forthcoming but sequence and assembly of repeat rich non-recombining sex determination regions remains challenging. With technological advances including long-read sequencing and physical mapping approaches, comparative analyses across dioecious lineages can be utilized to test whether parallels exist in sex chromsome evolution and structure across divergent taxa.

At the same time, comparative analyses within dioecious plant groups will allow researchers to elucidate conditions under which the size non-recombining regions may be expected to stable or dynamic.

Several plant systems have been particularly important to our understanding of the evolution of sex chromosomes and sex determination genes, though. The power of plants for studying the evolution of sex determination genes on sex chromosomes comes not from a single system, but by holistically studying the repeated evolution and subsequent diversity of sex chromosomes across the flowering plant phylogeny. While by no means an exhaustive or all-inclusive list, the establishment of genomic research into asparagus, papaya, strawberry, white campion, persimmon, *Rumex*, *Populus*, and *Salix* has driven significant discoveries about the early and progressive evolution of sex chromosomes in angiosperms. Here we give a concise overview of several recent advances in these systems, and in particular focus on the diverse set of approaches used to identify putative sex determination candidates.

Garden asparagus (*Asparagus officinalis*)

Garden asparagus is an agriculturally valuable monocot species with a young, homomorphic XX/XY sex chromosome pair. Dioecy evolved from hermaphroditism once or potentially twice roughly 2 million years ago in the genus Asparagus, concordant with a range shift from South Africa to North Africa, Europe, and Asia (Kuhl et al., 2005; Kubota et al., 2012; Norup et al., 2015). Developmentally, both males and females initiate hermaphroditic flowers, but XX females lose their anthers sometime during tapetum development and pollen meiosis while XY males never fully form a stylar tube (Caporali et al., 1994); this morphological difference is also seen in gene expression data following the anther development pathway, as XX females cease expression of anther-related genes sometime during tapetum development and before pollen microsporogenesis (Harkess et al., 2015). Colchicine treatment of anthers can produce viable YY doubled haploid genotypes, suggesting that little degeneration has

occurred on the Y (Ferrie and Caswell, 2010). The viability of the YY genotype allows for the production of all-XY male F1 progeny, which typically have a higher yield and lifespan relative to XX females (Falloon and Nikoloff, 1986).

The sex determination region has been genetically mapped to a single locus of the *A. officinalis* Y chromosome (Telgmann-Rauber et al., 2007), and several sex-specific markers and full bacterial artificial chromosomes (BACs) have been identified (Jamsari et al., 2004; Gao et al., 2007). These genomic regions are largely dominated by recently transposed LTR retroelements and are quite gene-poor, which hampers traditional genome walking (Vitte et al., 2013). The ability to culture entirely homozygous XX and YY individuals reduces the complexity of de novo genome assembly. A *de novo* genome assembly, optical map, and genetic map and annotation for a doubled haploid YY *A. officinalis* individual has identified a nearly 1.8 Mb Y-specific region containing 14 genes (Harkess et al., in preparation). Several independently-derived male-to-hermaphrodite mutants (including Co-60 gamma irradiation and spontaneous mutants) identify a single uncharacterized gene containing a Domain of Unknown Function 247 (DUF 247) in the 1.8Mb hemizygous non-recombining region. Also inside this male-specific region is a single copy tapetum-related gene, TAPETUM DEVELOPMENT AND FUNCTION 1, with an *Arabidopsis* knockout phenotype similar to anther sterility that occurs in asparagus females. This study is the first genic evidence that supports Deborah and Brian Charlesworths model for the origin of Y chromosomes involving linkage of at least one female suppressor with at least one male promoting gene (Charlesworth, 1978).

Papaya (*Carica papaya*)

Papaya is a sub-dioecious plant with females (XX), males (XY), and hermaphrodites (XYh) that is native to Costa Rica. The YY genotype is inviable which indicates that a greater degree of Y chromosome degeneration has occurred compared to garden asparagus. The Yh chromosome is an alternate form of the Y that specifies hermaphroditism, but the

Y and Yh chromosomes only differ by 1.32% at the DNA level (Yu et al., 2008). The sex determination region in papaya, otherwise known as the Hermaphrodite/Male Specific Region of the Y (HSY/MSY), is largely heterochromatic and spans greater than 10% (8Mb) of the chromosome, including the centromere (Zhang et al., 2008). BAC-by-BAC assembly and annotation of the HSY and the homologous X counterpart reveals more than 70 shared genes which were used to identify two inversion events on the Y form the evolutionary strata of suppressed recombination, the older of which is dated to around 7 million years ago (Wang et al., 2012).

There is a wealth of genomic resources, genetic markers, expression data, and population-level analyses for papaya (Ming et al., 2008; Zhang et al., 2008; Weingartner and Moore, 2012; Wai et al., 2012; Aryal et al., 2014). Owing to the complexity of floral development in papaya, no genes have yet been directly implicated in sex determination in papaya, though candidate genes have been proposed (Ueno et al., 2014).

White campion (*Silene latifolia*)

White campion (*Silene latifolia*), formerly known as *Melandrium album*, is a dioecious species with a highly divergent X/Y sex chromosome pair. It is perhaps no surprise that with such a rich ecological and genetic research history, *Silene latifolia* emerged as the model plant sex chromosome system for early genetic research in the 1990s. Beginning with PCR-based strategies such as random amplified polymorphic DNA (RAPD) (Mulcahy et al., 1992; Zhang et al., 1998), amplified fragment length polymorphism (AFLP), and sequence characterized amplified region (SCAR) markers paired with genomic and cDNA subtraction libraries (Straus and Ausbel, 1990; Donnison et al., 1996), *Silene latifolia* is undoubtedly the most historically established system for the identification of sex-specific and sex determination genes. The Y chromosome is cytologically distinct; it has expanded to nearly 2.5 times the length of the X and is replete with satellite DNA (Kubat et al.,

2008), nuclear-integrated organellar DNA (Kejnovsky et al., 2006), and active transposons (Pritham et al., 2003).

The non-recombining region of the Y is large, and there is evidence that Y-specific degeneration has occurred at the gene level (Guttman and Charlesworth, 1998; Filatov and Charlesworth, 2002; Papadopulos et al., 2015). The identification of sex determination genes in *S. latifolia* is particularly difficult given that the non-recombining region is so large, likely comprising many more genes than in garden asparagus and papaya. A suite of genes have been identified as Y-linked, such as APETALA3 (Cegan et al., 2010), SIY1 and SIY4 (Filatov and Charlesworth, 2002), and more recently several thousand SNPs were identified through transcriptome sequencing of controlled crosses (Muyle et al., 2012). There is also debate over whether or not, and to what degree, sex chromosome dosage compensation is occurring in *S. latifolia*, a phenomenon that previously excluded plant sex chromosomes (Muyle et al., 2012; Papadopulos et al., 2015).

Strawberry (*Fragaria* sp.)

Fragaria virginia, the Virginian wild strawberry, is an octoploid sub-dioecious species with female, male, and hermaphrodite sexes. A ZZ/ZW proto-sex chromosome controls gender with at least two sex determining regions linked in partial recombination (Spigler et al., 2008). The region that suppresses male development is dominant to the female fertility region, meaning that females are the heterogametic (ZW) sex. Significant QTL for sexually dimorphic traits such as variation in fruit set have been found on the proto-sex chromosome (Spigler et al., 2011). Given that the sex loci are not perfectly linked and can recombine, the strawberry proto-sex chromosome represents a critically important intermediate stage between gynodioecy and full dioecy (Moore, 2009). An array of sequence data has been generated for numerous diploid and octoploid gynodioecious Fragaria species and hybrids in an effort to more finely map sterility loci. A genome assembly and annotation is available for diploid *Fragaria vesca* ssp. vesca (Shulaev et al., 2011), in which male sterility in closely

related *Fragaria vesca* ssp. bracteata maps to a 338kb region with 57 annotated genes. None of the contained genes have known function in male sterility, which suggests that genes have previously unknown function or alternative mechanisms may be responsible for sterility. Interestingly, male sterility is not consistently mapped to the same syntenic block across all gynodioecious *Fragaria* species, suggesting that male sterility has either been repeatedly independently evolved, the same sterility locus been translocated to different regions across species, or the same sterility phenotype is under the control of different genes (Govindarajulu et al., 2013; Tennessen et al., 2013). The seemingly labile nature of sterility mutations in *Fragaria* may be related to repeated polyploidization events in the genus, which have been implicated in both the transition to dioecy as well as the reversion back to hermaphroditism (Ashman et al., 2013; Njuguna et al., 2013).

Poplars, cottonwoods, aspens (*Populus*) and Willows (*Salix*)

The dioecious *Populus* and *Salix* lineages diverged within the Salicaceae at least 45 million years ago (Manchester et al., 2006), suggesting a relatively ancient origin of separate genders within the family. *Populus* and *Salix* sex chromosomes are, however, homomorphic with small non-recombining regions. Whereas perhaps most *Populus* species have an XY sex determination system (Geraldes et al. 2015, Kersten et al., 2014) females are the hemizygous sex (i.e. ZW system) in other *Populus* (Yin et al. 2008, Paolucci et al. 2010) and all characterized *Salix* species (Hou et al. 2015, Pucholt et al. 2015). Further, within *Populus*, the position of the sex determination region has been reported as mapping toward the telomere in some species (e.g. Geraldes et al. 2015) and pericentromeric in others (Pakull et al., 2009, Paolucci et al. 2010). The small size of the non-recombining regions and variation in the hemizygous gender, together with synteny analyses have led some to hypothesize multiple independent origins for sex chromosomes within *Salix* and *Populus* (Hou et al. 2015, Pucholt et al. 2015). However, as acknowledged by the authors of these studies, translocation of currently unknown sex determination genes could result in the formation and perhaps

turnover of neo-sex chromosomes (e.g. Stck et al. 2011, Mank and Avis 2009, Bachtrog et al. 2008) and explain the observed variation. Genomic rearrangements and the origin of neo-sex chromosomes has also been hypothesized within *Silene* (Howell et al. 2009, Weingartner and Delph 2014).

Persimmon (*Diospyros lotus*)

Caucasian persimmon is the first XY sex chromosome system to have a well-characterized sex determination gene identified in a non-recombining region of the Y chromosome (Akagi et al., 2014), representing a hallmark in the advancement of sex chromosome functional genomics. In particular, this study shows the unique power of whole genome sequencing paired with novel comparisons of male and female kmer distributions to identify and assemble male-specific sequences in the absence of a fully sequenced and assembled genome. Through the male-specific kmer-based assembly of a pseudo-reference non-recombining region and gene annotation, with additional evidence from male and female RNA-seq differential expression, the OGI and MeGI class I homeodomain transcription factors were identified as a putative sex determinant; these genes are hypothesized to act together as OGI mRNA encodes a predicted hairpin structure that may be cleaved into small RNAs to suppress expression of MeGI in males. Heterologous transformations of MeGI overexpression constructs into Arabidopsis showed stunted and male-sterile phenotypes resulting from MeGI-driven suppression of anther development, elegantly displaying the power of comparative functional genomics in more tractable systems and the relative conservation of the anther development pathway across phylogenetically divergent angiosperms.

Rumex (*Rumex hastatulus*)

Rumex is a particularly interesting genus given that two monophyletic lineages have evolved separate sex determination configurations, a more ancient XX/XY chromosome pair and a neo-sex chromosome XX/XY1Y2 derived from an X:autosome fusion (Smith, 1964; Grabowska-Joachimiak et al., 2014). Just as in persimmon, the absence of an as-

sembled genome does not necessarily inhibit the identification of putative sex determination genes. By leveraging transcriptome sequencing and SNP calling in a diverse panel of related males and females, Hough et al. (2014) identified a set of sex-linked transcript assemblies that showed variable expression and synonymous site substitution rates between X and Y-linked copies and between the hermaphroditic outgroup *Rumex bucephalophorus*, indicating a rapid degeneration of Y-linked genes. This novel combination of cross-species comparisons, transcriptome-based SNP segregation analysis, expression analysis, and molecular evolutionary comparisons showcases the utility of transcriptome sequencing and assembly paired with thoughtful experimental designs to extract large volumes of sex-linked data from single experiments.

1.6 From informatics to functional genomics

A major impediment to the identification of sex determination genes is the ability to accurately assembly the full genomic sequence of a sex chromosome pair. Young, highly homomorphic sex chromosomes are typically not divergent enough to separately assemble using shotgun genomic reads, while older sex chromosomes often suffer from the increased accumulation of transposable elements which cannot currently be assembled without high coverage and longer read data. Flow sorting and laser microdissection have been used to specifically isolate heteromorphic sex chromosomes (Veuskens et al., 1995; Matsunaga et al., 1999), but is still not feasible in separating homomorphic sex chromosome pairs without the assistance of X and Y-specific fluorescent markers. Given these considerations, the previous gold standard of plant sex chromosome sequencing was the tedious BAC-by-BAC process, typified by the tiled BAC assembly of the papaya X and Yh chromosomes (Wang et al., 2012; Gschwend et al., 2012). This trend has rapidly evolved, though; with more modern de novo long read sequencing and assembly strategies (eg. Pacific Biosystems, Illumina TruSeq

Synthetic Long-Read) paired with optical mapping (eg. BioNano Genomics) and ultra long-range mate pair libraries (eg. Hi-C, Dovetail Genomics), genome assemblies that span entire non-recombining regions are more feasible and affordable than ever. Despite considerable advances in genome sequencing and assembly methods, the assembly of sex chromosomes will continue to be difficult. A consequence of suppressed or non-recombination in sex chromosomes is that repetitive DNA can rapidly proliferate. The proliferation of repetitive DNA can interfere with gene function, ultimately leading to the degeneration of a sex chromosome, and also makes them intractable regions for full-length genome assembly and annotation due to the high degree of similar or near-identical repetitive DNA. RNA-Seq can instead be utilized as a powerful and cost-effective approach to identifying sex-specific transcripts. Reconstructing a floral transcriptome *de novo* is much less demanding than assembling a complex plant genome, which are often rife with the remnants of polyploidization events and repetitive element proliferations (Kejnovsky et al., 2009). After isolating biologically replicated RNA from floral tissues during the arrest of male and female development, short read sequencing (eg., on the Illumina platform) can quickly generate hundreds of millions of mRNA-derived sequence reads with extremely low error rates. When performing crosses and sequencing parents and offspring, the RNA-Seq alignments can be mined for sex-specific SNPs. For example, RNA-Seq of small cross populations has been used to identify more than 7,000 sex-linked SNPs in the Silene latifolia transcriptome (Bergero and Charlesworth, 2011; Muyle et al., 2012) and to identify a set of sex-linked transcripts in *Rumex* (Hough et al., 2014). In addition to SNP marker development, transcript expression levels can be quantified and differentially expressed transcripts between sexes can be identified (e.g., Harkess et al., 2015). Transcripts that are both sex-linked and differentially expressed between sexes are likely promising candidate sex determination loci. Other high-throughput methods for sex-linked marker identification include Genotype-by-Sequence (GBS) (eg., Peterson et al. 2012), which can be used to rapidly genotype mapping populations or bulked segregants for

gender.

Once sex-linked genes have been identified, a large challenge remains in characterizing their functions and putative roles in floral development, with the ultimate goal of identifying the master regulator(s) of sterility. Fortunately, many floral gene regulatory networks appear to be conserved in androecium and gynoecium development across the angiosperms (Whipple et al., 2004; Davila-Velderrain et al., 2013), placing *Arabidopsis thaliana* as an invaluable study system for heterologous investigation of candidate sex determination genes as typified by the persimmon sex determination story. Loss-of-function mutations (eg., T-DNA insertions) in numerous *A. thaliana* genes can lead to the arrest of anther and ovule development at a variety of stages, reiterating the genetic complexity of proper floral organ development (Ma and Sundaresan, 2010; Ge et al., 2010; Chang et al., 2011). If orthologs of candidate genes can be identified in A. thaliana, T-DNA mutants can be crossed to homozygosity and phenotyped for the hypothesized sterility event. A major caveat to heterologous analyses is that while even distant orthologs are proposed to retain function (Nehrt et al., 2011; Chen and Zhang, 2012), the possibility of novel function in distant orthologs cannot be ignored.

A more direct approach to functional characterization of sex-determining gene candidates is through gene expression analyses in the species of interest. Functional analysis involving gene expression knockdowns of candidate sex determination genes is feasible in systems without transformation protocols. Virus-Induced Gene Silencing (VIGS) is one powerful option, where a target gene is expressed inside a viral vector that is known to infect the host plant. Upon successful infection, the host plants endogenous RNA-Induced Silencing Complex (RISC) produces small RNAs that target both the virus and target gene, degrading RNAs emanating from both and ultimately resulting in the transcriptional knockdown of a candidate gene (eg., Renner et al. 2009). Other options include forward screens such as chemically mutagenizing large seed populations with Ethyl methanesulfonate (EMS) to induce point mutations, or using gamma radiation to induce large chromosomal deletions

and specifically look for mutations in candidate sex determination genes. The concept of a model system is rapidly changing, though, and CRISPR-Cas9 gene editing and knockdown strategies are readily available for plant systems which are able to be transformed.

Ultimately, the identification of sex determination genes is difficult in every sex chromosome system because of the unifying theme that flowers are structurally heterogenous, transcriptionally active and developmentally complex organs. Due to the independent, repeated evolution of sex chromosomes across the land plants, the floral genes involved in sex determination are likely not the same across distant dioecious species. As the cost of sequencing continues to decrease while the quality and length of reads increases, we should expect to see more full-length plant sex chromosome assemblies. These full-length sex chromosome assemblies can be placed in a comparative genomics framework to assess changes relative to an autosomal ancestor, but should also be useful to compare the structures of divergent plant sex chromosomes for a high-level view of how sex chromosomes age. Genome assembly is just a first step, though, and sex chromosome biologists need to embrace a suite of bioinformatic, functional, phylogenetic, and developmental approaches to identify the causal sex chromosome sterility mutations.

1.7 References

Acosta IF, Laparra H, Romero SP, Schmelz E, Hamberg M, Mottinger JP, et al. (2009). Tasselseed1 Is a Lipoxygenase Affecting Jasmonic Acid Signaling in Sex Determination of Maize. Science 323: 2625.

Adam H, Collin M, Richaud F, Beul T, Cros D, Omor A, et al. (2011). Environmental regulation of sex determination in oil palm: current knowledge and insights from other species. Ann Bot 108: 152937.

Ainsworth C (2000). Boys and Girls Come Out to Play: The Molecular Biology of

Dioecious Plants. Ann Bot 86: 211221.

Akagi T, Henry IM, Tao R, Comai L (2014). A Y-chromosomeencoded small RNA acts as a sex determinant in persimmons. Science (80-) 346: 646650.

Allen CE (1917). A CHROMOSOME DIFFERENCE CORRELATED WITH SEX DIFFERENCES IN SPHAeROCARPOS. Science 46: 466467.

Aryal R, Jagadeeswaran G, Zheng Y, Yu Q, Sunkar R, Ming R (2014). Sex specific expression and distribution of small RNAs in papaya. BMC Genomics 15: 20.

Ashman T-L, Kwok a, Husband BC (2013). Revisiting the dioecy-polyploidy association: alternate pathways and research opportunities. Cytogenet Genome Res 140: 24155.

Bachtrog D (2014). Sex determination: why so many ways of doing it? PLoS Biol 12: e1001899.

Bergero R, Charlesworth D (2009). The evolution of restricted recombination in sex chromosomes. Trends Ecol Evol 24: 94102.

Bergero R, Charlesworth D (2011). Preservation of the Y transcriptome in a 10-million-year-old plant sex chromosome system. Curr Biol 21: 14704.

Bewick AJ, Chain FJJ, Zimmerman LB, Sesay A, Gilchrist MJ, Owens NDL, et al. (2013). A large pseudoautosomal region on the sex chromosomes of the frog Silurana tropicalis. Genome Biol Evol 5: 108798.

Blackburn (1923). Sex Chromosomes in Plants. Nature 112:687-688.

Blackburn (1924). The cytological aspects of the determination of sex in the dioecious forms of Lychnis. Brit. J. Exp. Biol. 1:413-430.

Caporali E, Carboni A, Galli M (1994). Development of male and female flower in Asparagus officinalis. Search for point of transition from hermaphroditic to unisexual developmental pathway. Sex Plant Reprod: 239249.

Cegan R, Marais GA, Kubekova H, Blavet N, Widmer A, Vyskot B, et al. (2010). Structure and evolution of Apetala3, a sex-linked gene in Silene latifolia. BMC Plant Biol

10: 180.

Chang F, Wang Y, Wang S, Ma H (2011). Molecular control of microsporogenesis in Arabidopsis. Curr Opin Plant Biol 14: 6673.

Charlesworth B (1978). Model for evolution of Y chromosomes and dosage compensation. Proc Natl Acad Sci U S A 75: 561822.

Charlesworth D (2015). Tansley review Plant contributions to our understanding of sex chromosome evolution.

Chen X, Zhang J (2012). The Ortholog Conjecture Is Untestable by the Current Gene Ontology but Is Supported by RNA Sequencing Data (CA Ouzounis, Ed.). PLoS Comput Biol 8: e1002784.

Chuck G, Meeley R, Irish E, Sakai H, Hake S (2007). The maize tasselseed4 microRNA controls sex determination and meristem cell fate by targeting Tasselseed6/indeterminate spikelet1. Nat Genet 39: 151721.

Correns, C (1907). Die Bestimmung und Vererbung des Geschlechtes, nach Versuchen mit hoheren Pflanzen. Verh. Ges. deutscher Naturforscher und Arzte 1907: 794802.

Davila-Velderrain J, Servin-Marquez A, Alvarez-Buylla ER (2013). Molecular Evolution Constraints in the Floral Organ Specification Gene Regulatory Network Module across 18 Angiosperm Genomes. Mol Biol Evol 31: 560573.

Diggle PK, Di Stilio VS, Gschwend AR, Golenberg EM, Moore RC, Russell JRW, et al. (2011). Multiple developmental processes underlie sex differentiation in angiosperms. Trends Genet 27: 36876.

Falloon PG, Nikoloff AS (1986). Asparagus: value of individual plant yield and fern characteristics as selection criteria. New Zeal J Exp Agric 14: 417420.

Ferrie a. MR, Caswell KL (2010). Isolated microspore culture techniques and recent progress for haploid and doubled haploid plant production. Plant Cell, Tissue Organ Cult 104: 301309.

Filatov D a, Charlesworth D (2002). Substitution rates in the X- and Y-linked genes of the plants, Silene latifolia and S. dioica. Mol Biol Evol 19: 898907.

Gao WJ, Li RL, Li SF, Deng CL, Li SP (2007). Identification of two markers linked to the sex locus in dioecious Asparagus officinalis plants. Russ J Plant Physiol 54: 816821.

Ge X, Chang F, Ma H (2010). Signaling and transcriptional control of reproductive development in Arabidopsis. Curr Biol 20: R98897.

Govindarajulu R, Liston a, Ashman T-L (2013). Sex-determining chromosomes and sexual dimorphism: insights from genetic mapping of sex expression in a natural hybrid Fragaria ananassa subsp. cuneifolia. Heredity (Edinb) 110: 4308.

Grabowska-Joachimiak A, Kula A, Ksiczyk T, Chojnicka J, Sliwinska E, Joachimiak AJ (2014). Chromosome landmarks and autosome-sex chromosome translocations in Rumex hastatulus, a plant with XX/XY1Y2 sex chromosome system. Chromosom Res: 187197.

Gschwend AR, Weingartner L a, Moore RC, Ming R (2012). The sex-specific region of sex chromosomes in animals and plants. Chromosome Res 20: 5769.

Guerrero RF, Kirkpatrick M, Perrin N (2012). Cryptic recombination in the ever-young sex chromosomes of Hylid frogs. J Evol Biol 25: 194754.

Guttman DS, Charlesworth D (1998). An X-linked gene with a degenerate Y-linked homologue in a dioecious plant. Nature 393: 2636.

Harkess A, Mercati F, Shan HY, Sunseri F, Falavigna A, Leebens-Mack J (2015). Sex-biased gene expression in dioecious garden asparagus (Asparagus officinalis). New Phytol 207: 883892.

Hartwig T, Chuck GS, Fujioka S, Klempien A, Weizbauer R, Prasad D, et al. (2011). Brassinosteroid control of sex determination in maize. 020028.

Hough J, Hollister JD, Wang W, Barrett SCH, Wright SI (2014). Genetic degeneration of old and young Y chromosomes in the flowering plant Rumex hastatulus. Proc Natl Acad Sci U S A 2014: 16.

Jamsari a, Nitz I, Reamon-Bttner SM, Jung C (2004). BAC-derived diagnostic markers for sex determination in asparagus. Theor Appl Genet 108: 11406.

Kejnovsky E, Hobza R, Cermak T, Kubat Z, Vyskot B (2009). The role of repetitive DNA in structure and evolution of sex chromosomes in plants. Heredity (Edinb) 102: 53341.

Kejnovsky E, Kubat Z, Hobza R, Lengerova M, Sato S, Tabata S, et al. (2006). Accumulation of chloroplast DNA sequences on the Y chromosome of Silene latifolia. Genetica 128: 167175.

Kubat Z, Hobza R, Vyskot B, Kejnovsky E (2008). Microsatellite accumulation on the Y chromosome in Silene latifolia. Genome 51: 350356.

Kubota S, Konno I, Kanno A (2012). Molecular phylogeny of the genus Asparagus (Asparagaceae) explains interspecific crossability between the garden asparagus (A. officinalis) and other Asparagus species. Theor Appl Genet 124: 34554.

Kuhl JC, Havey MJ, Martin WJ, Cheung F, Yuan Q, Landherr L, et al. (2005). Comparative genomic analyses in Asparagus. Genome 1060: 10521060.

Ma H, Sundaresan V (2010). Development of flowering plant gametophytes. Curr Top Dev Biol 91: 379412.

Martin A, Troadec C, Boualem A, Rajab M, Fernandez R, Morin H, et al. (2009). A transposon-induced epigenetic change leads to sex determination in melon. Nature 461: 11358.

Matsunaga S, Kawano S, Michimoto T, Higashiyama T, Nakao S (1999). Semi-Automatic Laser Beam Micro dissection of the Y Chromosome and Analysis of Y Chromosome DNA in a Dioecious Plant , Silene latifolia. 40: 6068.

Ming R, Bendahmane A, Renner SS (2011). Sex chromosomes in land plants. Annu Rev Plant Biol 62: 485514.

Ming R, Hou S, Feng Y, Yu Q, Dionne-Laporte A, Saw JH, et al. (2008). The draft genome of the transgenic tropical fruit tree papaya (Carica papaya Linnaeus). Nature 452:

991996.

Moore RC (2009). Sex chromosome evolution: a missing link in the evolution of sex chromosomes. Heredity (Edinb) 102: 2112.

Morgan TH (1914). Heredity and sex. Z Indukt Abstamm Vererbungsl 12: 159161.

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 (D Weigel, Ed.). PLoS Biol 10: e1001308.

Nag A, Jack T (2010). Sculpting the flower; the role of microRNAs in flower development. Curr Top Dev Biol 91: 34978.

Nehrt NL, Clark WT, Radivojac P, Hahn MW (2011). Testing the ortholog conjecture with comparative functional genomic data from mammals. PLoS Comput Biol 7: e1002073.

Niculita-Hirzel H, Stck M, Perrin N (2008). A key transcription cofactor on the nascent sex chromosomes of European tree frogs (Hyla arborea). Genetics 179: 17213.

Njuguna W, Liston A, Cronn R, Ashman T-L, Bassil N (2013). Insights into phylogeny, sex function and age of Fragaria based on whole chloroplast genome sequencing. Mol Phylogenet Evol 66: 1729.

Norup MF, Petersen G, Burrows S, Bouchenak-Khelladi Y, Leebens-Mack J, Pires JC, et al. (2015). Evolution of Asparagus L. (Asparagaceae): Out-of-South-Africa and multiple origins of sexual dimorphism. Mol Phylogenet Evol 92: 2544.

Papadopulos AST, Chester M, Ridout K, Filatov DA (2015a). Rapid Y degeneration and dosage compensation in plant sex chromosomes. Proc Natl Acad Sci 112: 201508454.

Peterson BK, Weber JN, Kay EH, Fisher HS, Hoekstra HE (2012). Double digest RADseq: an inexpensive method for de novo SNP discovery and genotyping in model and non-model species. PLoS One 7: e37135.

Pritham EJE, Zhang YH, Feschotte C, Kesseli RR V (2003). An Ac-like transposable element family with transcriptionally active Y-linked copies in the white campion, Silene

latifolia. Genetics 165: 799807.

Renner T, Bragg J, Driscoll HE, Cho J, Jackson AO, Specht CD (2009). Virus-induced gene silencing in the culinary ginger (Zingiber officinale): an effective mechanism for down-regulating gene expression in tropical monocots. Mol Plant 2: 108494.

Renner SS, Ricklefs RE, Ricklefs E (2012). DIOECY AND ITS CORRELATES IN THE FLOWERING PLANTS groups. 82: 596606.

Shulaev V, Sargent DJ, Crowhurst RN, Mockler TC, Folkerts O, Delcher AL, et al. (2011). The genome of woodland strawberry (Fragaria vesca). Nat Genet 43: 10916.

Shull, G (1910). Inheritance of sex in Lychnis. Lychnis Bot. Gazette 49:110-125.

Shull, G (1911). Reversible sex mutants in Lychnis dioica. Bot. Gazette 52:329-368.

Smith B (1964). The evolving karyotype of Rumex hastatulus. Evolution (N Y) 18: 93104.

Song Y, Ma K, Ci D, Zhang Z, Zhang D (2013). Sexual dimorphism floral microRNA profiling and target gene expression in andromonoecious poplar (Populus tomentosa). PLoS One 8: e62681.

Spigler RB, Lewers KS, Ashman T-L (2011). Genetic architecture of sexual dimorphism in a subdioecious plant with a proto-sex chromosome. Evolution 65: 111426.

Spigler RB, Lewers KS, Main DS, Ashman T-L (2008). Genetic mapping of sex determination in a wild strawberry, Fragaria virginiana, reveals earliest form of sex chromosome. Heredity (Edinb) 101: 50717.

Telgmann-Rauber A, Jamsari A, Kinney MS, Pires JC, Jung C (2007). Genetic and physical maps around the sex-determining M-locus of the dioecious plant asparagus. Mol Genet Genomics 278: 22134.

Tennessen J a, Govindarajulu R, Liston A, Ashman T-L (2013). Targeted sequence capture provides insight into genome structure and genetics of male sterility in a gynodioecious diploid strawberry, Fragaria vesca ssp. bracteata (Rosaceae). G3 (Bethesda) 3: 134151.

Ueno H, Urasaki N, Natsume S, Yoshida K, Tarora K, Shudo A, et al. (2014). Genome sequence comparison reveals a candidate gene involved in malehermaphrodite differentiation in papaya (Carica papaya) trees. Mol Genet Genomics 290: 661670.

Veuskens J, Marie D, Brown SC, Jacobs M, Negrutiu I (1995). Flow sorting of the Y sex chromosome in the dioecious plant Melandrium album. Cytometry 21: 36373.

Vitte C, Estep MC, Leebens-Mack J, Bennetzen JL (2013). Young, intact and nested retrotransposons are abundant in the onion and asparagus genomes. Ann Bot 112: 8819.

Wai CM, Moore PH, Paull RE, Ming R, Yu Q (2012). An integrated cytogenetic and physical map reveals unevenly distributed recombination spots along the papaya sex chromosomes. Chromosome Res.

Wang J, Na J, Yu Q, Gschwend AR, Han J, Zeng F (2012). Sequencing papaya X and Y h chromosomes reveals molecular basis of incipient sex chromosome evolution.

Warmke and Blakeslee (1939). Effect of polyploidy upon the sex mechanism in dioecious plants. Genetics 24:88-89.

Warmke, H (1946). Sex determination and sex balance in Melandrium. Am. J. Bot. 33:648-660.

Wehrli, L (1892). Uber einen Fall von "vollstandiger Verweiblichung" der mannlichen Katzchen von Corylus Avellana. Flora 76:245-264.

Weingartner L, Moore RC (2012). Contrasting Patterns of X/Y Polymorphism Distinguish Carica papaya from Other Sex Chromosome Systems. Mol Biol Evol.

Westergaard, M. (1940). Studies on cytology and sex determination in polyploid froms of Melandrium album. Dansk. botan. Arkiv. 10:1-131.

Westergaard, M (1946). Aberrant Y-chromosomes and sex expression in Melandrium album. Hereditas 32:419-443.

Westergaard, M (1948). The relationship between chromosome constitution and sex in the offspring of triploid Melandrium. Hereditas 34:257-279. Whipple CJ, Ciceri P, Padilla

CM, Ambrose B a, Bandong SL, Schmidt RJ (2004). Conservation of B-class floral homeotic gene function between maize and Arabidopsis. Development 131: 608391.

Wilson Z a, Song J, Taylor B, Yang C (2011). The final split: the regulation of anther dehiscence. J Exp Bot 62: 163349.

Winge, O (1923). On sex chromosomes, sex determination and preponderance of females in some dioecious plants. Compt. rend. trav. lab. Carslerg 15:1-26.

Winge, O (1929). On the nature of the sex chromosomes in Humulus. Hereditas 12:53-63.

Winge, O (1931). X- and Y-linked inheritance in Melandrium. Hereditas 15:127-165.

Winge, O (1937). Goldschmidts theory of sex determination in Lymantria. Journal of Genetics 34:81-89.

Yu Q, Navajas-Prez R, Tong E, Robertson J, Moore PH, Paterson AH, et al. (2008). Recent Origin of Dioecious and Gynodioecious Y Chromosomes in Papaya. Trop Plant Biol 1: 4957.

Zhang W, Wang X, Yu Q, Ming R, Jiang J (2008). DNA methylation and heterochromatinization in the male-specific region of the primitive Y chromosome of papaya. Genome Res 18: 193843.

Chapter 2

Retrotransposon proliferation in *Asparagus* coincident with the evolution of dioecy

Alex Harkess, Francesco Mercati, Loredana Abbate, Michael McKain, J. Chris Pires, Tea Sala, Francesco Sunseri, Agostino Falavigna, Jim Leebens-Mack. Submitted to *G3: Genes, Genomes, Genetics*. April 12, 2016.

2.1 Abstract

Current phylogenetic sampling reveals that dioecy and an XY sex chromosome pair evolved once or possibly twice in the genus *Asparagus*. Although there appear to be some lineage-specific polyploidization events, the base chromosome number of 2n=2x=20 is relatively conserved across the Asparagus genus. Regardless, dioecious species tend to have larger genomes than hermaphroditic species. Here we test whether this genome size expansion in dioecious species is related to a polyploidization and subsequent chromosome fusion or retrotransposon proliferation in dioecious species. We first estimate genome sizes or use published values for four hermaphrodites and four dioecious species distributed across the phylogeny and show that dioecious species typically have larger genomes than hermaphroditic species. Utilizing a phylogenomic approach we find no evidence for ancient polyploidization contributing to increased genome sizes of sampled dioecious species. We do find support for an ancient whole genome duplication event predating the diversification of the *Asparagus* genus. Repetitive DNA content of the four hermaphroditic and four dioecious species was characterized based on randomly sampled whole genome shotgun sequencing and common elements were annotated. Across our broad phylogenetic sampling, Ty-1 Copia retroelements in particular have undergone a marked proliferation in dioecious species. In the absence of a detectable whole genome duplication event, retrotransposon proliferation is the most likely explanation for the precipitous increase in genome size in dioecious Asparagus species.

2.2 Introduction

Fewer than 10% of flowering plant species are dioecious, the condition where individual plants are distinctly male or female (Ainsworth 2000). Gender in some dioecious plants can be governed by a sex chromosome pair, such as in papaya (*Carica papaya)*, white campion (*Silene latifolia*), persimmon, *Rumex*, and garden asparagus (*Asparagus officinalis*)

(Telgmann-Rauber et al. 2007; Ming et al. 2011; Akagi et al. 2014; Hough et al. 2014). The evolution of a distinct sex chromosome pair is hypothesized to necessitate the formation of a non-recombining region between the X and Y (or Z and W) where tightly linked sex determination genes reside (Charlesworth and Charlesworth 1978). Given the repeated and independent evolution of dioecy across the angiosperm phylogeny, the transition from autosome to sex chromosome is undoubtedly governed by different sex determination genes and evolutionary processes and consequently must be viewed in a species-specific context. Despite this diversity in sex chromosome evolution across the angiosperms, two particularly interesting associations can be seen in some dioecious systems coincident with variation in sexual system: the proliferation of repetitive elements and the occurrence of one or multiple whole genome duplication (polyploidy) events.

As a consequence of restricted recombination between regions of sex chromosomes, repetitive elements tend to persist and replicate in an unbalanced way, preferentially accumulating on hemizygous regions of Y and W chromosomes. Transposable elements can be broadly classified primarily by their means of transposition (Wicker et al. 2007); class I retrotransposons move by a copy and paste mechanism and replicate through an mRNA intermediate which ultimately results in a net increase of the elements copy number, whereas class II transposable elements move through a DNA intermediate in a cut and paste fashion. Since Class I retrotransposons can range in length from 5 to 20 kilobases or longer, their uncontrolled proliferation can lead to drastic and rapid changes in genome size. This accumulation, especially of active retroelements, can be clearly seen when comparing the relatively young papaya X and hermaphrodite-specific region of the Y (HSY) (VanBuren and Ming 2013). Unbalanced accumulation of transposons and other repetitive elements, paired with the inability for recombination to remove them along with other deleterious mutations, is likely a major factor that leads to the initial physical expansion and genic degeneration of a young, partially non-recombining Y or W chromosome. Transposons have also been directly im-

plicated in the evolution of sex determination genes through disruption of gene expression. In melon (*Cucumis melo*), a class II hAT DNA transposon insertion is responsible for promoter hypermethylation and transcriptional repression of the zinc-finger transcription factor CMWIP1, heritably inducing the transition from monoecy to gynodioecy (Boualem et al. 2008).

Equally puzzling is the association between polyploidy and variation in sexual system. In the *Fragaria* genus, at least four independent whole genome duplication events have occurred across all major clades, leading to an abundance of polyploid dioecious species phylogenetically placed as sister to dioecious hermaphrodites (Rousseau-Gueutin et al. 2009; Ashman et al. 2013). Losses of dioecy have also been associated with an increase in ploidy, such as in one clade of *Mercurialis* (Krhenbhl et al. 2002). The mechanisms that potentially relate whole genome duplication events to the evolution of sexually dimorphic populations are variable and poorly understood, though, again owing to the extreme complexity and species-specific nature of sex chromosome and dioecy.

Garden asparagus (*Asparagus officinalis* L.) is a particularly useful dioecious plant for studying sex chromosome evolution given that it has cytologically homomorphic X and Y sex chromosomes, suggesting that the transition from hermaphroditism to dioecy was recent (Telgmann-Rauber et al. 2007; Kubota et al. 2012). Coincident with the evolution of dioecy was a range shift from South Africa into North Africa, Europe, and Asia (Stajner et al. 2002; Kubota et al. 2012; Norup et al. 2015). It was previously reported that dioecious *Asparagus* species tend to have larger genomes than hermaphrodites, but there was no evidence supporting a whole genome duplication event that separates the dioecious species from the hermaphrodites (Kuhl et al. 2005). The base chromosome number of 2n=2x=20 is generally consistent across the genus except for instances of very recent polyploidization in some species (Kanno and Yokoyama 2011), which suggests two major hypotheses that could lead to an increase in genome size: one possibility is that a whole genome duplication

occurred in the last common ancestor of all dioecious species, followed by a chromosome fusion or reduction, and another possibility is that repetitive DNA has proliferated to drive the increase in the genome sizes of dioecious species. Here, we test both hypotheses by first leveraging transcriptome assemblies for one hermaphroditic and one dioecious species to identify the relative timing of whole genome duplication events in the genus *Asparagus*. We then use shallow Roche 454 whole genome shotgun sequencing from four hermaphrodites and four dioecious species that are sampled from the phylogeny to assess the repetitive element content of each species in relation to its genome size.

2.3 Results and Discussion

Genome size increase in dioecious Asparagus

Genome sizes and ploidy are highly variable across the order Asparagales, with 1C values ranging from 0.3-88.2 pg (Leitch et al. 2010). Diploid dioecious *Asparagus* species have been reported as having genome sizes nearly double the size of diploid hermaphroditic congeners (Stajner et al. 2002, Fukuda et al. 2005, Kubota et al., 2012). We first confirmed this by generating or supplementing published genome size estimations for eight *Asparagus* species, four hermaphrodites and four dioecious species, sampled across all major clades of the *Asparagus* phylogeny (Kubota et al. 2012) (Table 1). All individuals have been documented as diploids (2n=2x=20) except for *A. maritimus*, a hexaploid (Stajner et al. 2002; Kanno and Yokoyama 2011). Flow cytometry-derived genome sizes (pg/1C) for hermaphrodites range from 0.72 to 1.06, while dioecious species range from 1.09 to 1.37. Dioecious species tend to have larger genome sizes than hermaphroditic species (Unpaired t test, $p = 0.0173$). An outlier is the hermaphrodite *Asparagus asparagoides* with a relatively large genome size (1C = 2.40; Dixons Q test, $p = 0.074$).

Table 2.1: Genome sizes, 454 pyrosequencing and repetitive element clustering

Species	Sexual System	pg/nucleus (mean SD)	1C value	Raw Reads	Filtered Reads	Clustered Reads (%)
A. officinalis	Dioecious	2.74 ± 0.044	1.37	29,677	26,525	54.40
A. maritimus	Dioecious	7.87 ± 0.204a	1.31	49,616	45,036	53.70
A. aphyllus	Dioecious	2.49 ± 0.007	1.25	47,322	42,808	58.90
A. stipularis	Dioecious	2.17 ± 0.005	1.09	30,405	27,911	56.40
A. falcatus	Hermaphrodite	2.11 ± 0.007	1.06	26,836	24,304	60.40
A. virgatus	Hermaphrodite	1.66 ± 0.055	0.83	45,043	41,053	45.00
A. pyrimidalis	Hermaphrodite	1.44 ± 0.037a	0.72	56,197	51,293	53.80
A. asparagoides	Hermaphrodite	4.80 ± 0.062	2.40	41,952	37,435	59.20
Sum				247,755	224,804	
Average				41,293	37,467	

No evidence for a dioecy-specific polyploidy event

We employed a phylogenomics approach to test whether a whole genome duplication event separates the dioecious and hermaphroditic species in *Asparagus*. Transcriptome assemblies were generated for two species sampled broadly across the phylogeny: a basal diploid hermaphrodite (*A. asparagoides*; 2n=2x=20) and diploid dioecious garden asparagus (*A. officinalis*; 2n=2x=20). Intraspecific paralog pairs and interspecific orthologous gene pairs were inferred to generate *Ks* (synonymous substitution rate) distributions and assess the relative timing of whole genome duplication event relative to speciation events (Blanc and Wolfe 2004, Cui et al. 2006, McKain et al. 2012, Doyle and Egan 2010). Despite being an outlier in terms of genome size, *A. asparagoides* was utilized for the comparison given that it is a basal member of the genus, shares the same diploid chromosome count as *A. officinalis*, and that transcriptome-based Ks analyses are independent of genome size.

Transcriptome assembly and translation results for the two species are presented in Supplementary Table 1. One distinct, shared polyploidization event (Ks = 0.5) was inferred

from the *Ks* frequency distribution of paralogous pairs in both *Asparagus* species (Figure 1; Supplementary Table 2). Additionally, orthologous pairs exhibit a *Ks* peak close to 0, representing low divergence and suggestive of recent diversification of species and/or similar mutation rates. Comparison of orthologs and paralogs demonstrates that at least one detectable genome duplication event occurred before the diversification of the *Asparagus* genus (Figure 2.1).

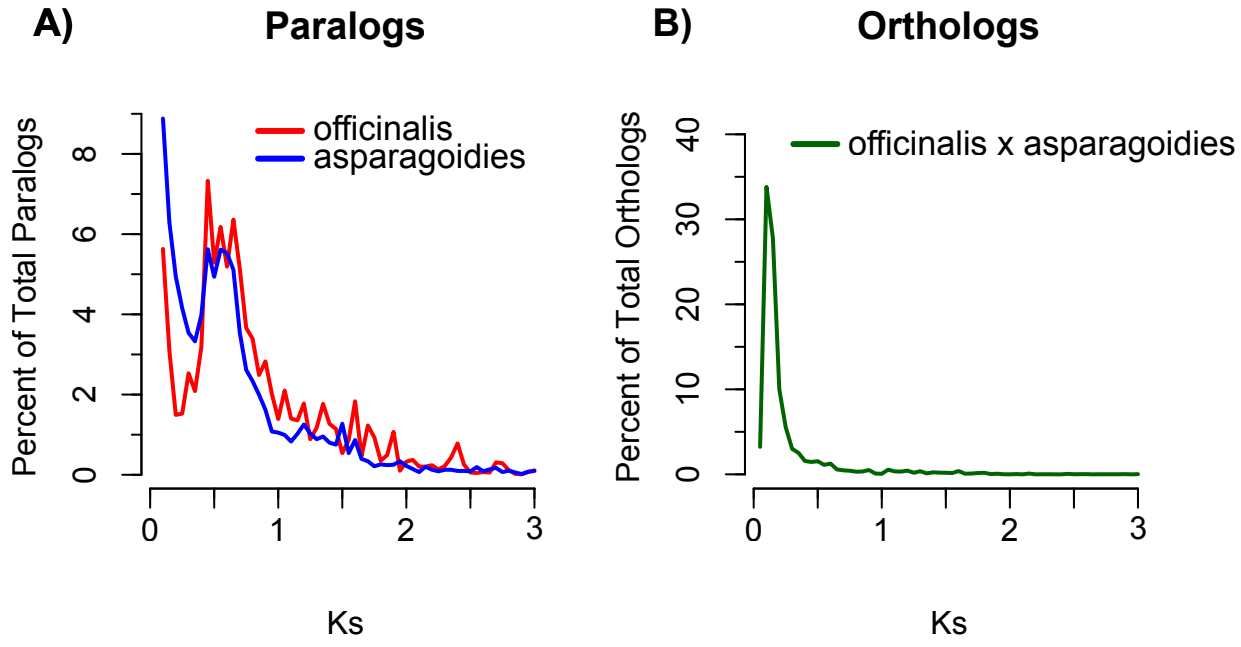

Figure 2.1: Transcriptome-based Ks frequency distributions for A) paralogous and B) orthologous pairs of dioecious *A. officinalis* and hermaphroditic *A. asparagoides*. Paralogous and orthologous Ks distributions suggest a shared whole genome duplication event at Ks 0.5 that occurred before the diversification of the Asparagus genus.

A major limitation with *Ks* analyses is that more recent duplication events are difficult to detect (Blanc and Wolfe 2004; Cui et al. 2006). This issue is exacerbated when using *de novo* transcriptome assemblies, where recently duplicated paralogs can be computationally mistaken as alleles and incorrectly collapsed into a single transcript sequence during the assembly process. Given that there are no current age estimates for the divergence of the

Asparagus lineage, we cannot exclude the possibility that a more recent duplication event, such as one that may have co-occurred with the evolution of dioecy, could be undetectable by transcriptome data. However, such a whole genome duplication event would need to be a followed by multiple chromosome fusion or loss events to reduce the chromosome number back to 2n=2x=20 found in most karyotyped dioecious *Asparagus* species (Kanno and Yokoyama 2011). Taken together, the overlapping paralogous and orthologous Ks distributions do not support the hypothesis that a whole genome duplication event occurred coincident with the evolution of dioecy.

Lineage-specific expansion of transposable elements

Given the lack of evidence that ancient polyploidy was responsible for the larger genome sizes of dioecious *Asparagus* species relative to hermaphroditic species, we assessed the alternative hypothesis that the genome size increase in dioecious species was due to transposon amplification. We utilized whole genome shotgun sequence reads to assess the repetitive content of hermaphrodite and dioecious *Asparagus* species using the RepeatExplorer Galaxy server (http://www.repeatexplorer.org). Briefly, this method utilizes all-by-all read comparisons followed by Louvain clustering (Blondel et al. 2008) to place reads into unbiased clusters of putative high copy elements, followed by a RepeatMasker annotation and cap3 assembly (Huang and Madan 1999).

A total of 327,048 raw reads were sequenced for the eight genomes using Roche 454 FLX chemistry, with genome coverages ranging from 0.0051X to 0.0234X (Supplementary Table 3). After removing duplicate reads that were likely clonal, 321,865 total reads remained for analysis. To improve clustering, we then removed reads less than 100nt long, yielding a filtered set of 296,365 reads (mean = 37,047 reads per species) with a mean length of 321nt. The filtered set of reads was concatenated and clustered using the RepeatExplorer pipeline, placing 162,435 reads into 29,643 repetitive element clusters (Table 1). Repetitive element

clusters were filtered by read count, requiring at least 0.01% of the total filtered reads (30 reads), amounting to 336 clusters for downstream analysis. These clusters were annotated against a custom RepeatMasker database generated with additional data for dioecious *A. officinalis*. For a given cluster of repetitive elements, the repetitive fraction of each species genome was calculated as the number of a given species reads in a cluster divided by the total number of reads sequenced for that species, represented as a percentage.

Multidimensional Scaling (MDS) analysis of the genomic proportions for all clusters shows that dioecious and hermaphroditic species form two distinct clusters (Figure 2). In general, Gypsy and Copia retrotransposons dominate the genomic landscape for all sampled *Asparagus* species (Figure 3). In all four dioecious species, Gypsy retrotransposons occupy a larger percentage of each genome than in the hermaphrodites, although Copia elements have distinctly expanded in the dioecious species (Figure 2). This suggests that both Gypsy and Copia elements increased in copy number in the dioecious species, and the proliferation of Copia elements was a more substantial contributor to the expansion of dioecious genome sizes.

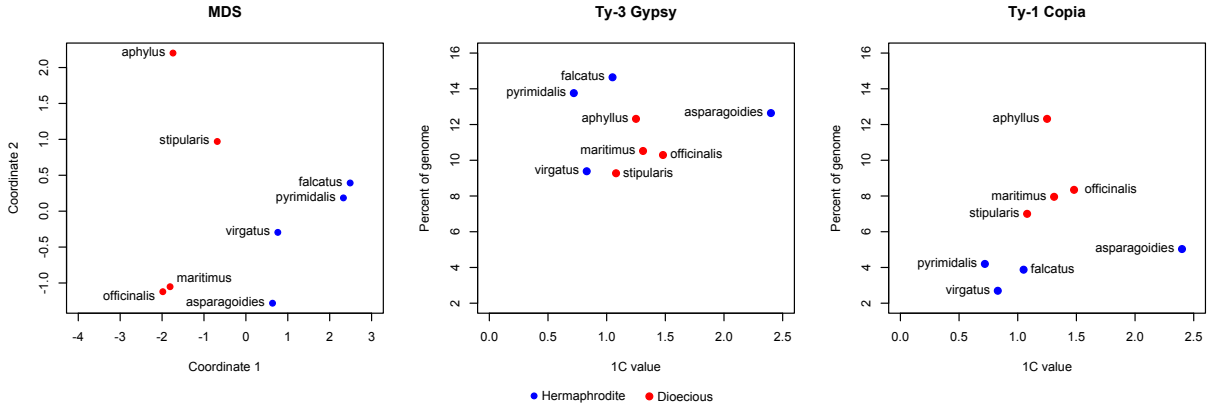

Figure 2.2: Multidimensional scaling (MDS) and relationship of genome size to Gypsy and Copia retroelement content for both dioecious and hermaphroditic genomes. Blue dots represent hermaphroditic species while red dots represent dioecious species

We identified 46 repetitive element clusters that were private to the dioecious species

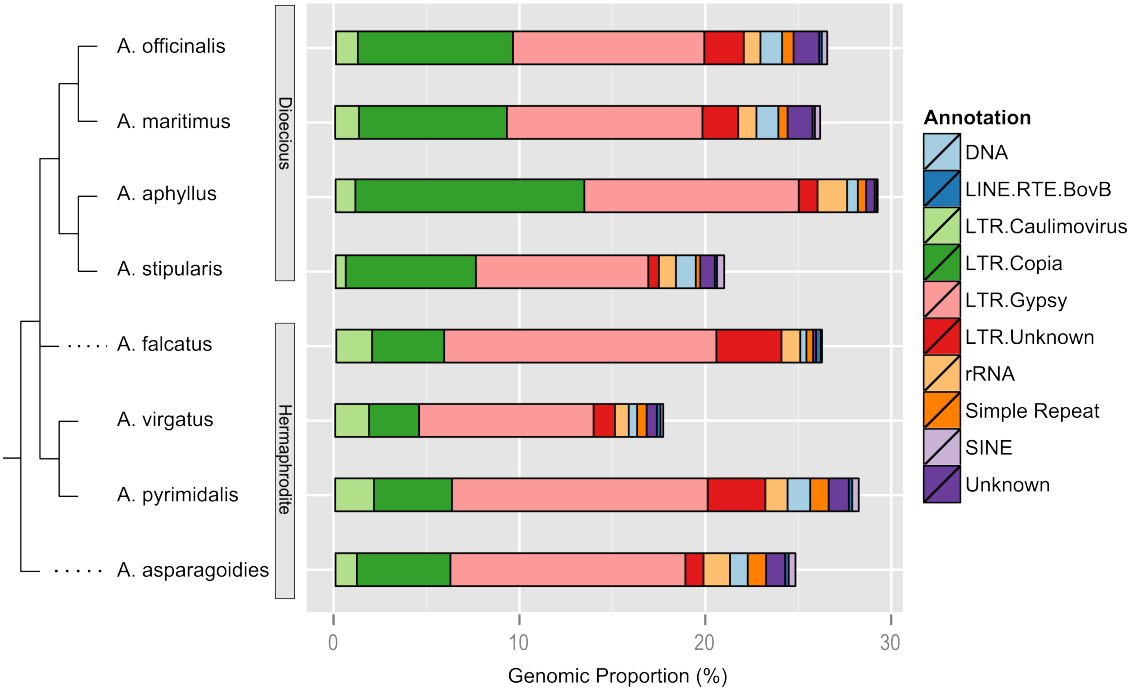

Figure 2.3: Cladogram of Asparagus species relationships with high copy repetitive elements. High copy elements refer to clusters with greater than 0.01% of the total read count in the multispecies analysis, able to be most confidently annotated against the custom *A. officinalis* repetitive element database. DNA transposons from several families were collapsed into a single annotation class.

and 37 clusters that were private to all hermaphroditic species. In the dioecious species, 26 clusters were Gypsy and 7 clusters were Copia, whereas in the hermaphroditic species, 12 clusters were Gypsy and 11 clusters were Copia. This suggests that there is active turnover of transposable elements in the Asparagus genus, perhaps coincident with the evolution of dioecy and a sex chromosome. Additionally, it is possible that a small number of Copia elements may be largely responsible for the genome size expansion in dioecious species, but this would require whole genome assemblies and annotations as RepeatExplorer is limited in ability to finely delimit elements.

One caveat for performing a single repeat clustering analysis including all species (as opposed to individually analyzing each species) is that low frequency or moderately diverged sequences from phylogenetically distant species may not cluster. Additionally, there could be less power for detecting species-specific transposon family proliferations. Consequently, these estimates of repetitive element content are certainly underestimates of the total proportion of repetitive element content in each species genome. To understand the level of difference in these two analysis types, we generated 893,623 additional 454 shotgun reads (mean length 526nt) for a mature double haploid YY *A. officinalis* individual and ran the RepeatExplorer pipeline with this single species. The repeat content was estimated at 71.1%, much greater than the 54.4% that was estimated by concatenating eight species in a single analysis. This result suggests that the genomic proportions of transposons estimated through multispecies read clustering in this study should be interpreted as being underestimates, biased towards high copy elements with lower divergence between species, and used mostly for comparisons of high copy element percentages between species. The advantage of this analysis is that direct comparisons for a given transposon cluster can be assessed across all species.

The method of repeat quantification and sequence read type also largely affects the estimated proportion of repetitive elements. Repetitive element content has previously been estimated for *A. officinalis* in at least three separate studies. Vitte et al. (2013) directly an-

notated garden asparagus Bacterial Artificial Chromosome (BAC) assemblies for transposon content. By comparing the sequence alignment identity of intact LTRs from retroelements and applying a clock estimation from rice retroelement divergence (Ma and Bennetzen 2004), Vitte et al. estimated that the majority of the asparagus genome is comprised of young, recently inserted (<6 million years ago) and nested retroelements. Li et al. (2014) took a high-throughput sequencing approach and inferred that the garden asparagus genome is 53% repetitive by de novo assembling genomic paired end 100nt Illumina reads into a 400Mbp assembly with a scaffold N50 of 1504nt. Hertweck (2013) took a similar approach with 80 bp Illumina read data and independently estimated 47% of the garden asparagus genome as comprising repetitive elements. We hypothesize that our much higher estimation of 71.1% repetitive content is largely due to the increased detection power coming with longer 454 reads relative to 80-100 bp Illumina reads and our use of RepeatExplorers unique assembly-free, graph-based clustering and annotation of individual long reads.

Transposon clustering yields phylogenetic signal

Clustering of the genomic proportions for the 100 largest Gypsy and Copia retrotransposon clusters also reveals phylogenetic signal in the data (Figure 4). The deepest branch divides the hermaphroditic and dioecious species from each other, and all species are paired with their closest phylogenetic neighbor given the current phylogeny and sampling from Kubota et al. (2012), with exception of *A. asparagoides*. The genomic proportions of repetitive elements have been used to identify phylogenetic signal in several plant species with species relationships that have been difficult to resolve with traditional low copy gene sequencing (Dodsworth et al. 2014). While our clustering approach is much less statistically robust compared to Dodsworth et al. (2014), here we show a complementary analysis that yields similar results using high copy transposon clusters.

Recently, Norup et al. (2015) proposed two origins of dioecy within *Asparagus*, an al-

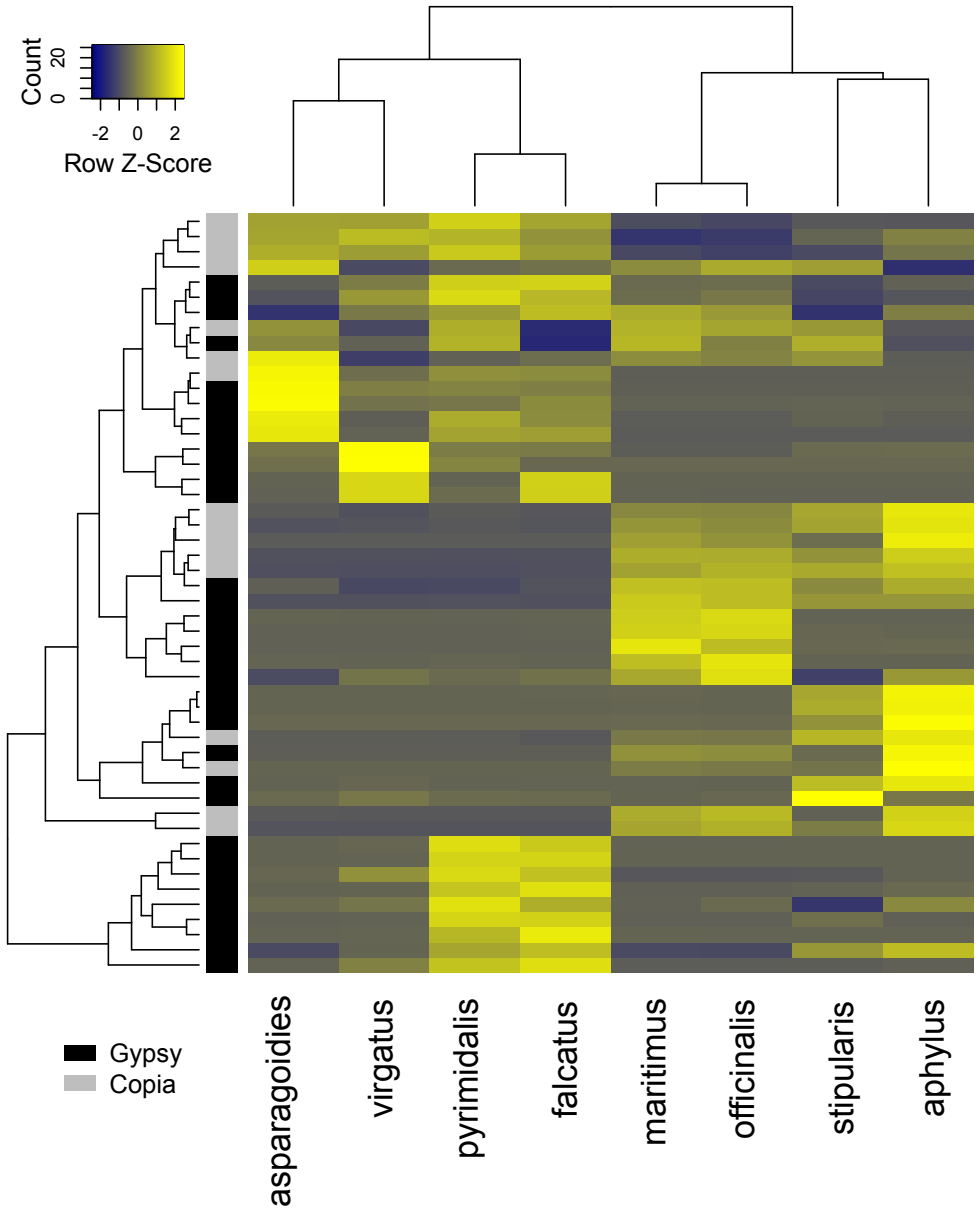

Figure 2.4: Heatmap clustering of 100 largest Gypsy or Copia element clusters. Rows represent individual clusters, annotated as Gypsy (black) and Copia (grey).

ternative to the previously hypothesized single origin (Kubota et al. 2012). Our sampling includes species derived from both of the hypothesized origins of dioecy from Norup et al. (2015), which indicates that dioecy evolved in one clade that includes *A. officinalis* and *A. maritimus*, as well as another clade that includes *A. stipularis* and *A. aphyllus*. In the case of multiple origins of dioecy, without hermaphroditic outgroup species for each origin, our limited sampling does not allow us to describe the potentially different repetitive element radiations.

Several mechanisms exist to remove repetitive element DNA from obese, transposon-dense genomes. One mechanism is the formation of small chromosomal deletions by illegitimate recombination, which usually occurs by slip-strand mispairing or non-homologous end joining (NHEJ) (Hawkins et al. 2009). Another mechanism is unequal intra-strand homologous recombination between the directly repeated LTRs of retrotransposons, resulting in a solo LTR remaining (Devos et al. 2002). The half-life for retrotransposon occupancy seems to be relatively short in several plant species (Ma et al. 2004; Wang and Liu 2008; Charles et al. 2008), suggesting that these removal mechanisms are actively purging transposons. In most angiosperm genomes, Gypsy and Copia-type retrolements typically dominate the repetitive landscape. Extreme cases of transposon purging and genome compression are evidenced by the *Utricularia gibba* genome, comprising only about 3% transposable elements (Ibarra-Laclette et al. 2013). Compared to Asparagus, similar cases of lineage-specific transposon expansion have been found in the Asteraceae, where a small number of Gypsy families have been expanding since the branch leading to the Asteraceae (Staton and Burke 2015). We hypothesize that the proliferation of both Gypsy and Copia retroelements in dioecious lineages may be explained by two coincident events in *Asparagus* evolution: range expansion and the origin of dioecy. As others have documented, range expansion out of South Africa is associated with a transition of ancestrally hermaphroditic Asparagus species to dioecy within a clade distributed across Europe and Asia. (Stajner et al. 2002; Kuhl et al. 2005; Kanno

and Yokoyama 2011; Kubota et al. 2012; Norup et al. 2015). Founder populations formed during this range expansion with small effective population sizes may have been especially susceptible to weakly deleterious transposon proliferation due to the reduced strength of purifying selection relative to populations with large effective sizes (Lynch et al. 2011). The origin of sex chromosomes alone may have promoted proliferation of retrotransposons. Suppressed recombination within the region of the sex chromosomes where gender determination genes reside in the first dioecious Asparagus species may have harbored active retrotransposons. Young and old plant Y chromosomes in *Silene* and papaya can be replete with or entirely missing tandem arrays and LTR retroelements that distinguish them from both the X and other autosomes (Pritham et al. 2003; Filatov et al. 2009; VanBuren and Ming 2013). Recombination is selected against in these regions of a sex chromosome given that recombination could break apart genes influencing male and female function, leading to the formation of neuters. This selection on young sex chromosomes may drive the maintenance and proliferation of LTR retrotransposons.

2.4 Methods

Flow cytometry genome size estimation

The genome sizes of *A. officinalis, A. virgatus* and *A. asparagoides* were estimated by flow cytometry at the Benaroya Research Institute at Virginia Mason in Seattle, Washington. Nuclei isolations from a single mature leaf were analyzed in three technical replicates for each species. The genome sizes of *A. aphyllus, A. stipularis*, and *A. falcatus* were estimated by flow cytometry using the known genome size of *A. officinalis* (1C-value = 1.37 pg) as a reference standard. Ten plants for each species, grown in greenhouse, were sampled and three randomly selected plants were analysed. The analysis was carried out with the Partec PAS flow cytometer (Partec, http://www.partec.de/), equipped with a mercury lamp. Fully

expanded leaves (0.1 g) were chopped in 300 l nuclei extraction buffer (CyStain ultraviolet Precise P Nuclei Extraction Buffer; Partec, Mnster, Germany) for 30-40 s. The solution was filtered through a 30 mm Cell-Trics disposable filter (Partec), and 1.2 ml of staining solution containing 4,6-diamidino-2- phenylindole was added. The relative fluorescence intensity of stained nuclei was measured on a linear scale and 4,000-5,000 nuclei for each sample were analysed (Galbraith et al., 1998). DNA content histograms were generated using the Partec software package (FloMax). Given that the X and Y chromosomes in garden asparagus (*Asparagus officinalis*) are cytologically homomorphic (Deng et al. 2012) representing a lack of degeneration and the relatively young age of the Y, we did not discern between potential sex differences in the dioecious species.

Transcriptome-based *Ks* analysis

Transcriptomes from dioecious *A. officinalis* and hermaphroditic *A. asparagoides* were used to infer a putative whole genome duplication event in the genus *Asparagus*. The transcriptome assembly and translation for *A. officinalis* was taken from Harkess et al. (2015) (http://datadryad.org/resource/doi:10.5061/dryad.92c60). We generated leaf RNA-Seq for *A. asparagoides* by first isolating total RNA from mature leaf tissue using a Qiagen RNeasy Plant Mini kit. Total RNA quantity and quality was assessed using an RNA Nano chip on the Bioanalyzer 2100. A sequencing library was generated using the TruSeq RNA Library Prep Kit v2 (Illumina) according to manufacturers instructions using 1ug of total RNA input. The library was sequenced with paired end 100nt reads on an Illumina HiSeq2000, generating 55,686,513 read pairs (nearly 11 gigabases of data). Reads were quality trimmed using Trimmomatic (v0.32), removing sequencing adapters and clipping 3 and 5 read ends with a quality score lower than Phred 5. Cleaned reads were assembled using Trinity (r20140717) with default parameters. We filtered transcript isoforms with low support by removing isoforms with less than 0.01% of the Trinity gene subcomponent read support. Coding sequence

and peptide translations were inferred using TransDecoder (r20140704) with default settings. Raw sequence reads for *A. asparagoides* has been deposited under BioProject (ID here after acceptance).

Using a pipeline from McKain et al. (https://github.com/mrmckain/FASTKs), we first identified putative paralogs in each filtered transcriptome assembly using all vs. all blastn (1e-40 cutoff). Peptide sequences for hit pairs longer than 100 amino acids were aligned using MUSCLE (v3.8.31), then codon alignments were inferred using PAL2NAL (v13) (Suyama et al. 2006). For each paralog pair, Ks was calculated using CodeML in PAML (Yang 2007) (v4.8).

454 pyrosequencing and transposon quantification

Whole genomic DNA was extracted from four hermaphroditic and four dioecious species using a CTAB method (Doyle and Doyle 1987). Sequencing libraries were prepared using the Roche 454 GS FLX Titanium library preparation kit according to manufacturer instructions. Raw reads were first de-duplicated to remove probable emulsion PCR sequencing artifacts, then filtered to remove reads less than 100nt long. Read names from all species were first prepended with a unique species identifier and concatenated. The RepeatExplorer (v0.9.7.4) pipeline (http://www.repeatexplorer.org) was then used to cluster, assemble, and annotate all filtered shotgun reads against a custom garden asparagus RepeatMasker database (see below) using otherwise default settings. Clustering and heatmap production of the 100 largest transposon clusters was performed using heatmap.2 in the gplot package in R (v3.2.1) using default settings; a distance matrix was generated using Euclidean distances, and hierarchical clustering was performed using complete clustering.

To improve the annotations of repetitive element clusters generated through the RepeatExplorer pipeline, we generated a much higher coverage of 454 reads for A. officinalis to build a comprehensive database of annotated exemplar repeats for the genus. The custom

garden asparagus RepeatMasker database was generated using similar methodology. A total of 893,623 454 FLX Titanium reads were generated from leaf tissue of a doubled haploid (YY) garden asparagus individual. Reads were more stringently filtered to a 150nt minimum length. The same version of RepeatExplorer was then run, and the resulting cap3 consensus assemblies for each cluster were annotated using RepeatClassifier, part of the RepeatModeler (v1.0.8) suite, with default settings. A total of 22,361 sequences greater than 150nt in length and with annotations were retained and are available at (Dryad link here after acceptance). Raw 454 shotgun sequence data for all individuals have also been deposited in Dryad.

2.5 References

Ainsworth, C., 2000 Boys and Girls Come Out to Play: The Molecular Biology of Dioecious Plants. Ann. Bot. 86: 211221.

Akagi, T., I. M. Henry, R. Tao, and L. Comai, 2014 A Y-chromosomeencoded small RNA acts as a sex determinant in persimmons. Science. 346: 646650.

Ashman, T.-L., a Kwok, and B. C. Husband, 2013 Revisiting the dioecy-polyploidy association: alternate pathways and research opportunities. Cytogenet. Genome Res. 140: 24155.

Blanc, G., and K. H. Wolfe, 2004 Functional divergence of duplicated genes formed by polyploidy during Arabidopsis evolution. Plant Cell 16: 167991.

Blondel, V. D., J. Guillaume, R. Lambiotte, and E. Lefebvre, 2008 Fast unfolding of communities in large networks. J. Stat. Mech. Theory Exp. 2008: P10008.

Boualem, A., M. Fergany, R. Fernandez, C. Troadec, A. Martin et al., 2008 A conserved mutation in an ethylene biosynthesis enzyme leads to andromonoecy in melons. Science 321: 836838.

Charles, M., H. Belcram, J. Just, C. Huneau, A. Viollet et al., 2008 Dynamics and dif-

ferential proliferation of transposable elements during the evolution of the B and A genomes of wheat. Genetics 180: 107186.

Charlesworth, B., and D. Charlesworth, 1978 A model for the evolution of dioecy and gynodioecy. Am. Nat. 112: 975997.

Cui, L., P. K. Wall, J. H. Leebens-Mack, B. G. Lindsay, D. E. Soltis et al., 2006 Widespread genome duplications throughout the history of flowering plants. Genome Res. 16: 738749.

Deng, C.-L. L., R.-Y. Y. Qin, N.-N. N. Wang, Y. Cao, J. Gao et al., 2012 Karyotype of asparagus by physical mapping of 45S and 5S rDNA by FISH. J. Genet. 91: 20912.

Devos, K. M., J. K. M. Brown, and J. L. Bennetzen, 2002 Genome size reduction through illegitimate recombination counteracts genome expansion in Arabidopsis. Genome Res. 12: 10759.

Dodsworth, S., M. W. Chase, L. J. Kelly, I. J. Leitch, J. Macas et al., 2014 Genomic Repeat Abundances Contain Phylogenetic Signal. Syst. Biol. 64: 112126.

Doyle, J. J., and J. L. Doyle, 1987 A rapid DNA isolation procedure for small quantities of fresh leaf tissue. Phytochem. Bull. 19: 1115.

Doyle, J. J., and A. N. Egan, 2010 Dating the origins of polyploidy events. New Phytol. 186: 7385.

Filatov, D. a, E. C. Howell, C. Groutides, and S. J. Armstrong, 2009 Recent spread of a retrotransposon in the Silene latifolia genome, apart from the Y chromosome. Genetics 181: 8117.

Harkess, A., F. Mercati, H. Y. Shan, F. Sunseri, A. Falavigna et al., 2015 Sex-biased gene expression in dioecious garden asparagus (Asparagus officinalis). New Phytol. 207: 883892.

Hawkins, J. S., S. R. Proulx, R. a Rapp, and J. F. Wendel, 2009 Rapid DNA loss as a counterbalance to genome expansion through retrotransposon proliferation in plants. Proc. Natl. Acad. Sci. U. S. A. 106: 178116.

Hertweck, K. L., 2013 Assembly and comparative analysis of transposable elements from low coverage genomic sequence data in Asparagales1. Genome 56: 487494.

Hough, J., J. D. Hollister, W. Wang, S. C. H. Barrett, and S. I. Wright, 2014 Genetic degeneration of old and young Y chromosomes in the flowering plant Rumex hastatulus. Proc. Natl. Acad. Sci. U. S. A. 2014: 16.

Huang, X., and A. Madan, 1999 CAP3: A DNA sequence assembly program. Genome Res. 9: 86877.

Ibarra-Laclette, E., E. Lyons, G. Hernndez-Guzmn, C. A. Prez-Torres, L. Carretero-Paulet et al., 2013 Architecture and evolution of a minute plant genome. Nature 498: 948.

Kanno, A., and J. Yokoyama, 2011 Wild Crop Relatives: Genomic and Breeding Resources (C. Kole, Ed.).

Krhenbhl, M., Y. M. Yuan, and P. Kpfer, 2002 Chromosome and breeding system evolution of the genus Mercurialis (Euphorbiaceae): Implications of ITS molecular phylogeny. Plant Syst. Evol. 234: 155169.

Kubota, S., I. Konno, and A. Kanno, 2012 Molecular phylogeny of the genus Asparagus (Asparagaceae) explains interspecific crossability between the garden asparagus (A. officinalis) and other Asparagus species. Theor. Appl. Genet. 124: 34554.

Kuhl, J. C., M. J. Havey, W. J. Martin, F. Cheung, Q. Yuan et al., 2005 Comparative genomic analyses in Asparagus. Genome 1060: 10521060.

Leitch, I. J., J. M. Beaulieu, M. W. Chase, A. R. Leitch, and M. F. Fay, 2010 Genome Size Dynamics and Evolution in Monocots. J. Bot. 2010: 118.

Li, S. F., W. J. Gao, X. P. Zhao, T. Y. Dong, C. L. Deng et al., 2014 Analysis of transposable elements in the genome of Asparagus officinalis from high coverage sequence data. PLoS One 9: 18.

Lynch, M., L.-M. Bobay, F. Catania, J.-F. Gout, and M. Rho, 2011 The repatterning of eukaryotic genomes by random genetic drift. Annu. Rev. Genomics Hum. Genet. 12:

347366.

Ma, J., and J. L. Bennetzen, 2004 Rapid recent growth and divergence of rice nuclear genomes. Proc. Natl. Acad. Sci. U. S. A. 101: 1240412410.

Ma, J., K. M. Devos, and J. L. Bennetzen, 2004 Analyses of LTR-Retrotransposon Structures Reveal Recent and Rapid Genomic DNA Loss in Rice. 860869.

McKain, M. R., N. Wickett, Y. Zhang, S. Ayyampalayam, W. R. McCombie et al., 2012 Phylogenomic analysis of transcriptome data elucidates co-occurrence of a paleopolyploid event and the origin of bimodal karyotypes in Agavoideae (Asparagaceae). Am. J. Bot. 99: 397406.

Ming, R., A. Bendahmane, and S. S. Renner, 2011 Sex chromosomes in land plants. Annu. Rev. Plant Biol. 62: 485514.

Norup, M. F., G. Petersen, S. Burrows, Y. Bouchenak-Khelladi, J. Leebens-Mack et al., 2015 Evolution of Asparagus L. (Asparagaceae): Out-of-South-Africa and multiple origins of sexual dimorphism. Mol. Phylogenet. Evol. 92: 2544.

Pritham, E. J. E., Y. H. Zhang, C. Feschotte, and R. R. V Kesseli, 2003 An Ac-like transposable element family with transcriptionally active Y-linked copies in the white campion, Silene latifolia. Genetics 165: 799807.

Rousseau-Gueutin, M., A. Gaston, A. A??nouche, M. L. A??nouche, K. Olbricht et al., 2009 Tracking the evolutionary history of polyploidy in Fragaria L. (strawberry): New insights from phylogenetic analyses of low-copy nuclear genes. Mol. Phylogenet. Evol. 51: 515530.

tajner, N., B. Bohanec, and B. Javornik, 2002 Genetic variability of economically important Asparagus species as revealed by genome size analysis and rDNA ITS polymorphisms. Plant Sci. 162: 931937.

Staton, S. E., and J. M. Burke, 2015 Evolutionary transitions in the Asteraceae coincide with marked shifts in transposable element abundance. BMC Genomics 16: 623.

Suyama, M., D. Torrents, and P. Bork, 2006 PAL2NAL: Robust conversion of protein sequence alignments into the corresponding codon alignments. Nucleic Acids Res. 34.:

Telgmann-Rauber, A., A. Jamsari, M. S. Kinney, J. C. Pires, and C. Jung, 2007 Genetic and physical maps around the sex-determining M-locus of the dioecious plant asparagus. Mol. Genet. Genomics 278: 22134.

VanBuren, R., and R. Ming, 2013 Dynamic transposable element accumulation in the nascent sex chromosomes of papaya. Mob. Genet. Elements 15.

Vitte, C., M. C. Estep, J. Leebens-Mack, and J. L. Bennetzen, 2013 Young, intact and nested retrotransposons are abundant in the onion and asparagus genomes. Ann. Bot. 112: 8819.

Wang, H., and J.-S. Liu, 2008 LTR retrotransposon landscape in Medicago truncatula: more rapid removal than in rice. BMC Genomics 9: 382.

Wicker, T., F. Sabot, A. Hua-Van, J. L. Bennetzen, P. Capy et al., 2007 A unified classification system for eukaryotic transposable elements. Nat. Rev. 8: 973982.

Yang, Z., 2007 PAML 4: Phylogenetic analysis by maximum likelihood. Mol. Biol. Evol. 24: 15861591.

Chapter 3

Sex-biased gene expression in dioecious garden asparagus

A. Harkess, F. Mercati, H. Y. Shan, F. Sunseri, A. Falavigna, J. Leebens-Mack. 2015. Sex-biased gene expression in dioecious garden asparagus (*Asparagus officinalis*). *New Phytologist* 207: 883-892. Reprinted here with permission of publisher.

3.1 Summary

- Sex chromosomes have evolved independently in phylogenetically diverse flowering plant lineages. The genes governing sex determination in dioecious species remain unknown, but theory predicts that the linkage of genes influencing male and female function will spur the origin and early evolution of sex chromosomes. For example, in an XY system, the origin of an active Y may be spurred by the linkage of female suppressing and male promoting genes.

- Garden asparagus (*Asparagus officinalis*) serves as a model for plant sex chromosome evolution, given that it has recently evolved an XX/XY sex chromosome system. In order to elucidate the molecular basis of gender differences and sex determination, we used RNA- sequencing (RNA-Seq) to identify differentially expressed genes between female (XX), male (XY) and supermale (YY) individuals.

- We identified 570 differentially expressed genes, and showed that significantly more genes exhibited male-biased than female-biased expression in garden asparagus. In the context of anther development, we identified genes involved in pollen microspore and tapetum develop- ment that were specifically expressed in males and supermales.

- Comparative analysis of genes in the *Arabidopsis thaliana, Zea mays* and *Oryza sativa* anther development pathways shows that anther sterility in females probably occurs through interruption of tapetum development before microspore meiosis.

3.2 Introduction

In contrast to vertebrates, species with separate sexes are relatively rare in flowering plants. Nearly 90% of angiosperms are hermaphroditic, with bisexual flowers that harbor both male

(staminate) and female (pistillate) organs (Ainsworth, 2000). At the same time, a diverse set of alternative mating strategies have evolved across the angiosperms, including the temporal separation of male and female organ function in a bisexual flower (dichogamy), the separation of unisexual male and female flowers on a single plant (monoecy), the production of both hermaphrodite and male flowers on the same plant (andromonoecy), and entirely distinct male and female plants (dioecy). Dioecy is a rare yet phylogenetically widespread phenomenon in the flowering plants; it has been described in less than 6% of angiosperm species and in nearly half (43%) of all angiosperm families, but it has evolved at least 100 independent times (Charlesworth, 2002).

From a developmental standpoint, the differentiation of males and females in dioecious systems is highly variable across flowering plant species with many species forming both stamen and pistil primordia that are selectively aborted before maturity; pistil development is arrested in male flowers, and stamens abort in female flowers. Given the repeated evolution of dioecy and the genetic complexities of androecium and gynoecium development, the timing and genetic mechanisms governing the cessation of sex organ development in ancestrally bisexual flowers may well be lineage-specific with alterations in the function of different genes spurring the transition from hermaphroditism to dioecy (Diggle et al., 2011). A closer genetic dissection of the male and female gametophyte development pathway in various model systems reveals a complex assortment of genes (Ma et al., 2008; Ma & Sundaresan, 2010; Chang et al., 2011; Zhang & Yang, 2014; Chettoor et al., 2014), hormonal controls (Wilson et al., 2011; Orozco-Arroyo et al., 2012), small RNAs (Guo et al., 2009), and epigenetic marks (Song et al., 2012) that interact to contribute to proper flower development, ultimately complicating the identification of the mutations responsible for the evolutionary transition from hermaphroditism to dioecy. Recently, a small RNA producing gene has been identified as a sex-linked candidate for regulation of anther development in dioecious persimmon species (*Diospyros* sp.; Akagi et al. 2014) but this gene has not been implicated as influencing female

function.

In dioecious plant and animal systems, a sex chromosome pair can genetically control the differentiation of male and female sexes. While the genetic basis of this mechanism has not yet been fully elucidated for any dioecious plant species, theory predicts that at least two dominant genes, one suppressing female organogenesis (ie., pistil development) and one promoting male organogenesis (ie., anther and pollen production), must be tightly linked in a non-recombining region on the Y/W chromosome (Charlesworth & Charlesworth, 1979). The identification of these sex-linked gender determination genes is difficult given the complexity of floral development paired with the highly repetitive, transposon-rich nature of non-recombining sex determination regions. As a consequence, even in dioecious species with sequenced genomes including poplar (*Populus trichocarpa*; Tuskan et al., 2006), papaya (*Carica papaya*; Ming et al., 2008), cannabis (*Cannabis sativa*; van Bakel et al., 2011) and date palm (*Phoenix dactylifera*; Al-Dous et al., 2011; Al-Mssallem et al., 2013), the causal sex determination genes remain elusive. The recent identification of a putative regulator of anther development in persimmon (Akagi et al. 2014) is consistent with the two gene model of chromosome Charlesworth & Charlesworth (1978) in that there is no evidence for the gene having any influence on female function. However, the existence and identity of a female suppressor has yet to be identified in any dioecious plant species.

Garden asparagus (*Asparagus officinalis*, 2n=2x=20) is a particularly useful system for investigating the early stages of sex chromosome evolution. It is a dioecious species with an XX/XY sex determination system, where the presence of a Y chromosome in males dominantly suppresses female organogenesis and promotes complete development of fertile anthers. The X and Y are cytologically homomorphic (Deng et al., 2012), suggesting that the conversion of an autosome pair to a sex chromosome is a recent event. Co-dominant markers verify a non-recombining, male-specific region on the Y chromosome that distinguishes males from females (Jamsari et al., 2004; Nakayama et al., 2006; Kanno et al., 2014). Moreover,

the viability of the YY supermale genotypes derived through anther culture or selfing of andromonoecious plants (Franken, 1970; Peng & Wolyn, 1999; Falavigna & Casali, 2002) suggests an early stage of sex chromosome evolution in Asparagus.

Developmentally, the temporal separation of male and female organ abortion in *A. officinalis* suggests that at least two genes are involved in sex determination. Early in floral development, female (XX), male (XY) and supermale (YY) genotypes have indistinguishable hermaphroditic floral meristems. The first observable gender difference occurs early in XY male and YY supermale development, when the tricarpellate gynoecium fails to fully elongate and fuse to form a mature stylar tube (Caporali et al., 1994). Later in XX female development, anther degeneration occurs when the tapetum breaks down before microsporogenesis completes, ultimately leading to anther sterility (Caporali et al., 1994).

The genes responsible for observed differences in male and female asparagus flower development remain unknown. One approach to identifying them is to sequence and assemble whole genome shotgun data, identifying gender-specific non-recombining and hemizygous regions, annotating genes in these regions and experimentally testing the functions of these genes. This work is underway for garden asparagus, but the transposon-rich complexity of the *A. officinalis* genome, particularly in the non-recombining sex-determination region poses challenges (Telgmann-Rauber et al., 2007). An alternative approach is to use RNA-Seq to identify gender-specific SNPs with coding regions and genes exhibiting gender-biased expression (Muyle et al., 2012) that are known to play some role in the anther and ovule developmental pathways. Here we investigate these pathways through comprehensive RNA-Seq expression profiling of developing female, male, and supermale floral and vegetative tissues in order to identify genes with shared male and female expression in addition to male and female-specific expression. The results of our analyses effectively narrow the developmental stages during which sex determination genes may alter flower development. We assembled a comprehensive transcriptome for garden asparagus including transcripts from supermales,

males and females and used it to identify expression differences among these three gender types. Genes exhibiting gender-biased expression were compared to known anther and ovule development genes in order to narrow the window during which sex-linked gender determination genes may be acting to promote anther development and gynoecium suppression in males or regulate suppression of anther development and promotion of gynoecium maturation in females.

3.3 Methods

Spear tip RNA isolation and sequencing

Recently emerged (>6 inch spears) *Asparagus officinalis* spear tips were harvested from male, female and supermale plants of four distinct breeding lines. Total RNA was isolated from the spear tips using a Qiagen RNeasy Plant Mini Kit and treated with Qiagen RNase-free DNase. RNA quality (RIN >8.0) was assessed using an Agilent Bioanalyzer RNA nano chip (Agilent, Wilmington, DE). Sequencing libraries were prepared using the Illumina TruSeq RNA Sample Preparation Kit v2 according to manufacturers specifications, and quality and insert size distribution were assessed using an Agilent Bioanalyzer DNA 1000 chip. Sequencing libraries were qPCR quantified, pooled in equimolar concentration, and analyzed on an Illumina HiSeq 2000 producing 2x100nt paired-end reads.

De novo transcriptome assembly and annotation

Raw reads were adapter clipped and assessed using FASTQC (http://www.bioinformatics.babraham.ac.uk/
to identify adapter contamination. We trimmed the 3 ends of reads with <20 Phred scores using FASTX-Toolkit (http://hannonlab.cshl.edu/fastx_toolkit/), requiring that the resulting read be greater than 40nt long. Following end trimming, reads were filtered by ap-

plying a threshold of Phred score ¿20 over at least 80% of the read length. To reduce memory and runtime, redundant paired reads were normalized to 30x coverage with normalize_by_kmer_coverage.pl, Trinitys in silico read normalization script (Grabherr et al., 2011). This reduced the dataset to nearly 40 million normalized, paired reads. Unpaired reads were added to the normalized paired reads and the entire set was assembled in Trinity (version r2012-10-05) with default parameters. Functional annotation of transcript assemblies was done using the included Trinotate pipeline (http://trinotate.sourceforge.net/) to identify open reading frames and assign best hits to UniprotKB (1e-03), PFAM-A (1e-03), Gene Ontology (GO) and eggNOG categories. Gene Ontology (GO) over-representation analysis was performed using the DAVID web server, utilizing a custom background gene set derived from the non-redundant *A. officinalis* whole transcriptome UniprotKB annotations.

Transcript abundance estimation and differential expression

Cleaned reads from each library were aligned to the assembly using Bowtie (v0.12.8) wrapped by Trinitys alignReads.pl script, and transcript abundance was estimated in the RSEM program v1.2.0 (Li & Dewey, 2011). Gene-level abundance estimations from RSEM were rounded up to the nearest integer and used as input for differential expression analysis in edgeR (Robinson et al., 2010). Initial filtering required transcripts to be represented in at least 3 libraries, each with greater than 1 read count per million mapped reads (CPM). Given the nested experimental design, an additive model (line + sex genotype) was fit in edgeR to remove the underlying differences between lines and to identify the genes that consistently display differential expression in multiple lines for a given pairwise sex comparison. Generalized Linear Model (GLM) common dispersions, trended dispersions, and tagwise dispersions were estimated before a GLM model was fit. In order to simplify and visualize multidimensional patterns in our expression data, Multidimensional Scaling (MDS) analysis was performed using the edgeR function plotMDS() with pair-wise log fold-changes

estimated on read counts (normalized to library size) for the 250 most heterogeneously expressed genes. Differentially expressed genes with a False Discovery Rate (FDR) ¡ 0.05 are reported in Supplemental Table 1.

Isolation and RNA-Seq of developmentally staged flower buds

Asparagus officinalis flower buds were harvested from female (XX) and double haploid androgenetic supermale (YY) clones. The flower buds were sampled on several plants of each clone at three approximate development stages: a) pre-meiotic, b) microspore and c) pollen. Total RNA was isolated from flower buds using a NucleoSpin RNA Plant (Macherey-Nagel GmbH & Co. KG, 52355 Dren, Germany) and treated with RNase-free DNase. RNA quantity and purity were assessed using a Nanodrop 1000 spectrophotometer (Thermo Scientific, Waltham, MA); the RNA integrity (¿8.0) was determined using a Bioanalyzer 2100 (Agilent, Wilmington, DE) with an RNA 6000 Nano chip (Agilent, Wilmington, DE). Synthesis of cDNA and library preparation were carried out starting with 2.5 g of total RNA using the TruSeq RNA Sample Prep Kit v2 (Illumina, San Diego, CA) according to manufacturer instructions. Size selection of cDNA libraries was performed by excising and eluting 350 to 550nt fragments using a MinElute Gel Extraction kit (Qiagen, Hilden, Germany). Following size selection, the cDNA was quantified using qPCR with KAPA Library Quantification kit (KAPA Biosystems, Woburn, MA) and RNA-Seq libraries were pooled and sequenced on an Illumina HiSeq 2000 generating 100nt paired-end sequences.

3.4 Results and Discussion

Transcriptome sequencing and assembly

We generated RNA-Seq to assemble transcripts and quantify spear tip gene expression patterns in all females, males, and supermales. Whole transcriptome Illumina shotgun sequencing of each gender type in four inbred lines yielded nearly 648 million paired-end reads. These data were combined with two existing YY supermale spear tip libraries for a total of nearly 763 million paired-end reads (Table 1). Given the large number of sequenced reads, we used Trinity to digitally normalize and de novo assemble the reads into 276,556 transcripts representing 120,061 subcomponents (referred to here as genes or loci) with an N50 of 2,386nt (Supplemental Figure 1).

To better assess the number of full-length transcripts assembled, we compared the transcripts to 768,340 peptide annotations from 22 sequenced and annotated plant genomes (Amborella Genome Project, 2013) using blastx (1e-10, best hit only) and found that 13,336 unique peptides aligned over at least 80% of their length to an *A. officinalis* transcript (Supplemental Figure 2). Additionally, we performed a gene family analysis to determine the diversity of sampled transcripts. Using OrthoMCL-derived estimated gene families from these 22 genomes, we translated and sorted 121,704 Asparagus transcripts into 13,498 distinct gene families using a pipeline described in McKain et al. (2012).

We annotated transcripts through comparisons with the Swiss UniprotKB complete proteome databases and other databases using the Trinotate pipeline (see methods for details). Nearly 49% of the assembled transcripts had a blastx (e-value 1e-03) hit against the UniprotKB database. Similarly, 51% of the transcripts had open reading frames greater than 100 amino acids long. We extracted 7,774 non-redundant Gene Ontology (GO) terms from 41% of the transcripts and summarized them into 103 GOslim plant categories using CateGOrizer (Hu et al., 2008) (Supplemental Table 2). The complete list of transcript annotations is pre-

Table 3.1: Sequencing read counts and alignment statistics for all sequenced garden asparagus individuals. Mapped read percentages only include those reads with concordant, paired alignments.

Line	Sex	Tissue Type	Clean pairs	% Mapped
8A	Female	Spear tip	11,506,758	88.72
8A	Male	Spear tip	33,433,776	87.53
8A	Supermale	Spear tip	37,066,238	88.37
8B	Female	Spear tip	33,711,613	87.84
8B	Male	Spear tip	6,677,641	87.43
8B	Supermale	Spear tip	24,710,491	87.81
10	Female	Spear tip	40,348,458	87.50
10	Male	Spear tip	63,871,037	88.39
10	Supermale	Spear tip	1,463,330	87.18
9	Female	Spear tip	36,909,297	87.41
9	Male	Spear tip	28,551,654	86.68
1	Supermale	Pre-meiotic*	25,929,261	82.44
1	Supermale	Meiotic*	21,421,286	80.36
1	Supermale	Post-meiotic*	28,222,100	79.84
2	Female	Pre-meiotic*	28,077,734	82.50
2	Female	Meiotic*	27,316,690	82.59
2	Female	Post-meiotic*	18,451,370	81.34

* signifies tissue derived from flower bud isolations

sented in Supplemental Table 3. Taken together, these analyses suggest that we sequenced and accurately assembled a diverse set of transcripts expressed in *A. officinalis* spear tips which include developing vegetative and reproductive organs.

Diverse lines reveal male-biased gene expression

Following transcript annotation, transcript abundances for every library were calculated using RSEM. We filtered the transcriptome assembly by requiring that each tested gene be expressed at greater than 1 CPM in at least three libraries, which reduced the number of tested genes to 19,986. The first two dimensions of the MDS analysis on the 250 most heterogeneously expressed genes revealed clustering of individuals by breeding line rather than gender (Figure 1). At the same time, there is substantial among-gender transcriptional variation within lines (coefficient of variation = 0.33). After accounting for the underlying differences between lines (see methods), tests for differential gene expression in pairwise Male:Female, Supermale:Female, and Supermale:Male comparisons identified a total of 570 differentially expressed genes (DEGs) (Figure 2). Almost 91% of DEGs are upregulated in males and supermales compared to females, indicating that gene expression is generally male-biased in spear tip tissue. Additionally, nearly 34% of DEGs were found in both male and supermale transcript pools (Figure 3). All differentially expressed genes and their annotations are presented in Supplemental Table 1.

Sex-biased gene expression is a common phenomenon in both dioecious plant and animal species, as documented in humans (Dimas & Nica, 2012), *Drosophila* (Zhang et al., 2007), emus (Vicoso et al., 2013), and *Silene latifolia* (Zemp et al., 2014) among others. Genes that exhibit sex-biased expression might be under unique evolutionary constraints, evolving more rapidly both in terms of sequence and expression (Meiklejohn et al., 2003). The developmental stage in which sex-biased genes are expressed may also impact their rate of evolution given that they often have reproductively and developmentally specialized

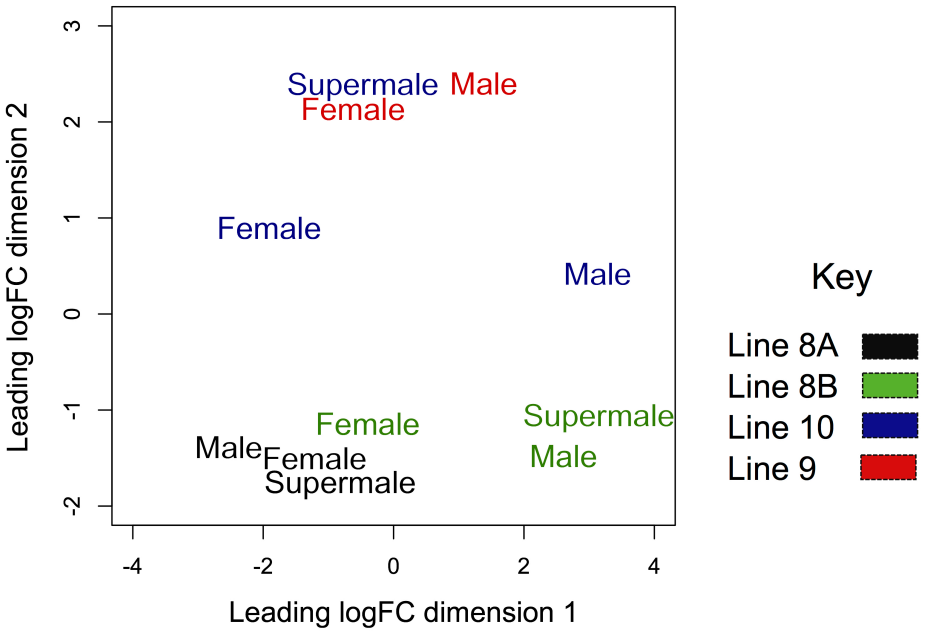

Figure 3.1: Multidimensional scaling (MDS) plot showing the distance between samples as the root-mean-square deviation (Euclidean distance) for the top 250 most heterogeneously expressed genes in independent, unsupervised pairwise comparisons between samples.

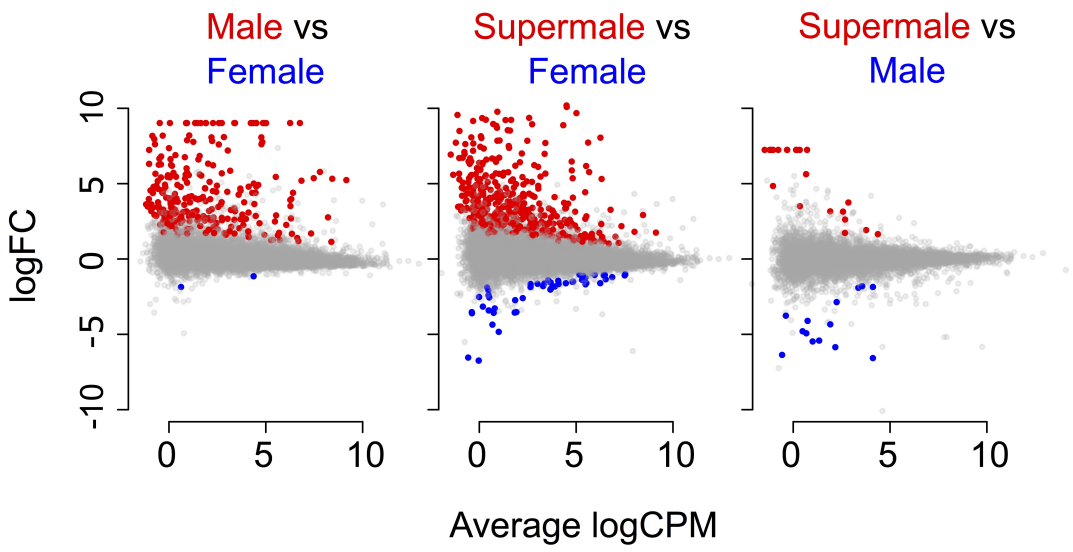

Figure 3.2: Pairwise comparisons of gene expression (average log fold change vs. average log of counts per million mapped reads) among female, male and supermale genes across line replicates. Red and blue dots indicate genes exhibiting significantly biased expression levels at a False Discovery Rate (FDR) ¡ 0.05. Colored dots spread horizontally at the y-axis upper limit represent genes not expressed in one comparison, and are plotted at the minimum CPM across all samples. Grey dots represent genes that are not significantly differentially expressed.

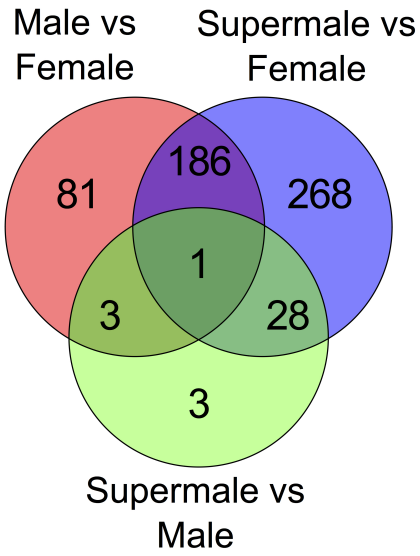

Figure 3.3: Venn diagram showing the overlap of differentially expressed genes between the three pairwise comparisons.

expression domains (Perry et al., 2014).

Sex-biased gene expression in dioecious species is strongly influenced by the developmental stage at which gynecium develop is arrested in male flowers and male function is interrupted in anthers of female flowers. In garden asparagus anthers are initiated in female flowers and develop to a stage just prior to microspore meiosis (Caporali et al., 1994). Therefore, pollen development genes functioning downstream of Y-linked sex-determination genes are expected to be expressed in males but not female flowers. Males form a tricarpellate gynoecium with development of the stylar tube typically arrested before the stigma forms. Ovule development and abortion in males is not well characterized but variation in male style lengths (Galli et al., 1993) and the occurrence of andromonecy in some genotypes (Sneep, 1953) suggests a continuum of points over which female development is interrupted in male flowers. The expression profiles of many genes vary over the course of development of both male and female flowers, underscoring the need for developmentally staged sampling of transcripts.

Heterogenous gene expression among lines for each gender

As indicated in the MSD plot of the 570 most differentially expressed genes (Figure 1), there is significant among-line gene expression variation for each gender. This is especially true for males and supermales. Whereas both axes of the MDS plot exhibit variation among males and supermales, all females for all four lines have similar values on the first axis of the MDS plot. Heterogeneous gene expression among males and supermales is also evident in a heatmap (Figure 4). Strikingly, the line 8A male supermale expression profiles clusters with the four female individuals (Figure 4). We hypothesized that the observed male and supermale expression heterogeneity is likely related to developmental differences between the lines and the sexes; despite all spear tips being collected at the same vegetative height (6 inches above ground), males and supermales will usually flower earlier than females

(Kahara et al., 1940; Caporali et al., 1994). To better understand this heterogeneity, we sectioned and imaged spear tips from female, male, and supermale spears of the same spear heights. Unfortunately, we were not able to section the lines included in the gene expression study (they are no longer available), but our observations matched those of previous studies (Kahara et al., 1940; Caporali et al., 1994). All gender types exhibited a developmental gradient along the length of the spear tip, with young floral buds at the top and more mature buds towards the bottom. Moreover, supermale spear tips at each height have the most developmentally advanced flower buds, female flower buds are least developmentally advanced at each height and males are intermediate (Supplemental Figure 3). We surmise that similar developmental variation exists among lines and flower bud development is slow in line 8A relative to the other lines.

In order to better resolve whether the heterogeneous expression patterns among males and supermales might be due to differences in flower bud development, we generated additional targeted RNA-Seq derived from flower buds at three developmental stages (pre-meiotic, meiotic, and post-meiotic) in one XX female and one YY supermale line (Table 1). Among-sample correlations in the 570 most differentially expressed genes identified in the spear tip gene expression profiles suggest that the among-line variation for each gender may indeed be due variation in the frequency of flower buds at each stage (Figure 5; Supplemental Figure 4). The female pre-meiotic and meiotic flower bud expression profiles are similar to the line 8A supermale spear tips, whereas the female spear tips transcript profiles are more correlated with the line 8A male spear tip profile. Interestingly, the supermale pre-meiotic flower bud expression profile is quite distinct from the other flower bud profiles and most similar to lines 10 and 8B male profiles (Figure 5).

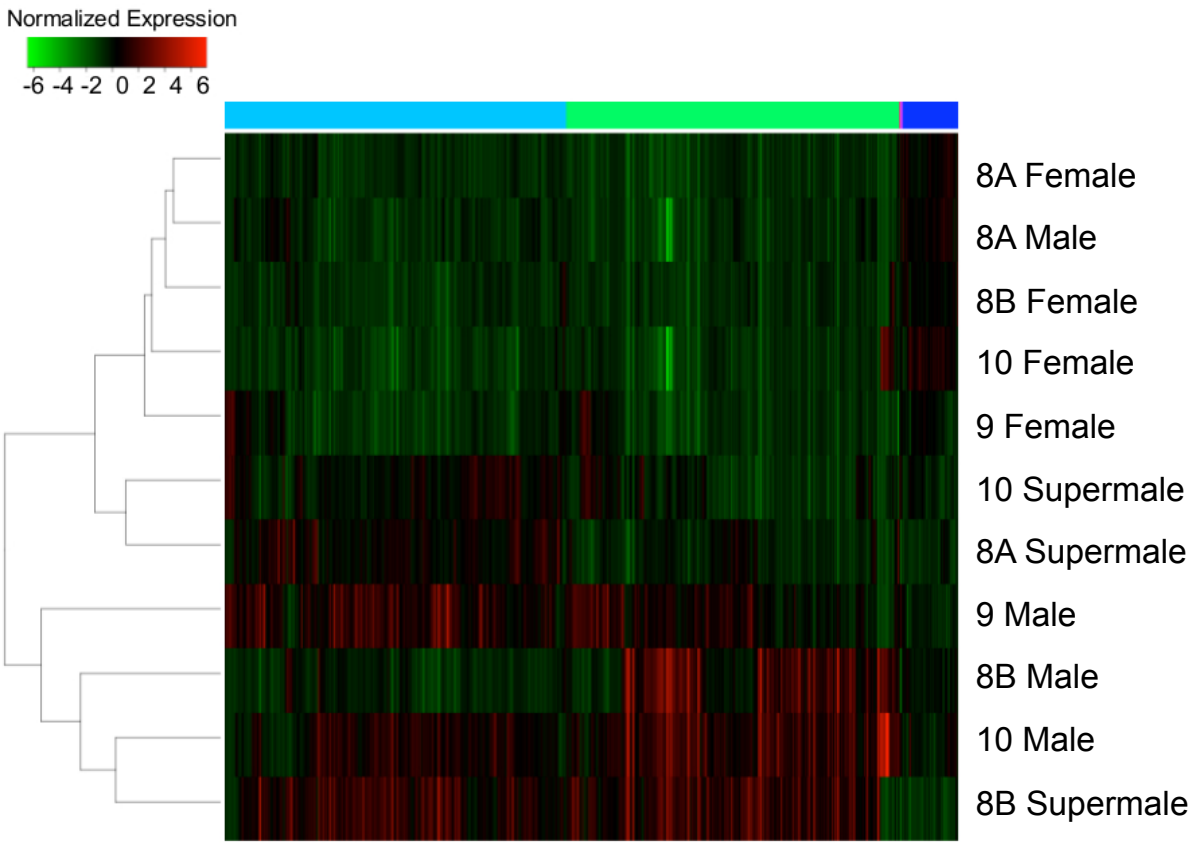

Figure 3.4: Heatmap clustering of the 570 differentially expressed genes in all spear tip tissues. TMM-normalized FPKM values were log transformed before ward clustering, cutting the tree into k=4 clusters.

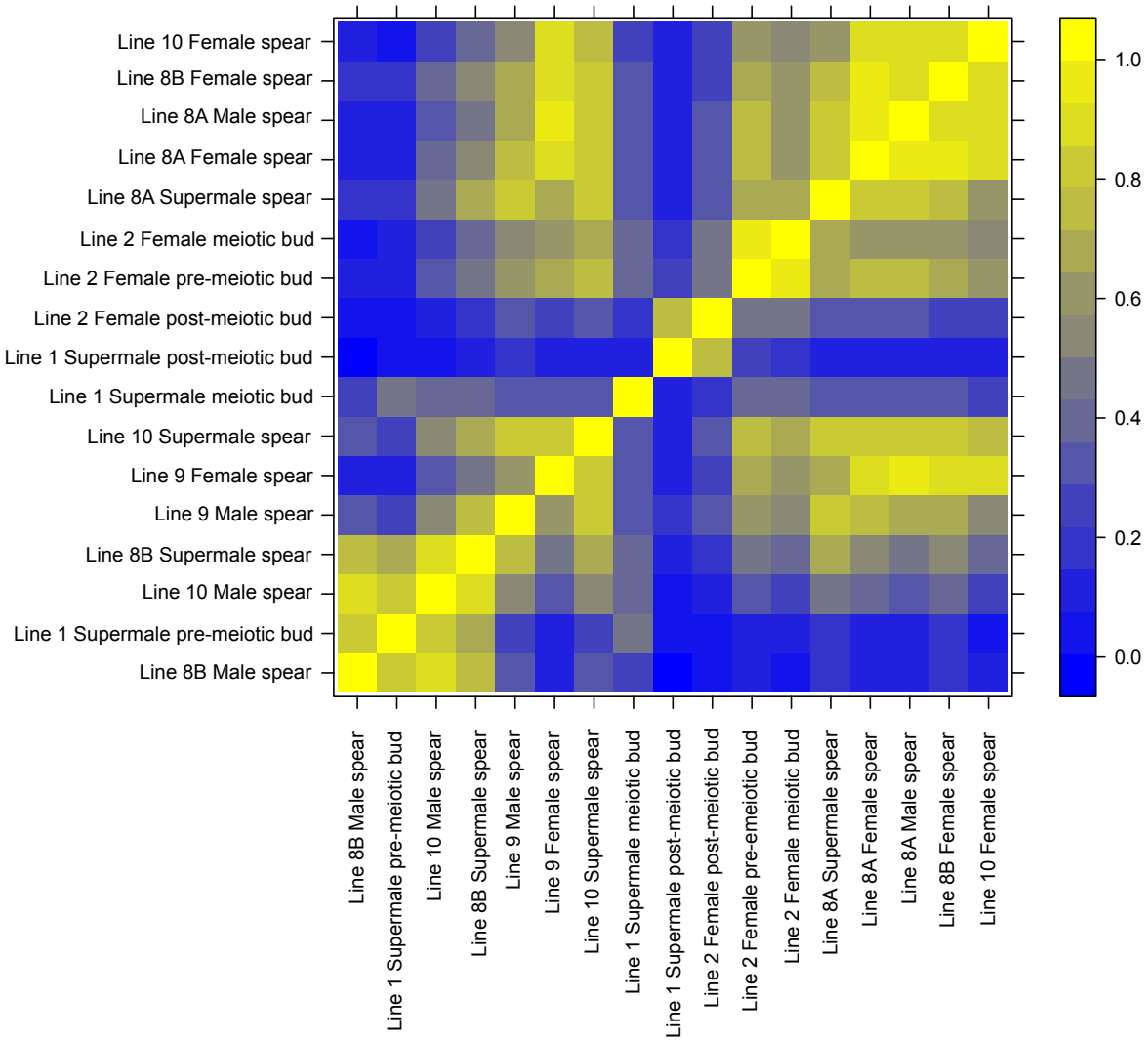

Figure 3.5: Spearman correlation matrix of the 570 differentially expressed genes in all comparisons for spear tip tissues and the developmentally staged flower buds. TMM-normalized FPKM values were used to generate a spearman correlation matrix, and individuals were ordered according to a complete clustering dendrogram using that matrix

Differential expression in the anther development pathway

Gene expression patterns in all samples were compared to expression profiles for known *Arabidopsis*, *Zea* and *Oryza* anther development genes (Ma et al., 2008; Ma & Sundaresan, 2010; Wilson et al., 2011; Zhang & Yang, 2014). Previous developmental studies in garden asparagus revealed that anther development in female anthers arrests after initiation of tapetum development but before formation of tetrads in microspore meiosis (Caporali et al., 1994). By placing the differences between male and female gene expression within the context of known anther development pathway (Figure 6), we aimed to distinguish differentially expressed sex-determination genes from downstream male function genes.

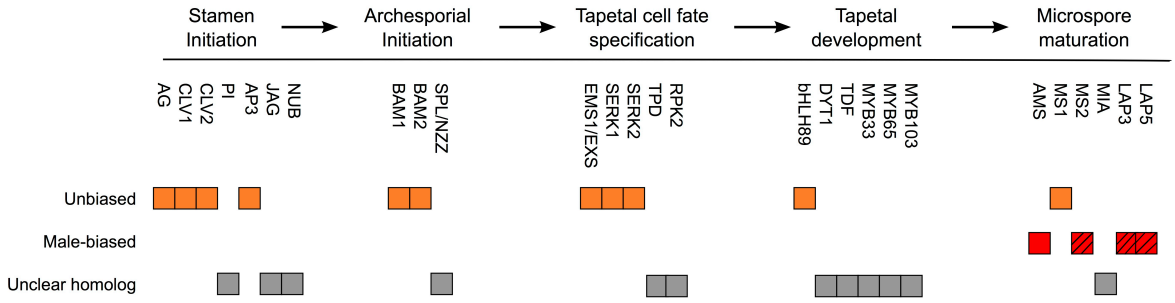

Figure 3.6: A sampling of genes involved in the A. thaliana anther development pathway and their expression patterns in female and male/supermale *A. officinalis*. Using 0.1 FPKM as a cutoff for identifying a gene as expressed, we identified genes that showed no gender-biased expression between female, male, and supermale sexes (orange blocks). Genes involved in stamen initiation, archesporial initiation, and tapetal cell fate specification show expression in all genders. Later in tapetal development and microspore maturation, some genes such as AMS show male-biased expression (red blocks without hatching), while others such as MS2, LAP3 and LAP5 show male-specific expression (red blocks with hatching). Genes that were unable to be confidently annotated in our transcriptome assembly are represented by grey blocks.

A total of 516 genes were identified as exhibiting male-biased (XY and YY) expression relative to females. Of these, 46 displayed no expression in any XX female spear (Supplemental Table 1). Genes with male-specific expression in *A. officinalis* included homologs of Quartet1 (QRT1), Polygalacturonase (PG), and other pollen related genes. QRT1 is

a pectinesterase expressed in *A. thaliana* anthers shortly after microspore meiosis is completed. QRT1 is required for proper microspore separation (Francis et al., 2006). The qrt1 mutant will properly deposit and degrade callose on the secondary pollen mother cell (PMC) wall, but the primary PMC is left mostly intact (Preuss et al., 1994; Rhee & Somerville, 1998). Polygalacturonase (PG), a member of a family of cell wall loosening enzymes that are specifically involved in the degradation of pectin and cell walls, is involved in pollen exine development. Several PG homologs in *Brassica napus*, *Zea maize*, and *Lillium longiflorium* are specifically expressed in late-stage pollen development (Niogret et al., 1991; Robert et al., 1993; Chiang et al., 2006). Several other pollen cell wall-related genes are significantly more abundant in males and supermales relative to females, including genes encoding Laccase 7, pollen-specific leucine-rich repeat extension-like protein 1, oligopeptide transporter 5 (OPT5), and glucan endo-1,3-beta-glucosidase, respectively. This transcriptional pattern suggests that genes required for proper pollen wall exine formation are expressed weakly in female anthers relative to males, which agrees with previous microscopic analysis of anther degeneration in XX female plants (Caporali et al., 1994; Mariziani et al., 1999). While each of these genes could be considered promoters of male function, their reduced transcription in female flowers does not necessarily imply that they are ultimately responsible for gender determination.

Since the expression data suggested that the gene(s) ultimately responsible for female anther sterility occurs upstream of pollen exine formation, we next identified genes that tend to show high expression in *A. thaliana* tapetal tissues. Male Sterility 2 (MS2) exhibits male-specific expression in *A. officinalis*, encoding a fatty acid reductase that is primarily expressed in the *A. thaliana* tapetum during pollen tetrad release (Aarts et al., 1997). Another tapetum-specific gene is Aborted Microspores (AMS), a gene that encodes a MYC class basic helix-loop-helix (bHLH) transcription factor that is expressed specifically in the tapetum of *A. thaliana* anthers. Tapetal tissue in ams mutants degenerate prematurely (Xu

et al., 2010), similar to developmental defects found in *A. officinalis* females (Caporali et al., 1994). AMS can also modulate the expression of a suite of anther development genes, binding to promoter regions in vivo (Xu et al., 2010). Although exhibiting significantly higher expression in males, we do find female expression for AMS. Similar patterns of male and supermale-biased expression emerge for other tapetum-related transcription factors like bHLH89, which has been shown to interact with AMS during anther development (Ma et al., 2012).

Finally, we identified genes expressed further upstream of tapetum and pollen development, with the expectation that they would be expressed in both males and females. Indeed, we find very similar female, male and supermale expression levels for genes involved in development at multiple stages far upstream of microsporogenesis (stamen initiation, archesporial initiation, tapetal cell fate specification, and tapetal development; Figure 6). Given the developmental patterns described by Caporali et al. (1994) and above, genes involved in initiation and early development of male and female organs were expected to show similar expression patterns. For example, in agreement with earlier work (Park et al., 2003) no difference was observed between male and female flower buds in expression of homeotic MADS box genes.

Differential expression in the phenylpropanoid biosynthesis pathway

We also found numerous genes in the phenylpropanoid biosynthesis pathway that were differentially expressed between sexes. The general phenylpropanoid and flavonoid pathway is well described, and has been shown to affect pollen development and male sterility (van der Meer et al., 1992). Significant male or supermale-biased expression occurs in genes encoding Cinnamic 4-hydroxylase acid (C4H), Chalcone Synthase (CHS), Flavanone-3-

hydroxylase (F3H), Stilbene Synthase 2 (STS2), Stilbene Synthase 4 (STS4), and Shikimate O-hydroxycinnamoyltransferase (HCT). Several genes in the phenylpropanoid pathway exhibit significantly higher transcript abundance in females, too, including Cinnamoyl-CoA reductase 1 and 2 (CCR1/CCR2) and Leucoanthocyanidin deoxygenase (LDOX).

We identified two differentially expressed genes annotated as Chalcone Synthase. One of these transcript assemblies (comp32958_c0) was annotated based on a match to a *Malus* x *domestica* gene model in UniprotKB, and showed even closer similarity to the *A. thaliana* LESS ADHESIVE POLLEN 5 (LAP5) gene, a chalcone and stilbene synthase family member. LAP5 is essential for proper pollen exine development and shows co-expression with other asparagus male anther-specific genes like Male Sterility 2 (MS2), providing additional transcriptional evidence for the lack of proper pollen development in female asparagus plants (Dobritsa et al., 2010). Further, the differential expression of numerous phenylpropanoid genes in dioecious asparagus is intriguing given that disruptions in the phenylpropanoid and downstream anthocyanin pathways have classically been implicated in nuclear and cytoplasmic male sterility (CMS) phenotypes (van der Meer et al., 1992; Matsuda et al., 1996).

Overall, the observed patterns of gender biases in gene expression leads us to hypothesize that a Y-linked male promoting gene, perhaps complementing a deficiency on the non-recombining portion of the X chromosome, plays some role in late development of the tapetum, in early microsporocyte development, likely before or during pollen wall exine formation and pollen tetrads disassociate. This inference is consistent with the developmental investigations of Caporali et al. (1994) and the model of sex chromosome evolution proposed by Charlesworth and Charlesworth (1978).

Differential expression in ovule development genes

Gynoecium and female gametophyte development pathways have been investigated through analysis of homeotic mutants (e.g. see Ferrandiz et al., 1999; Alvarez & Smyth, 1999). A

small number of asparagus transcripts with *Arabidopsis* homologs influencing ovule development exhibited female-biased expression relative to males or supermales. In total, just 51 genes exhibited significantly higher transcript abundance in females relative to males or supermales. The smaller number of transcripts exhibiting female-biased expression relative to those exhibiting male-biased expression is not surprising given the massive diversity genes expressed during anther and pollen (the male gametophyte) development (e.g. Ma et al., 2008). GO term enrichment for transcripts with female-biased expression included phenylpropanoid biosynthesis and cellular amino acid derivative processes (Supplemental Table 4). Among these differentially expressed genes, we found AINTEGUMENTA (ANT) homolog transcripts as significantly more abundant in females than males or supermales. ANT is an APETALA2 (AP2)-domain containing ethylene-responsive transcription factor that regulates many aspects of female growth, including ovule and gynoecium development, inner and outer integument development, and petal epidermal cell identity (Elliott et al., 1996; Klucher et al., 1996; Mizukami & Fischer, 2000). Previous studies in Arabidopsis have shown that ant mutants are female-sterile, with undeveloped or malformed integuments and a failure for megasporogenesis to proceed past the tetrad stage (Elliott et al., 1996).

Several other genes with female-biased expression and homologs involved in female gametophyte development and flowering time were also identified including PIN2 and FD. The PIN gene family is intriguing given its involvement in auxin transport, and some members show expression early in ovule development in *A. thaliana* (Fraga et al., 2002; Ceccato et al., 2013). FD is another interesting gene involved in the transition from vegetative to floral development, encoding for a bZIP transcription factor that directly regulates FLOWERING LOCUS T (FT), the floragen gene which promotes the transition to flowering at the shoot apex (Abe et al., 2005).

3.5 Conclusions

De novo RNA-Seq transcriptome assembly and expression analysis can inform investigation of gender determination in dioecious species, but whereas many genes may exhibit sex-based expression few are expected to regulate gender determination. Most genes exhibiting male are female biased expression function after gender has been specified in the developing flower bud. Here we place genes exhibiting gender-biased expression within anther and carpel developmental pathways in order to identify the earliest points in these pathways that are influenced by Y-linked Asparagus gender determination genes. As expected, we observed male-specific expression for many known pollen development genes, but similar expression levels in male and female spear tips for organ specifying genes and several tapetum development genes, corroborating previous findings (Caporali et al. 1994) in suggesting that female asparagus anthers initiate the abortion process during tapetum and young microspore development. The differentially expressed genes placed within the anther and gynecium developmental pathways will guide ongoing work to identify Y-linked genes ultimately responsible for sex determination in dioecious Asparagus species.

Unsurprisingly, Asparagus homologs of the HD-Zip genes hypothesized to regulate gender differences in anther development in dioecious persimmon species (Akagi et al., 2014), were not implicated as possible gender determination genes. Class-I HD-Zip transcripts were observed in the transcript assembly, but none exhibited sex-biased expression. Given that dioecy has evolved independently in many flowering plant lineages (Charlesworth, 2002; Renner 2014) and suppression or loss of function mutations of many genes in the anther development pathway would convert hermaphroditic flowers to female, we expect that the identity of sex determination genes will vary among unrelated dioecious species.

3.6 Acknowledgements

This work was supported with funding from the National Science Foundation grants DEB-0951071 and DEB-0841988, the University of Georgia and Limgroup (Horst, Netherlands). We are particularly grateful for the insights provided by Ron van der Hulst of Limgroup. Sequencing was performed through the Georgia Genomics Facility (dna.uga.edu) and computational resources were provided by the Georgia Advanced Computing Resource Center (gacrc.uga.edu).

3.7 References

Aarts M, Hodge R, Kalantidis K. 1997. The Arabidopsis MALE STERILITY 2 protein shares similarity with reductases in elongation/condensation complexes. The Plant Journal 12:615623.

Abe M, Kobayashi Y, Yamamoto S, Daimon Y, Yamaguchi A, Ikeda Y, Ichinoki H, Notaguchi M, Goto K, Araki T. 2005. FD, a bZIP protein mediating signals from the floral pathway integrator FT at the shoot apex. Science 309:10521056.

Ainsworth C. 2000. Boys and Girls Come Out to Play: The Molecular Biology of Dioecious Plants. Annals of Botany 86:211221.

Akagi T, Henry I, Tao R, Comai L. 2014. A Y-chromosome-encoded small RNA acts as a sex determinant in persimmons. Science 346:646-50.

Al-Dous EK, George B, Al-Mahmoud ME, Al-Jaber MY, Wang H, Salameh YM, Al-Azwani EK, Chaluvadi S, Pontaroli AC, DeBarry J, et al. 2011. De novo genome sequencing and comparative genomics of date palm (Phoenix dactylifera). Nature biotechnology 29:5217.

Al-Mssallem IS, Hu S, Zhang X, Lin Q, Liu W, Tan J, Yu X, Liu J, Pan L, Zhang T, et al. 2013. Genome sequence of the date palm Phoenix dactylifera L. Nature communications 4:2274.

Alvarez J & Smyth D. 1999. CRABS CLAW and SPATULA, two genes that control carpel development in parallel with AGAMOUS. Current Topics in Developmental Biology 45:155-205.

Van Bakel H, Stout JM, Cote AG, Tallon CM, Sharpe AG, Hughes TR, Page JE. 2011. The draft genome and transcriptome of Cannabis sativa. Genome Biology 12:R102.

Caporali E, Carboni A, Galli M. 1994. Development of male and female flower in Asparagus officinalis. Search for point of transition from hermaphroditic to unisexual developmental pathway. Sexual Plant Reproduction 7:239249.

Ceccato L, Masiero S, Sinha Roy D, Bencivenga S, Roig-Villanova I, Ditengou FA, Palme K, Simon R, Colombo L. 2013. Maternal control of PIN1 is required for female gametophyte development in Arabidopsis. PloS ONE 8:e66148.

Chang F, Wang Y, Wang S, Ma H. 2011. Molecular control of microsporogenesis in Arabidopsis. Current Opinion in Plant Biology 14:6673.

Charlesworth B, Charlesworth D. 1978. A model for the evolution of dioecy and gynodioecy. American Naturalist 112:975997.

Charlesworth D. 2002. Plant sex determination and sex chromosomes. Heredity 88:94101.

Charlesworth D, Charlesworth B. 1979. The evolutionary genetics of sexual systems in flowering plants. Proceedings of the Royal Society of London B. 205:513530.

Chettoor AM, Givan SA, Cole RA, Coker CT, Unger-Wallace E, Vejlupkova Z, Vollbrecht E, Fowler JE, Evans MM. 2014. Discovery of novel transcripts and gametophytic functions via RNA-seq analysis of maize gametophytic transcriptomes. Genome Biology 15(7):414.

Chiang JY, Shu SW, Ko CW, Wang CS. 2006. Biochemical characterization of a pollen-specific cDNA encoding polygalacturonase in Lilium longiflorum. Plant Science 170:433440.

Deng C-L, Qin R-Y, Wang N-N, Cao Y, Gao J, Gao W-J, Lu L-D. 2012. Karyotype of asparagus by physical mapping of 45S and 5S rDNA by FISH. Journal of Genetics 91:20912.

Diggle PK, Di Stilio VS, Gschwend AR, Golenberg EM, Moore RC, Russell JRW, Sinclair

JP. 2011. Multiple developmental processes underlie sex differentiation in angiosperms. Trends in Genetics 27:36876.

Dimas A, Nica A. 2012. Sex-biased genetic effects on gene regulation in humans. Genome Research 22:23682375.

Dobritsa A, Lei Z, Nishikawa S-I, Urbanczyk-Wochniak E, Huhman D V, Preuss D, Sumner LW. 2010. LAP5 and LAP6 encode anther-specific proteins with similarity to chalcone synthase essential for pollen exine development in Arabidopsis. Plant Physiology 153:93755.

Elliott R, Betzner A, Huttner E. 1996. AINTEGUMENTA, an APETALA2-like gene of Arabidopsis with pleiotropic roles in ovule development and floral organ growth. The Plant Cell 8:155168.

Falavigna, A. & Casali, P.E. 2002. Practical aspects of breeding program of asparagus based on in vitro anther culture. Acta Horticulturae 589:201-210

Ferrndiz C, Pelaz S, Yankofsky MF. 1999. Control of carpel and fruit development in Arabidopsis. Annual Reviews of Biochemistry 68:328-54. Flory WS. 1932. Genetic and cytological investigations on Asparagus officinalis L. Genetics 17:432467.

Fraga MF, Rodrguez R, Caal MJ. 2002. Genomic DNA methylation-demethylation during aging and reinvigoration of Pinus radiata. Tree Physiology 22:8136.

Francis KE, Lam SY, Copenhaver GP. 2006. Separation of Arabidopsis pollen tetrads is regulated by QUARTET1, a pectin methylesterase gene. Plant Physiology 142:100413.

Franken AA. 1970. Sex characteristics and inheritance of sex in asparagus (Asparagus officinalis L.). Euphytica 19:277287. Galli MG, Bracale M, Falavigna A, Raffaldi F, Savini C, Vigo A. 1993.

Different kinds of male flowers in the dioecious plant Asparagus of officinalis L. Sexual Plant Reproduction 6(1):16-21.

Grabherr MG, Haas BJ, Yassour M, Levin JZ, Thompson DA, Amit I, Adiconis X, Fan L, Raychowdhury R, Zeng Q, Chen Z, Mauceli E, Hacohen N, Gnirke A, Rhind N, di Palma F,

Birren BW, Nusbaum C, Lindblad-Toh K, Friedman N, Regev A. Full-length transcriptome assembly from RNA-seq data without a reference genome. Nature Biotechnology 29(7):644-52.

Guo X, Su B, Zhou Z, Sha J. 2009. Rapid evolution of mammalian X-linked testis microRNAs. BMC Genomics 10:97.

Hu Z-L, Bao J, Reecy J. 2008. CateGOrizer: A Web-Based Program to Batch Analyze Gene Ontology Classification Categories. Online Journal of Bioinformatics 9:108112.

Jamsari A, Nitz I, Reamon-Bttner SM, Jung C. 2004. BAC-derived diagnostic markers for sex determination in asparagus. TAG. Theoretical and applied genetics. Theoretische und angewandte Genetik 108:11406.

Kahara K, Hagiwara T, Hamada K, Tamura S. 1940. A study of sex in young asparagus plantings. Journal of the Horticular Association of Japan 11:150-161

Kanno A, Kubota S, Ishino K. 2014. Conversion of a male-specific RAPD marker into an STS marker in Asparagus officinalis L. Euphytica 197:39-46

Klucher KM, Chow H, Reiser L, Fischer RL. 1996. The AINTEGUMENTA gene of Arabidopsis required for ovule and female gametophyte development is related to the floral homeotic gene APETALA2. The Plant Cell 8:13753.

Langmead B, Trapnell C, Pop M, Salzberg S. 2009. Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. Genome Biology 10:R25

Li B, Dewey CN. 2011. RSEM: accurate transcript quantification from RNA-Seq data with or without a reference genome. BMC Bioinformatics 12:323.

Loptien H. 1979. Identification of the sex chromosome pair in asparagus (Asparagus officinalis L.). Z. Pflanzenzuecht. 82:162 173.

Ma X, Feng B, Ma H. 2012. AMS-dependent and independent regulation of anther transcriptome and comparison with those affected by other Arabidopsis anther genes. BMC Plant Biology 12:23.

Ma H, Sundaresan V. 2010. Development of flowering plant gametophytes. Current Topics in Developmental Biology 91:379412.

Ma J, Morrow DJ, Fernandes J, Walbot V. 2008. Comparative profiling of the sense and antisense transcriptome of maize lines. Genome Biology 7(3):R22

Mariziani G, Caporali E, Spada A. 1999. Search for genes involved in asparagus sex determination. Sex determination in plants. BIOS Scientific Publishers, Oxford. 149162.

Matsuda N, Tsuchiya T, Kishitani S, Tanaka Y, Toriyama K. 1996. Partial Male Sterility in Transgenic Tobacco Carrying Antisense and Sense PAL cDNA under the Control of a Tapetum-Specific Promoter. Plant and Cell Physiology 37:215222.

McKain MR, Wickett N, Zhang Y, Ayyampalayam S, McCombie WR, Chase MW, Pires JC, dePamphilis CW, Leebens-Mack J. 2012. Phylogenomic analysis of transcriptome data elucidates co-occurrence of a paleopolyploid event and the origin of bimodal karyotypes in Agavoideae (Asparagaceae). American Journal of Botany 99:397406.

Van der Meer IM, Stam ME, van Tunen a J, Mol JN, Stuitje a R. 1992. Antisense inhibition of flavonoid biosynthesis in petunia anthers results in male sterility. The Plant Cell 4:25362.

Meiklejohn CD, Parsch J, Ranz JM, Hartl DL. 2003. Rapid evolution of male-biased gene expression in Drosophila. Proceedings of the National Academy of Sciences of the United States of America 100:98949.

Ming R, Hou S, Feng Y, Yu Q, Dionne-Laporte A, Saw JH, Senin P, Wang W, Ly B V, Lewis KLT, et al. 2008. The draft genome of the transgenic tropical fruit tree papaya (Carica papaya Linnaeus). Nature 452:991996.

Mizukami Y, Fischer RL. 2000. Plant organ size control: AINTEGUMENTA regulates growth and cell numbers during organogenesis. Proceedings of the National Academy of Sciences of the United States of America 97:9427.

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:e1001308.

Nakayama H, Ito T, Hayashi Y, Sonoda T, Fukuda T, Ochiai T, Kameya T, Kanno A, 2006. Development of sex-linked primers in garden asparagus (Asparagus officinalis L.). Breeding Science 56:327-330.

Niogret MF, Dubald M, Mandaron P, Mache R. 1991. Characterization of pollen polygalacturonase encoded by several cDNA clones in maize. Plant Molecular Biology 17:11551164.

Orozco-Arroyo G, Vzquez-Santana S, Camacho A, Dubrovsky JG, Cruz-Garca F. 2012. Inception of maleness: auxin contribution to flower masculinization in the dioecious cactus Opuntia stenopetala. Planta 236:22538.

Park JH, Ishikawa Y, Yoshida R, Kanno A, Kameya T. 2003. Expression of AODEF, a B-functional MADS-box gene, in stamens and inner tepals of the dioecious species Asparagus officinalis L. Plant Molecular Biology 51(6):867-75.

Peng M, Wolyn DJ. 1999. Development of a microspore culture method to produce haploid and doubled-haploid asparagus (asparagus officinalis l.) plants. Acta Horticulturae 479:357363.

Perry JC, Harrison PW, Mank JE. 2014. The Ontogeny and Evolution of Sex-Biased Gene Expression in Drosophila melanogaster. Molecular biology and evolution 31:120619.

Preuss D, Rhee SY, Davis RW. 1994. Tetrad analysis possible in Arabidopsis with mutation of the QUARTET (QRT) genes. Science 264:145860.

Renner SS. 2014. The relative and absolute frequencies of angiosperm sexual systems: Dioecy, monoecy, gynodioecy, and an updated online database. American Journal of Botany 101(10):1588-96

Rhee SY, Somerville CR. 1998. Tetrad pollen formation in quartet mutants of Arabidopsis thaliana is associated with persistence of pectic polysaccharides of the pollen mother cell wall. The Plant Journal 15:7988.

Robert LS, Allard S, Gerster JL, Cass L, Simmonds J. 1993. Isolation and characterization of a polygalacturonase gene highly expressed in Brassica napus pollen. Plant Molecular Biology 23:12731278.

Robinson MD, McCarthy DJ, Smyth GK. 2010. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. Bioinformatics 26:139140.

Sneep J. 1953. The significance of andromonoecy for the breeding of Asparagus officinalis L. Euphytica 2(2):89-95

Song Y, Ma K, Bo W, Zhang Z, Zhang D. 2012. Sex-specific DNA methylation and gene expression in andromonoecious poplar. Plant Cell Reports 31:1393405.

Telgmann-Rauber A, Jamsari A, Kinney MS, Pries JC, Jung C. 2007 Genetic and physical maps around the sex-determining M-locus of the dioecious plant asparagus. Molecular Genetics and Genomics 278:221234.

Tuskan GA, Difazio S, Jansson S, Bohlmann J, Grigoriev I, Hellsten U, Putnam N, Ralph S, Rombauts S, Salamov A, et al. 2006. The genome of black cottonwood, Populus trichocarpa (Torr. & Gray). Science 313:15961604.

Vicoso B, Kaiser VB, 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 110:64538.

Wilson ZA, Song J, Taylor B, Yang C. 2011. The final split: the regulation of anther dehiscence. Journal of Experimental Botany 62:163349.

Xu J, Yang C, Yuan Z, Zhang D, Gondwe MY, Ding Z, Liang W, Zhang D, Wilson Z a. 2010. The ABORTED MICROSPORES regulatory network is required for postmeiotic male reproductive development in Arabidopsis thaliana. The Plant Cell 22:91107.

Zemp N, Minder A, Widmer A. 2014. Identification of internal reference genes for gene expression normalization between the two sexes in dioecious white Campion. PloS ONE 9:e92893.

Zhang D, Yang L. 2014. Specification of tapetum and microsporocyte cells within the anther. Current Opinion in Plant Biology 17:4955.

Zhang Y, Sturgill D, Parisi M, Kumar S, Oliver B. 2007. Constraint and turnover in sex-biased gene expression in the genus Drosophila. Nature 450:2337.

Chapter 4

The asparagus genome and the evolution of a sex chromosome

Alex Harkess, Chunyan Xu, Ron Van der Hulst, John Bowers, Saravanaraj Ayyampalayam, Francesco Mercati, Michael R. McKain, Atul Kakrana, Haibao Tang, Jeremy Ray, Paolo Riccardi, John Groenendijk, Tow Arikit, Sandra Mathioni, Mayumi Nakuno, Hongyan Shan, Alexa Telgmann-Rauber, Akira Kanno, Zhen Yue, Haixin Chen, Wenqi Li, Xiangyang Xu, J. Chris Pires, Bob Schmitz, Dave Kudrna, Rod A. Wing, Blake C. Meyers, Hongzhi Kong, Pierre Lavrijsen, Francesco Sunseri, Agostino Falavigna, Yin Ye, Guangyu Chen, James H. Leebens-Mack. To be submitted to *Science*.

4.1 Abstract

The majority of flowering plant species are classically described as being perfect, reproducing with flowers that harbor both male and female organs. In contrast, relatively few plant species are dioecious, where individual plants will produce exclusively either male or female flowers. In some cases dioecy can be controlled by a sex chromosome, as is the case for plants such as strawberry (ZW), asparagus (XY), papaya (XY), and white campion (XY), but the full suite of genes underlying gender determination on any plant sex chromosome have largely remained unknown. Here we present a genome assembly and annotation for garden asparagus (*Asparagus officinalis*), a plant with a young homomorphic XX/XY sex chromosome pair, and identify a 1.8Mb hemizygous, Y-linked region by comparing sibling doubled haploid YY and XX individuals. Through physical mapping, expression analysis, and mutant resequencing, we present strong evidence that sex determination in garden asparagus is mediated by two genes in a hemizygous, non-recombining region on the Y.

4.2 Introduction

On the shores of Naples in 1904, Geoffrey Smith discovered that while male and female *Inachus mauritanicus* crabs were parasitized at the same frequency, males that were castrated by a parasite developed secondary female characteristics and became hermaphroditic. In conjunction with early cytological evidence of sex differences (Morgan, 1914; Allen, 1917), this result framed the hypothesis that sex inheritance was a Mendelian trait inherited as homozygous in one sex and heterozygous in the other. Over the next hundred years of research, cytological and genetic evidence implicated the zygotic configuration of a sex chromosome pair as the dominant control for gender determination in many dioecious species across plants, animals, and fungi (Morgan, 1914, 1926).

Despite nearly a century of research in the genetic and ecological causes and consequences

of sex chromosome evolution across plants, several fundamental questions about their origins still remain unresolved. In particular, how are regions of non-recombination between X/Y (or Z/W) established from an ancestral autosome pair, how does that region change over time, and do we find parallels across divergent plant taxa with sex chromosomes? Sex chromosomes have evolved from autosomes hundreds of independent times across the flowering plant phylogeny (Ming et al., 2011), and researchers have developed several modern model systems to test the evolution of sex chromosomes and sex determining genes. Despite this diversity in model systems across plants, the exact suite of genes involved in gender determination remains unknown in any angiosperm.

A hypothesis formalized by Deborah and Brian Charlesworth (1978) suggests that the conversion from an autosome pair to a sex chromosome pair requires at least two genes linked perfectly in a non-recombining region; one gene must suppress female organogenesis, and another must promote the formation of anthers. Under this hypothesis, gynodioecy, or the co-occurrence of female and hermaphrodite forms, likely serves as an intermediate form between hermaphroditism and dioecy. Here we directly test this hypothesis about the origin of sex determination genes on a young sex chromosome using garden asparagus (*Asparagus officinalis* L.) as a model system, which is ideal as a sex chromosome pair appears to have evolved only once or potentially twice in the Asparagus genus (Kubota et al., 2012; Norup et al., 2015). The X and Y chromosomes are cytologically homomorphic (Deng et al., 2012), suggesting that this conversion occurred recently. Further, in garden asparagus, anther and pistil sterility events occur in two temporally separate events, suggesting the possibility of a two gene sex determination system.

Table 4.1: Initial garden asparagus SOAPdenovo2 assembly statistics

Statistic	Scaffold length (nt)	Contig length (nt)
N90	73,440	6,762
N80	128,075	9,634
N70	178,416	13,335
N60	235,886	17,100
N50	301,040	21,179
Longest	1,874,998	176,610
Total Size	1,162,656,628	1,109,535,501

4.3 Methods

Genome sequencing, assembly, and genetic mapping

To assemble the 1.3Gb garden asparagus genome into pseudomolecules, we utilized a combination of physical mapping, genetic mapping, and optical mapping (Figure 4.1). A YY doubled haploid *A. officinalis* individual (DH00/086) was sequenced to nearly 123X coverage using a diversity of short insert size and large mate pair Illumina libraries (Table 4.1). Reads were assembled using SOAPdenovo2 to produce a draft initial assembly.

To improve the initial assembly, roughly 20X of PacBio longread coverage and 1X of Roche 454 coverage was generated for gap-filling and scaffolding. High molecular weight DNA isolated from the reference DH00/086 individual was used as input for PacBio library preparation. The prepared library was size selected for >20kb fragments using the BluePippin (Sage Science, Beverly, MA) and sequenced on a PacBio RS II at the University of Florida Interdisciplinary Center for Biotechnology Research (ICBR). The assembly was subsequently improved utilizing PBJelly2.

To further improve the contiguity of the assembly, a 79X coverage BioNano Genomics optical map (using enzyme BspQI) was generated for the same YY reference individual. Molecules were assembled in IrysView and anchored onto the sequence scaffolds using the

Compare tool in IrysView.

To place sequence contigs into genetic bins on pseudomolecules representing each of the 10 haploid *A. officinalis* chromosomes, we generated a SNP-based genetic map by resequencing a population of 71 doubled haploid XX or YY individuals derived from performing colchicine anther culture on a single XY individual (K323). Each of the 71 individuals were resequenced to roughly 3X coverage using standard Illumina library protocols. The XY parent of the population was resequenced to 20X coverage, and the two doubled haploid parents of that XY individual (that is, the grandparents of the mapping population) were also resequenced to 10X coverage. This sampling strategy ensured that SNP validation in the mapping population can be verified through both the parent and one of the grandparents, which should significantly reduce the false positive SNP detection rate. Illumina paired-end 150nt reads from the mapping population individuals were quality assessed and adapter trimmed using batch scripts archived at https://github.com/alexharkess/bread_and_butter/tree/master/readCleaning and aligned to the reference pseudomolecules using bwa v0.7.5a. SNPs were called using samtools v1.2 mpileup and bcftools v1.2, filtering for homozygous SNPs with at least a read coverage of 2. Genetic map construction utilizing spreadsheet software (Microsoft Excel 2010) was performed following Bowers et al. (2010).

Gene and repetitive element annotation

To annotate the assembled pseudomolecules for gene content, we utilized a combined approach of RNA-seq based, ab initio, and ortholog-based gene identification. Illumina-based RNAseq reads from Harkess et al. (2015), which includes vegetative spear tips and floral bud isolations, were aligned to the unmasked reference pseudomolecules using TopHat2 (v2.0.14) with default parameters. Gene models were assembled with Cufflinks (v2.2.1) with default parameters. We used these gene models to develop a 5,000 gene training set for ab initio gene prediction with SNAP (v2010-07-28), PASA (r20130907), and AUGUSTUS (v2.5.5). We also

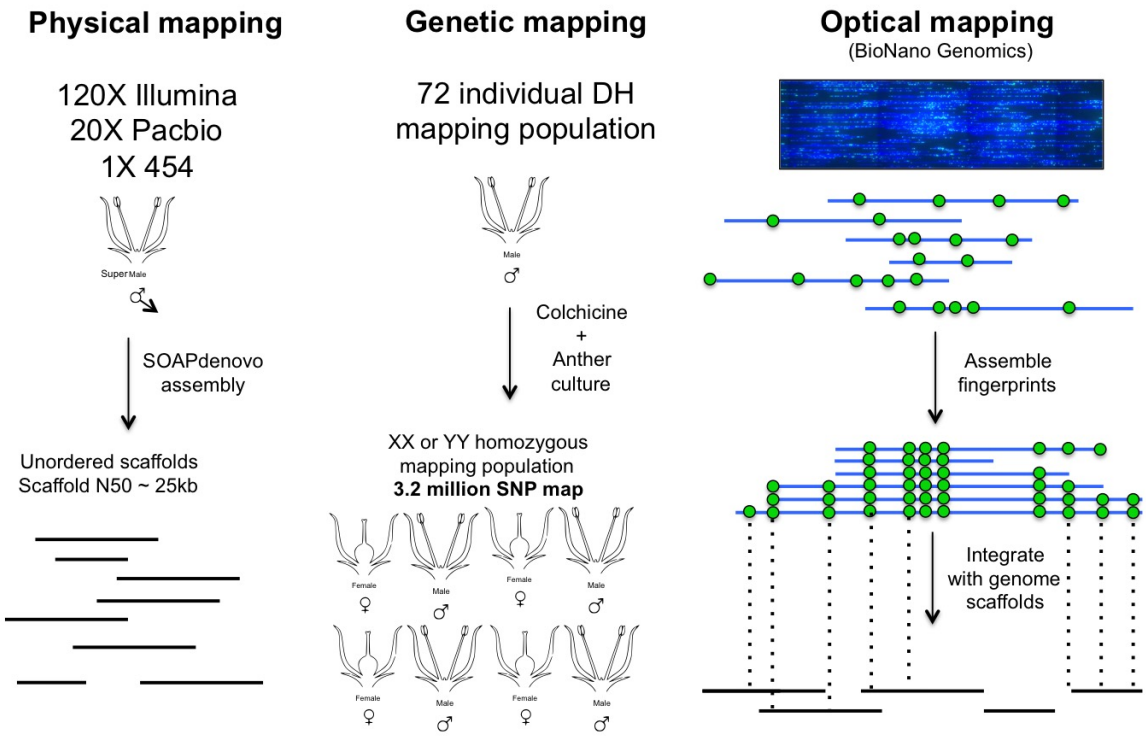

Figure 4.1: Overview of combined physical, genetic, and optical mapping strategy for *Asparagus officinalis* genome assembly into pseudomolecules

leveraged orthologous protein models from *Vitis vinifera*, *Arabidopsis thaliana*, *Phaealenopsis*, *Oryza sativa*, and *Musa acuminata* using exonerate (https://github.com/nathanweeks/exonerate) with the prot2genome alignment model to generate additional evidence for gene models. Evidence from the RNA-Seq, *ab initio*, and orthologous gene model annotations was combined using EvidenceModeler (r2012-06-25) to produce a unified set of gene model annotations.

To annotate the pseudomolecules for repetitive DNA content, we first utilized a custom RepeatMasker database generated by Harkess et al. (in review). We ran RepeatMasker v4.0.5 with parameters "-e wublast -pa 32 -lib Asparagus.repeatexplorer.annotatedRepeats.fa -no_is -nolow -html -gff". To annotate intact LTR retroelements, we used LTRharvest from GenomeTools v1.5.1 with default parameters except for "-similar 55 -maxdistltr 40000" to identify older LTR insertions as large as 40kb. The genome pseudomolecules were masked prior to generating the final gene set with EvidenceModeler.

Mutant generation and analysis of sex determination genes

We utilized a reverse genetic approach to generate male-to-hermaphrodite and male-to-female mutants in garden asparagus to identify the sex determination genes on the Y. To identify a gene that putatively suppresses femaleness on the Y, we first focused on the identification of a mutant that switched from an XY male to producing hermaphroditic flowers, indicating that a putative female suppression gene had been knocked out. In collaboration with Limgroup (Horst, Netherlands), a population of 40,000 XY male individuals from the K323 genotype were treated with low levels of Cobalt-60 gamma irradiation. All 40,000 individuals were planted and phenotyped for the presence of hermaphroditic flowers at Limgroup facilities. Additional populations of >5,000 seeds were also treated with Co-60 gamma irradiation, leading to the generation of five additional mutants (Lim_mut1 - Lim_mut5) that include both male-to-hermaphrodite mutants and male-to-female mutants. All individuals were resequenced utilizing standard Illumina library preparation protocols with the KAPA

DNA Library kit and sequenced on an Illumina NextSeq at the Georgia Genomics Facility to generate paired-end or single-end 150nt reads. Short reads were quality assessed, adapter trimmed, and quality trimmed utilizing batch FASTQC and Trimmomatic scripts available at https://github.com/alexharkess/bread_and_butter/tree/master/readCleaning. All cleaned Illumina reads were aligned with bwa v0.7.10 with default parameters.

4.4 Results and Discussion

Genome assembly and annotation

The genome of a doubled haploid YY asparagus individual (DH00/086) was sequenced and de novo assembled using 120X coverage Illumina reads, 20X coverage Pacific Biosciences (PacBio) reads, and a 66X BioNano Genomics optical map to yield an initial assembly. Contigs comprising nearly 94% of the estimated genome size were anchored and oriented onto 10 pseudomolecules (and an additional unordered bin) using a nearly 3.2 million SNP genetic map derived from a 72 individual doubled haploid XX and YY mapping population. Contigs within pseudomolecules were ordered and oriented utilizing combined evidence from the physical map scaffolds, the optical map scaffolds, and the genetic map bins. The combination of the optical and physical map identified nearly 3,200 misjoins in the assembly that were manually broken. 93.8% of the genome assembly contigs were placed into pseudomolecule bins, 29.3% of the contigs were able to be properly ordered and oriented on the optical and genetic maps, and 40.4% of the contigs were able to be ordered but not oriented on the map. To compare X and Y chromosomes, a sibling doubled haploid XX female (DH00/94) was resequenced to nearly 40X Illumina coverage and an additional >70X BioNano optical map was generated.

Utilizing a combination of expression-based, ab initio, and orthology-based annotations we predicted 27,656 genes in the *A. officinalis* genome. To judge the quality and com-

Table 4.2: Filtered Illumina Read Statistics for DH00/086 Reference

Insert size (bp)	Read Length (bp)	Total Data (Gb)	Depth (X coverage)
170	100	13.9	10.51
200	100	10.88	8.23
500	100	26.16	19.77
800	100	25.5	19.27
2,000	90/49	37.73	28.52
5,000	90	12.96	9.8
10,000	90	9.72	7.34
20,000	90/49	21.13	15.97
40,000	90	5.18	3.91
Total:		163.15	123.32

pleteness of the gene models, we compared the annotations to the 458 *Arabidopsis thaliana* genes in the Core Eukaryotic Gene Model Annotation (CEGMA) database and recovered hits for 95% the CEGMA genes, 66% of which were 100% full-length hits. Flow cytometry estimation of the genome size is approximately 1,330 megabases (Mb), nearly 72% of which is estimated by clustering random shotgun reads to be repetitive and largely dominated by a recent proliferation of Ty3-Gypsy and Ty1-Copia Long Terminal Repeat (LTR) retrotransposons coincident with the evolution of dioecy in the genus (Harkess et al., in review (Kuhl et al., 2005)). Pairwise nucleotide similarity comparisons of annotated LTR repeat ends also suggests that repeats have recently proliferated in the genus (Figure 4.2). Overall, utilizing a custom *A. officinalis* repetitive element database (from Harkess et al. in review), 72.38% of the genome pseudomolecule bases were masked using RepeatMasker, independently confirming the shotgun read approach to characterizing repetitive element content in the genome.

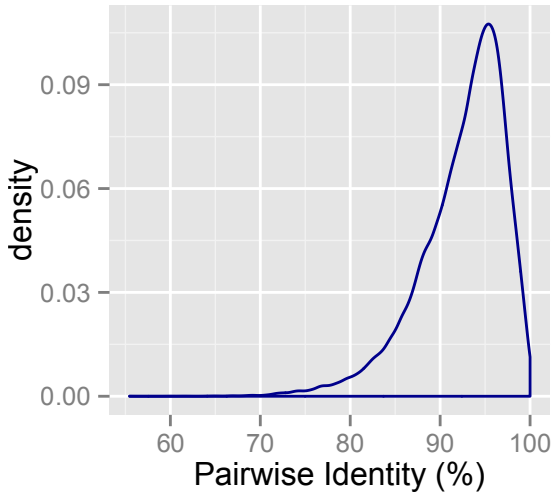

Figure 4.2: Distribution of percent identities from pairwise long terminal repeat (LTR) alignments across all annotated, full-length LTR retrotransposons

Identifying two sex determination genes on the Y

Several markers have previously been closely linked to the non-recombining, sex determination region in *A. officinalis*, but the exact delimitation and structure of the non-recombining region on the Y remained unclear (Jamsari et al., 2004; Telgmann-Rauber et al., 2007). We used a combination of SNP-based genetic mapping, physical mapping, optical mapping, and XX and YY accession resequencing to identify the exact boundaries of a nearly 1.8Mb hemizygous region of the Y chromosome that is largely missing from the X chromosome based on resequencing. A total of 14 genes were predicted in this region of the Y, 10 of which have annotations against Swiss-Prot (blastx, 1e-03 cutoff) Table 4.3.

Based on the model developed by Deborah and Brian Charlesworth (1978), we hypothesized that sex determination in garden asparagus may be mediated by two linked genes on the Y; one gene must suppress female organogenesis and another gene must promote the formation of anthers. These genes must also be missing or other non-functional or pseu-

Table 4.3: UniprotKB annotations for hemizygous, Y-linked genes

Gene model	Evalue	% ID	Annotation
01.231	9.00E-28	25.52	UPF0481 protein At3g47200
01.233	2.00E-04	33.33	Probable transcription-associated protein 1 tra1
01.235	6.00E-33	61.86	Ethylene-responsive transcription factor ERF086
01.236	1.00E-13	41.38	Nudix hydrolase 22, chloroplastic
01.240	6.00E-69	77.27	Transcription factor MYB35
01.241	4.00E-87	54.92	Jasmonic acid-amido synthetase JAR1
01.243	1.00E-38	34.11	Probable WRKY transcription factor 41
01.244	3.00E-124	34.34	Pentatricopeptide repeat-containing protein At2g13600
01.245	6.00E-45	41.23	Chitinase domain-containing protein 1
01.246	8.00E-92	75.57	Abscisic acid receptor PYL9

dogenized on the X chromosome. We identified several independent lines of evidence that implicate a single gene on the Y as being responsible for the suppression of female organogenesis. Three independently derived XY male plants that converted to hermaphroditism were used to deletion map several loss-of-function mutations that mutually include a single Y-linked gene; two mutants were derived from Co-60 gamma irradiation (Figure 3A, 3B), and a third was a spontaneous mutant identified by collaborators Agostino Falavigna, Francesco Mercati, and Francesco Sunseri. Due to the high concentration of repetitive elements in the non-recombining region of the Y, identifying the exact boundaries of the gamma radiation-induced deletions is informatically challenging with larger deletions. Regardless, the two gamma irradiation deletion mutants clearly implicate a single gene as being responsible for the gender conversion through blind analysis of read counts across all gene models. We first took a sequence coverage-based approach to identifying deletions in the mutants relative to the wild types. In the G033 mutant in which this gene was initially discovered, we screened all gene annotations containing 0 read counts in the mutant, and ¿5 read counts in the wild type male; this threshold was initially based on our expectation of this gene being in a repetitive region, coupled with the rather low sequencing depth of 8X coverage. A single gene

model across the entire sex-linked annotation had a 0 read count in its 2nd exon and ¿8 read counts in the wild type K323 individual. In a second, independent hermaphrodite mutant (Lim_mut2), the irradiation-induced deletion is also mapped to a roughly 300nt region in exon 2 of this same single gene. In the spontaneous mutant, there is a single nucleotide polymorphism in exon 2 that leads to a premature stop codon in the transcript. This is the only annotated gene in the region with clear deletions or premature stop codons across all hermaphrodite mutants relative to their wild type progenitors. The gene does not have a functionally annotated ortholog but contains a DUF247 domain (Domain of Unknown Function 247, PF03140), a domain that has been implicated as the male component to the self-incompatibility S-locus in perennial ryegrass. These hermaphroditic individuals display properly formed pistils, and are able to self-pollinate and produce viable berries, strongly suggesting that at least one more linked gene is likely required for male sterility in females.

To identify a gene on the Y that promotes anther development, we screened all 14 gene annotations in the hemizygous region for anther-related genes and identified a homolog of the Arabidopsis TAPETAL DEVELOPMENT AND FUNCTION 1 (TDF1, MYB35) gene. TDF1 is a Myb-class transcription factor that is required for early tapetum and pollen microspore development in Arabidopsis (Zhu et al., 2011; Gu et al., 2014). Anther sterility in asparagus XX females has previously been shown to occur sometime before or during pollen meiosis, and is marked by a clear destruction of the tapetal layer and the failure for pollen microsporogenesis to fully initiate (Caporali et al., 1994). Using RNA-seq expression analysis of the anther development pathway in XX female asparagus plants, we similarly hypothesized that TDF1 or a developmentally related gene is likely an upstream regulator of sex determination in *A. officinalis* (Harkess et al., 2015). Further, TDF1 is the only gene model in this region with an Arabidopsis T-DNA knockout that yields a male sterile phenotype or has previously been implicated in anther development.

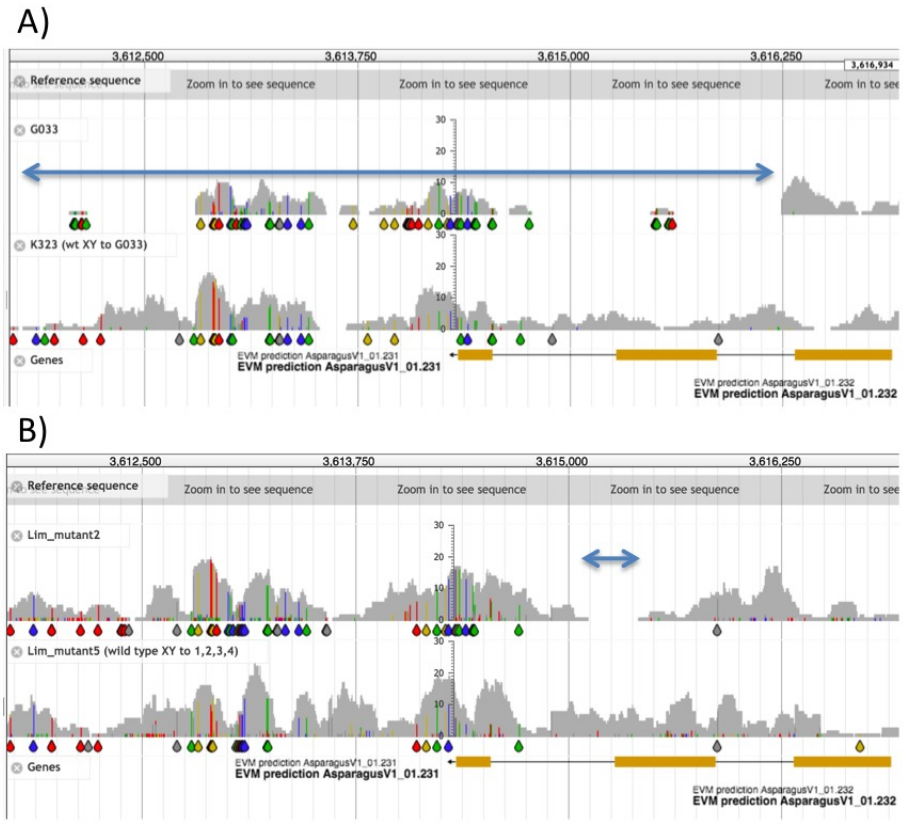

Figure 4.3: Short read alignments for gamma irradiation-derived male-to-hermaphrodite mutants for the sex-linked DUF247 domain-containing gene.

Structure and evolution of the Y chromosome

With a genome assembled into pseudomolecules, and the identification of two sex determination genes on the Y, the *Asparagus* genus now offers a unique system to study the conversion of an autosome into a sex chromosome. Specifically, by what mechanism does non-recombination between X and Y occur? By anchoring sequence scaffolds onto the YY optical map, we have identified a nearly 1.8Mb region that delimits the boundaries of the non-recombining, hemizygous region of the Y where the two sex determination genes reside. This region is flanked by recombining, pseudoautosomal region sequence (PAR) (Figure 4).

To understand the origin of the Y chromosome, we performed syntenic block analysis to identify remnants of whole genome duplications. Within-asparagus syntenic block analysis shows that a large portion of the Y chromosome is syntenic with chromosome 5, and this syntenic region spans the non-recombining sex determination region (Figure 5). Through *Ks* analysis, this syntenic block is largely derived from one of the two detectable whole genome duplication events in asparagus, which transcriptome-based analysis clearly indicates occurred before the evolution of dioecy and a sex chromosome in the genus (Harkess et al., in review).

Given that the TDF1 gene is single copy in the garden asparagus genome, only present on the hemizygous region of the Y, this leads to two possible hypotheses about the evolution of the male sterility component of the Y: either the TDF1 homolog was present on chromosome 5 before the whole genome duplication event, and has since been deleted on chromosome 5 and retained on the Y, or that the TDF1 gene has been recruited to the non-recombining region of the Y. Clarifying this hypothesis requires an assembly and annotation of a hermaphroditic ancestor, though.

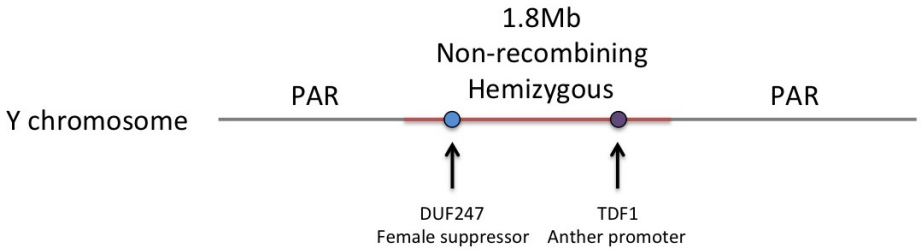

Figure 4.4: Depiction of the sex hemizygous, non-recombining sex determination region on the Y chromosome flanked by pseudoautosomal regions (PAR)

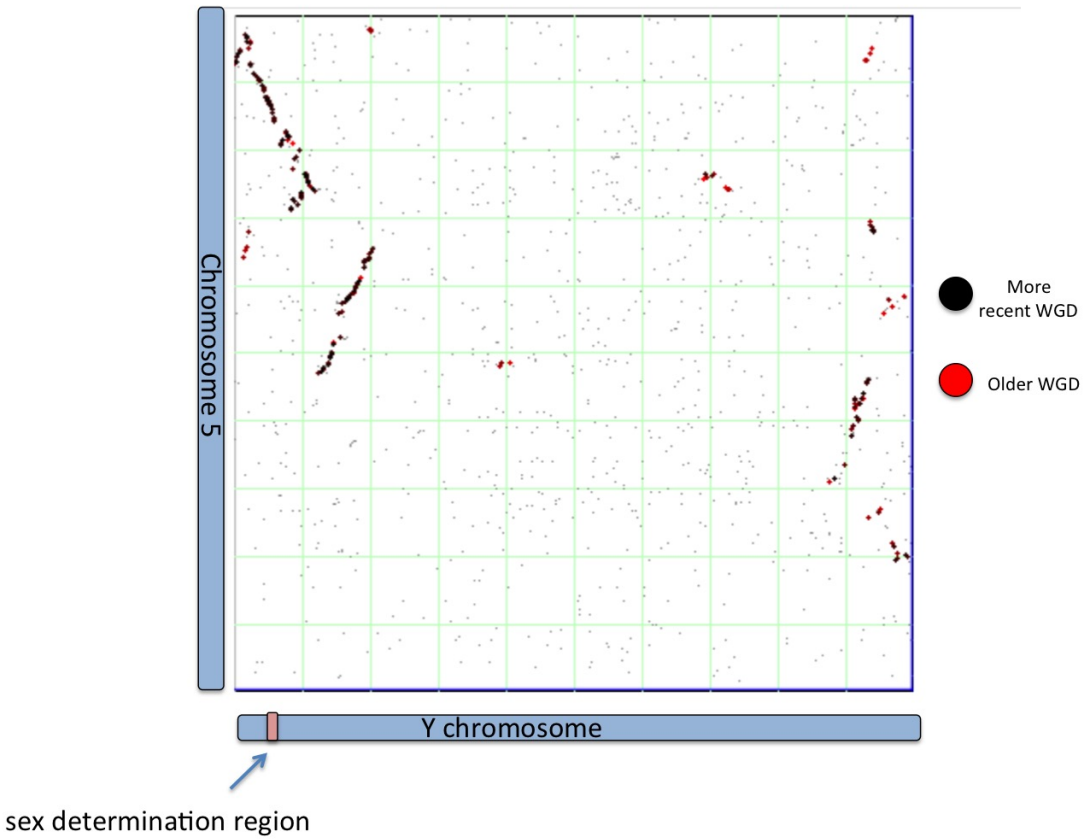

Figure 4.5: Synteny of gene model annotations between the Y chromosome and chromosome 5

4.5 Conclusions

The flower is a unifying reproductive structure across the angiosperms. Despite this general conservation, angiosperms have evolved myriad ways to alter the flower to yield massive variation in morphology, sexual systems, and mating systems. In particular, the conversion from hermaphroditic flowers to unisexual ones has repeatedly occurred across the angiosperm phylogeny. This conversion is often accompanied by the evolution of a sex chromosome that actively determines gender, allowing for the characterization of sex chromosomes across several different systems and time scales. This comes in great contrast to the mammalian Y, which shares a single origin nearly 180 million years ago, and is evident by its highly degenerate form relative to the X chromosome. All evidence of the earliest events in the evolution of the mammalian Y have been lost due to this degeneration.

Our high quality garden asparagus reference genome serves as an ideal non-grass monocot genome for studying both monocot evolution and the evolution of young sex chromosomes. In particular, the identification of two sex determination genes in a relatively small, 1.8Mb non-recombining region of the Y chromosome is the first genic confirmation of a nearly 40 year old hypothesis regarding the conversion of autosome to sex chromosomes (Charlesworth and Charlesworth, 1978). Given the homomorphism of the X and Y, this conversion from autosome to sex chromosome likely occurred recently; regardless, the initial degeneration of the Y chromosome is already evident by the expansion of young retrotransposons in this sex determination region. While the TDF1 gene has been functionally characterized as a tapetum-specific gene through mutant analysis in orthologous species, the DUF247 domain-containing gene remains functionally mysterious.

4.6 References

Allen CE (1917). A CHROMOSOME DIFFERENCE CORRELATED WITH SEX DIFFERENCES IN SPHAEROCARPOS. Science 46: 466467.

Caporali E, Carboni A, Galli MGM, Rossi G, Spada A, Marziani Longo GP (1994). Development of male and female flower in Asparagus officinalis. Search for point of transition from hermaphroditic to unisexual developmental pathway. Sex Plant Reprod 7: 239249.

Charlesworth, B., and D. Charlesworth, 1978 A model for the evolution of dioecy and gynodioecy. Am. Nat. 112: 975997.

Deng C-LL, Qin R-YY, Wang N-NN, Cao Y, Gao J, Gao W-JJ, et al. (2012). Karyotype of asparagus by physical mapping of 45S and 5S rDNA by FISH. J Genet 91: 20912.

Gu JN, Zhu J, Yu Y, Teng XD, Lou Y, Xu XF, et al. (2014). DYT1 directly regulates the expression of TDF1 for tapetum development and pollen wall formation in Arabidopsis. Plant J 80: 10051013.

Harkess A, Mercati F, Shan HY, Sunseri F, Falavigna A, Leebens-Mack J (2015). Sex-biased gene expression in dioecious garden asparagus (Asparagus officinalis). New Phytol 207: 883892.

Jamsari a, Nitz I, Reamon-Bttner SM, Jung C (2004). BAC-derived diagnostic markers for sex determination in asparagus. Theor Appl Genet 108: 11406.

Kubota S, Konno I, Kanno A (2012). Molecular phylogeny of the genus Asparagus (Asparagaceae) explains interspecific crossability between the garden asparagus (A. officinalis) and other Asparagus species. Theor Appl Genet 124: 34554.

Kuhl JC, Havey MJ, Martin WJ, Cheung F, Yuan Q, Landherr L, et al. (2005). Comparative genomic analyses in Asparagus. Genome 1060: 10521060.

Ming, R., A. Bendahmane, and S. S. Renner, 2011 Sex chromosomes in land plants. Annu. Rev. Plant Biol. 62: 485514.

Morgan TH (1914). Heredity and sex. Z Indukt Abstamm Vererbungsl 12: 159161.

Morgan TH (1926). Recent Results Relating to Chromosomes and Genetics. Q Rev Biol 1: 186.

Norup MF, Petersen G, Burrows S, Bouchenak-Khelladi Y, Leebens-Mack J, Pires JC, et al. (2015). Evolution of Asparagus L. (Asparagaceae): Out-of-South-Africa and multiple origins of sexual dimorphism. Mol Phylogenet Evol 92: 2544.

Telgmann-Rauber A, Jamsari A, Kinney MS, Pires JC, Jung C (2007). Genetic and physical maps around the sex-determining M-locus of the dioecious plant asparagus. Mol Genet Genomics 278: 22134.

Zhu J, Lou Y, Xu X, Yang Z-N (2011). A genetic pathway for tapetum development and function in Arabidopsis. J Integr Plant Biol 53: 892900.

Chapter 5

Conclusions and Discussion

5.1 Dioecy across the Asparagus genus

Our understanding of the evolution of dioecy and sex chromosomes is fragmented across several key model systems, such as asparagus, papaya, strawberry, persimmon, *Rumex*, *Populus*, *Salix*, and white campion. Each of these systems provides a unique opportunity to understand the conversion from autosome to sex chromosome as it accompanies the conversion from hermaphroditic flowers to dioecious, androdioecious, or gynodioecious flowers. Here we place the conversion from hermaphrodite to dioecious flowers in Asparagus within several contexts: the evolution of genomic content across the entire genus, the specific expression differences in male and female flower development in garden asparagus, and the identification of master sex determination genes on that Y chromosome.

Each of the chapters in this dissertation present different ways to view dioecy, all of which are critical to a more holistic understanding of the causes and consequences of the evolution of a sex chromosome. In addition, each chapter undoubtedly raises more questions than it answers. In chapter II, we observe that genome sizes in dioecious species precipitously increase in concert with an increase in repetitive elements, but we can only speculate as to

the role of sex chromosome evolution in that repetitive element proliferation. In Chapter III, we identify differential expression between male and female flowers, but can only hypothesize where in the anther development pathway might sterility occur in XX females. Our genome assembly and annotation in Chapter IV serves to synthesize these major questions in garden asparagus, though, as we reconstruct the Y chromosome and identify the genes involved in sex determination.

5.2 A two locus model comes to light

Nearly 40 years ago, Deborah and Brian Charlesworth hypothesized the necessity of mutations in two genes involved in the conversion from autosome to sex chromosome. As shown in Chapter I, this hypothesis had ecological and genetic roots stretching back more than 50 years earlier, based on the cytological, genetic, and morphological observations from botanists. Though this model is elegantly simple, it has long been under scrutiny given that the genes underlying these putative loci remained unknown in any dioecious species. Although we have long known that two "loci" control sex determination in *Silene latifolia*, which in many ways validates the Charlesworths' model, the genes underlying these loci have remained entirely mysterious. Our identification of two sex determination genes, TDF1 and a DUF247 domain-containing gene, is the first genic confirmation of the Charlesworth's 1978 model.

It is important to note that given the complexities of floral development, this model is certainly not the only way for dioecious species to evolve from a hermaphroditic ancestor. As we learn more about how gender is controlled in species such as cucurbits (Boulaem et al. 2008) and persimmon (Akagi et al. 2014), for example, we find that the genetic controls of dioecy can certainly be more complicated than a simple two gene model. To reiterate a point made in Chapter I by Winge (1937), The views as to where the genes are located and

whether a single gene or several genes be present in the respective [sex] chromosome suggest that the matter is more complicated than was at first supposed.

5.3 Remaining questions

This dissertation has largely answered a long-term question in the evolution of sex chromosomes, but this finding must be placed in an appropriate context. While we have identified the master sex determination genes in a single species, this must include the caveat that we have observed a single data point across the thousands of dioecious plants in the angiosperms. Consequently our results describe the evolution of a sex chromosome and sex determination genes only in *Asparagus*, which certainly differs from the evolution of other diverse sex chromosomes in both plants and animals. These findings will continue to grow in their importance as we place them in a comparative context with other dioecious systems. In particular, as we learn more about sex determination in dioecious genomes, do we find parallels between phylogenetically distant systems? While the sex determining genes are likely different across plants, are certain parts of the anther and ovule development pathways more prone to sterility mutations? We might speculate that traditionally single copy genes in this pathway may be especially subject to mutations, and these are hypotheses we can soon begin to test given the increase in available expression data for dioecious species.

Several burgeoning questions are perhaps obvious next steps in this research: what is the structure of the ancestral autosome, and the X chromosome? We are currently *de novo* assembling a sibling XX doubled haploid genome to identify the corresponding X region to the hemizygous region of the Y, identifying both evolutionary strata and the potential presence of X-specific genes. Simultaneously, we are assembling an additional dioecious genome for both an XY male and XX female Asparagus cochinchensis individual to test whether there has been expansion and/or contraction of the borders of the X chromosome relative to the Y.

These genomes are only useful in the context of phylogeny, though, and we are utilizing high-throughput gene capture approaches to build more complete coalescent-based phylogenies across the entire genus.

The Asparagus genus, and in particular garden asparagus, is poised to become a model system sex chromosome evolution. Future work must include forward functional genetics, such as the transformation of a CRISPR-Cas9 locus, for targeted genome editing and the functional verification of these sex determination genes. The future for garden asparagus is bright, though, especially as we uncover the genomic and ecological causes and consequences for the evolution of sex chromosomes in other species.

5.4 References

Akagi T, Henry IM, Tao R, Comai L (2014). A Y-chromosomeencoded small RNA acts as a sex determinant in persimmons. Science 346: 646650.

Boualem A, Fergany M, Fernandez R, Troadec C, Martin A, Morin H, et al. (2008). A conserved mutation in an ethylene biosynthesis enzyme leads to andromonoecy in melons. Science 321: 836838.

Winge, O (1937). Goldschmidts theory of sex determination in Lymantria. Journal of Genetics 34:81-89.

Chapter 6

Appendices

6.1 Chapter II supplemental figures and tables

Table 6.1: Transcriptome assembly and annotation

Species	Sexual System	Number of transcripts	Translated transcripts	Translated cDNA N50 (nt)
A. asparagoidies	Hermaphrodite	118,517	40,928	1,431
A. officinalis	Dioecious	276,556	158,386	1,482

Table 6.2: Number of paralogous and orthologous transcript pairs analyzed

Comparison	Pairs analyzed
A. officinalis x A. officinalis	225,035
A. asparagoidies x A. asparagoidies	11,818
A. officinalis x A. asparagoidies	326,687

Table 6.3: Estimated shotgun sequencing genome coverage

Species	Sequence data (nt)	Genome size (nt)	Estimated coverage
A. officinalis	7,955,329	1,370,000,000	0.0058
A. maritimus	14,586,484	1,310,000,000	0.0111
A. aphyllus	13,805,190	1,250,000,000	0.0110
A. stipularis	9,569,792	1,090,000,000	0.0089
A. falcatus	8,238,184	1,060,000,000	0.0078
A. virgatus	13,374,059	830,000,000	0.0161
A. pyrimidalis	16,872,384	720,000,000	0.0234
A. asparagoides	12,304,847	2,400,000,000	0.0051