CHARACTERIZATION OF THE SEX CHROMOSOMES IN SPINACH

BY

WILLIAM HERRING WADLINGTON

DISSERTATION

Submitted in partial fulfillment of the requirements
for the degree of Doctor of Philosophy in Plant Biology
in the Graduate College of the
University of Illinois at Urbana-Champaign, 2017

Urbana, Illinois

Doctoral Committee:

Professor Ray Ming, Chair, Director of Research
Professor Stephen Downie
Associate Professor Katy Heath
Assistant Professor Yoshie Hanzawa
Professor Patrick Tranel

ABSTRACT

Plant sex chromosomes have evolved from the autosomes of hermaphroditic species to maintain dioecy. The major barriers to sex chromosome characterization have been the inherent difficulties of working with a non-recombining sex determining region that make sequencing and mapping the Male-Specific region of the Y chromosome (MSY) painstaking. Here we use YY spinach (*Spinacia oleracea* L.) to characterize the MSY and the X-specific region of the X chromosome. A screen of 395 accessions in the USDA germplasm collection found a single accession that segregates YY progeny. The YY genotype was verified by a genetic cross and a novel X-specific marker. Potential novel X-specific sequences were found by depth of coverage analysis comparing alignments of male and female sequences to an XX reference. Of the 19 candidates found by depth of coverage, only one was verified as X-specific. The marker SpoX amplifies products from XX and XY but not YY templates. Pooled genomic DNA of 16 YY individuals selected by SpoX was sequenced at 63X using PacBio. Sequence data was assembled into an 823 Mbp assembly using CANU so the YY assembly could be compared to a 911 Mbp XX assembly. Seven genes that were non-repetitive sequences were found on a 1.14 X-specific contig from the XX assembly, and were queried by BLAST to the YY assembly to find their MSY homologs, which totaled 427 kb of novel Y chromosome sequence. On the Y contigs, the percentage uncovered by female k-mers decreased from 9.5% to 0.6% between positions 65.8 Mbp and 66.7 Mbp on the X chromosome, indicating that it is the location of one of the boundaries of the non-recombining region. The discovery of an accession which reliably segregates YY plants makes for unprecedented opportunity to study X and Y chromosomes in spinach. By comparing YY and XX genomes, the genomic basis of X and Y chromosome differentiation and evolution that gave rise to dioecy can be elucidated.

TABLE OF CONTENTS

CHAPTER 1: *SPINACIA OLERACEA* L. (SPINACH) IS AN IDEAL SPECIES FOR STUDYING PLANT SEX DETERMINATION

Abstract

Plant sex determination research has been an ongoing field for decades, but research of even the best studied dioecious species has been slow because of the difficulties of working with a non-recombining sex determining region, the hallmark of sex chromosomes. Spinach is an ideal species for sex determination research because it is YY viable and has a short generation time. Spinach is a dioecious species and has been studied for plant sex chromosomes since the mid-twentieth century. Classical genetic experiments indicated spinach has an XY sex chromosome system. However, there are exceptions to dioecy in spinach, including naturally occurring monoecy, and artificially induced sex changes. Genetic mapping of monoecy (M) using DNA markers located the M locus on the sex chromosome, but 12 cM away from the sex determining region. The spinach B and C class genes *SpPI* and *SpAG* are expressed in the primordia of male and female flowers, respectively, indicating that sex determination of the flower is set before the first whorl develops. Plant hormones affect sex expression of spinach, with gibberellic acid promoting male and cytokinin promoting female. Auxin and abscisic acid appear to promote female as well. The YY viability, variation in genetically controlled sex, and influence of hormones on sex expression make spinach an attractive system for studying sex chromosomes and sex determination in flowering plants.

Why Spinach?

By studying the structure of sex determining regions (SDRs) and identifying the sex determination genes on plant sex chromosomes, researchers can uncover the complex, stepwise evolutionary process that gives rise to dioecy (Ming and Moore 2007; Zhang et al. 2014). In flowering plants, using the XY system as an example, sex chromosome evolution begins with the development of two Y-linked sex determination genes. One gene supports the growth of stamen; this gene loses function on the X chromosome, thus making XX plants female, while XY plants can still produce pollen. The other gene suppresses carpel development, thus making flowers on XY plants female sterile. When the two sex determination genes are linked closely in one autosome, Stage 1 of sex chromosome evolution has been reached.

After Stage 1, a chromosomal rearrangement arrests recombination between the male and female sterility genes (Charlesworth and Charlesworth 1978). This results in a Male-Specific region of the Y chromosome (MSY) and a homologous X-specific region of the X chromosome that do not recombine with each other. After the arrest of recombination, the chromosomes are at Stage 2. During Stage 2, the sex chromosomes diverge in sequence as the MSY degrades. When the Y degrades to the point that the YY genotype is no longer viable, the sex chromosomes are defined as Stage 3. The repression of recombination between X and Y is necessary for maintaining dioecy but presents experimental difficulties for characterizing sex chromosomes.

After Stage 3, the lack of recombination between the MSY and X-specific region makes the sex determination genes unmappable, and the homologous yet divergent sequence on that region makes sequencing the XY genotype challenging. Unfortunately, the best studied plant sex chromosomes are at Stage 3. However, by working with a species with Stage 2 sex chromosomes, in which the YY genotype is still viable, both pitfalls of working with XY

genomes can be avoided. Spinach was recorded as being YY viable in early literature, and thus is of interest. In addition to YY viability, the practicality to grow, coexistence between dioecy with monoecy, and abundant data about factors that influence sex expression, make spinach a preferable system for plant sex chromosome research.

Spinach maintains dioecy with X and Y sex chromosomes

Although most flowering plant species have only hermaphrodite flowers, most spinach are dioecious, meaning individual plants have only unisexual female or male flowers. Female spinach flowers have 4-5 white stigmas and large sepals which persist and encase the fruit. Male flowers have 4-5 stamens with yellow anthers at the end of filaments. The small sepals of male flowers are hardly visible after anthesis. Spinach flowers have no petals (**Figure 1.1, Figure 1.2**). Flowers of dioecious spinach develop no trace of the opposite sex reproductive organs (Sherry et al. 1993)**.**

Table 1.1 Definitions related to plant flower types

Unisexual	Flower or individual is carpellate or staminate, female or male
Dioecious	Individual is exclusively male or female
Monoecious	Individual with both male and female unisexual flowers
Hermaphrodite	Flower or individual with both carpels and stamen
Gynomonoecious	Plant that develops female and hermaphroditic flowers

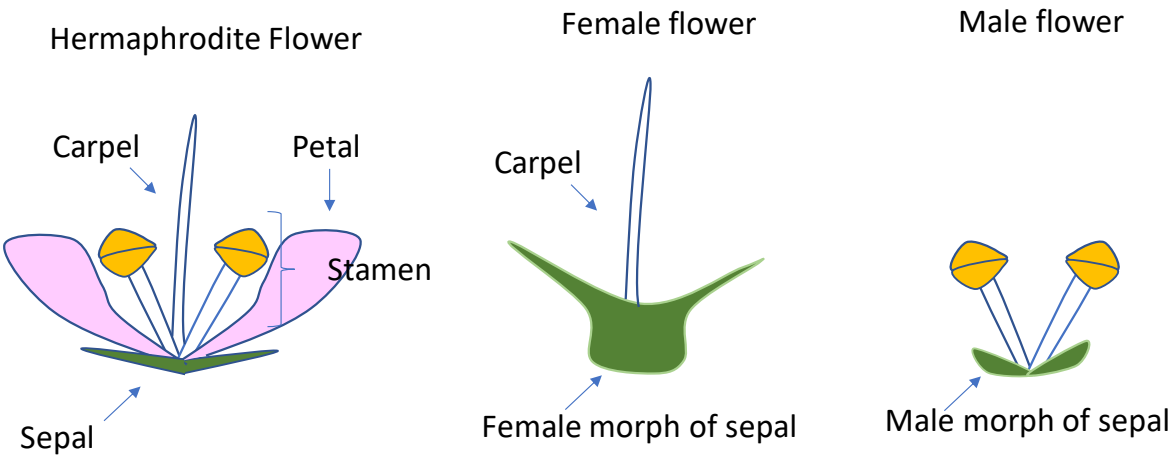

Figure 1.1 Cartoon of a typical complete flower and female and male spinach flowers. A hermaphrodite flower consists of sepals, petals, stamens and carpels. In comparison, the unisexual female flowers of spinach are comprised of persistent large sepals and carpels. The unisexual male spinach flowers have anthers and relatively small sepals. Neither sex of spinach flowers have petals.

In addition to sexual dimorphism in the flowers, spinach is also dimorphic in non-reproductive organs. Many plant species display sexual dimorphism to some degree (Geber et al. 2012), but few plants display the degree of sexual dimorphism that spinach does. Males flower before females. After bolting, males develop a branched inflorescence with elongated spike-like peduncles. Females bear axillary inflorescences. After anthesis, males senesce while females are relatively long lived. These adaptations allow males to excel as pollen parents.

Crosses between dioecious male and female spinach produce progeny with a 1:1 sex ratio. The approximately 1:1 sex ratio of dioecious spinach is maintained by sex chromosomes. In the 1950s, genetic experiments using polyploid spinach identified the male as the heterogametic parents, indicating an XY sex chromosome system (Mahoney et al. 1959). One copy of the Y is sufficient to enforce 100% staminate phenotype in XY, XXY and XXXY genotypes (Mahoney et al. 1959; Ellis and Janick 1960). Sex was genetically mapped to a locus

on the longest chromosome (Khattak et al. 2006; Yamamoto et al. 2014). The YY genotype is viable (Stevenson 1954). As is typical of Stage 2 plant sex chromosomes, the majority of the spinach sex chromosomes make up the recombining psuedoautosomal regions.

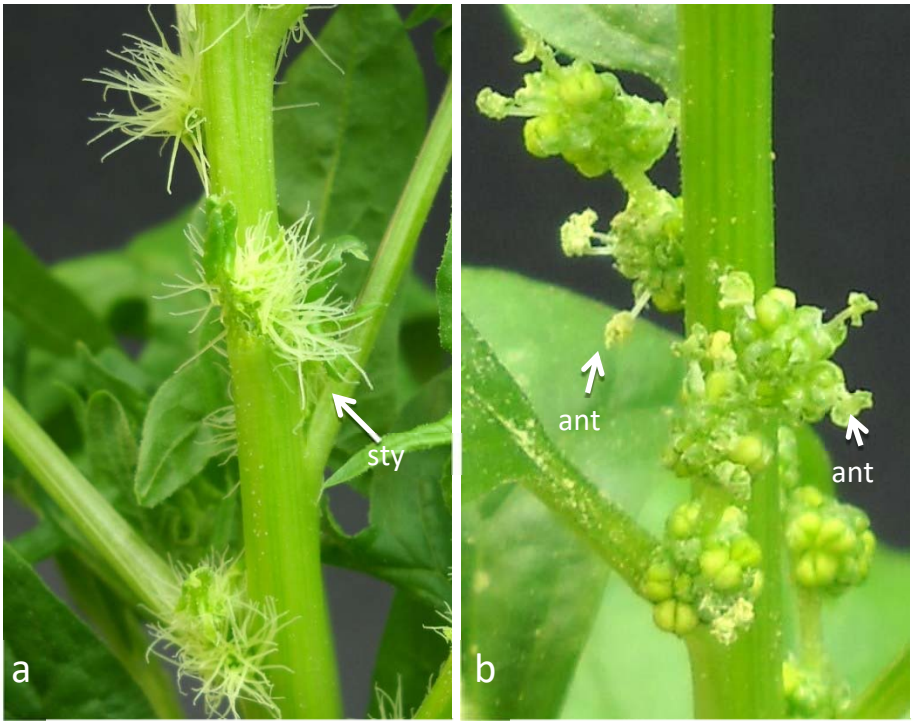

Figure 1.2 Female and male flowers on dioecious spinach. Inflorescences on female a) and male b) spinach from accession PI 604782. White styles (sty) are indicated on the female flowers, while extended anthers (ant) on filaments indicate male flowers

The pachytene X and Y chromosomes cannot be discerned under a microscope. This indicates spinach has homomorphic sex chromosomes (Ramanna 1976). Although the X and Y chromosomes are the same size, the sequences have begun diverging as is evidenced by the discovery of male-specific sequences. PCR markers with male-specific banding pattern target the MSY (Akamatsu and Suzuki 1999). Sex chromosome probes used in Fluorescence In Situ Hybridization (FISH) has shown that the X and Y loci have differences in probe binding patterns (Deng et al. 2012). The existence of male-specific markers targeting the diverging X and Y

sequences indicates the presence of a nonrecombining region. The nonrecombining MSY contains the sex determination genes and the sexual dimorphism genes, which are responsible for the numerous morphological differences between male and female dioecious spinach**.** However, there are some spinach that do not use the XY sex chromosomes system to regulate sex expression.

Spinach is not strictly dioecious

Spinach cannot be called a dioecious species without a modifier because individuals in this species are not always unisexual. Some frequency of monoecy (**Table 1.1, Figure 1.3**) was observed in at least 125 of the 398 spinach accessions in the 2013 USDA germplasm collection (https://npgsweb.ars-grin.gov/gringlobal/descriptordetail.aspx?id=219023). Monoecy is heritable and has been research by two independent groups. Jules Janick's group at Indiana University worked in the 1950s to characterize the inheritance of the trait using morphological markers (Janick and Stevenson 1955a), and a group at Hokkaido University has been working for the last decade to map monoecy using SSR and RFLP markers (Onodera et al. 2011; Yamamoto et al. 2014). Although the two groups are used different genetic backgrounds, their results on the heritability of monoecy are concordant.

Figure 1.3 Monoecious inflorescences. Inflorescences of a monoecious spinach from accession Ames 20168 baring both sexes Styles (sty) indicate female flowers, while anthers (ant) on filaments (fil).

Both groups found that monoecy is controlled by a single locus (M). The M allele functions by supporting the development of male flowers in XX individuals. In the XX background, M is partially dominate, as MM individuals have a higher percentage (50-60%) of male flowers than do Mm plants (30-40%); mm are female (Janick and Stevenson 1955a; Onodera et al. 2011). In an XY background, the monoecious gene has no expressed phenotype because all the flowers on the XY plant are male even without the contribution of the M allele.

Monoecy is not the only heritable exception to dioecy found in spinach. True-breeding gynomonoecious **(Table 1.1)** spinach have also been reported (Onodera et al. 2008). Gynomonoecy was found to be controlled by a recessive gene. Initial attempts to map the gene found the trait is not sex linked, but it has not yet been mapped to any other linkage group. Although dioecy and monoecy are likely to have evolved naturally before domestication, gynomonoecy is most likely caused by a recent, loss-of-function mutation (Onodera et al. 2008). Although, the genetic architecture describing the epistatic dominance of the MSY over monoecy is well understood, the molecular mechanism which controls either breeding system is unknown.

However, there are many clues to what type of genes may be involved in sex determination, revealed by numerous horticultural and molecular studies on sex expression in spinach. These studies give clues to which metabolic pathways or developmental processes the sex determination genes effect.

Physiology affects sex expression

In addition to genetically determined sex, external factors, including hormone treatment, photoperiod, heat stress, soil moisture, and pruning have been shown to cause significant biases in sex ratios (Freeman et al. 1980; Al-Khayri et al. 1992; Komai et al. 2003). In the 1970s, Mikhail Chailakyan, who coined the term florigen, experimented with spinach to find if phytohormones affected sex in dioecious species. Although untreated populations of spinach have a 1:1 sex ratio, populations of spinach grown in hydroponic solution supplemented with gibberellic acid were 78.8% male, while plants treated with artificial cytokinins were 86.7% female (Chailakhyan 1979). In artificial auxin and abscisic acid treated solution, plants were 76.0% and 71.0% female, respectively. This simple set of studies is strong support that hormone signaling may play a role in spinach sex determination. Studies also found that pruning can caused biased sex ratios. Spinach populations with primary and adventitious root tissue completely removed were 85% male, while populations where primary roots were removed then allowed to grow adventitious roots were 85% female (Chailakhyan and Khryanin 1979). The effect of defoliation on sex expression was also tested. When populations were defoliated, leaving only the top three flowers, 84.5% of plants were female.

ABC genes are expressed in a sex-specific fashion

The sex determination genes in the MSY or M locus are likely upstream factors that control if flowers are male or female. To identify the actual genes that determine male or female organ development, two of the spinach ABC MADS-box transcription factors were studied. Transcript of the spinach B class gene *SpPISTELLATA* (*SpPI*) was found in the stamen (Pfent et al. 2005) with only minor transcript detected in the carpels, while the C class gene *SpAGAMOUS* (*SpAG*) was detected in stamens and carpels (Sather et al. 2005). Ectopic expression *SpPI* was sufficient to control the sex of a plant, regardless of the background sexual genotype (Sather et al. 2010). Both genes are expressed in the floral primordia of the respective sex, indicating that sex determination of the flower is set before the first whorl develops. This would be expected as the male and female flowers differ in morphology even at the early stage of development.

In this dissertation

Because of the advantages discussed above, spinach is a desirable plant species for sex determination research. Surprisingly, when this project began, no meaningful effort was underway to sequence the X or Y chromosomes. In this dissertation, two research chapters describe the most comprehensive study of the spinach sex chromosomes to date. Chapter 2 describes the discovery and characterization of a novel sexual phenotype found in the diverse USDA spinach germplasm collection. In that accession, XY plants have some hermaphrodite flowers. The discovery of an androdioecious accession of spinach allowed for several avenues of research not before possible because this accession segregates YY spinach, which was previously reported only in one occurrence in 1955 (Janick and Stevenson 1955b). The reliable availability of YY DNA allows for research on the MSY without the presence of the X

chromosome. YY template was used to validate a novel X-specific marker, called SpoX, which was useful to identify the first X-specific sequence of the spinach sex chromosomes and used to identify YY spinach for whole genome sequencing.

Chapter 3 describes the X-specific region identified by SpoX and compares it to homologous sequences on the Y chromosome. There is a paucity of genes on the 1.1 Mbp contig identified by SpoX. The region SpoX locates was also found to contain the physical boundary of the MSY with the PAR. The contributions described here, including the discovery of a novel phenotype, novel marker, and the YY assembly lay the groundwork to bring spinach to the forefront of plant sex chromosome research.

CHAPTER 2: A NOVEL X-SPECIFIC MARKER IDENTIFIES YY INDIVIDUALS IN AN ANDRODIOECIOUS ACCESSION OF *SPINACIA OLERACEA* L. (SPINACH)

Abstract

The molecular mechanisms that maintain dioecious breeding systems have been studied in numerous plant species. However, the difficulties of working with non-recombining sex determination region (SDR) of XY individuals have hindered the progress towards sequencing sex chromosomes of most dioecious species. Here we present important advances toward characterizing the sex chromosomes of spinach. Of nearly 400 screened accessions, we identified a single accession of spinach in which androdioecious XY individuals segregate YY spinach. Genomic templates extracted from YY individuals were used as a negative control to validate an X-specific marker found by depth of coverage analysis. Of 19 possible X chromosome sequences found by depth of coverage analysis, one was verified to target X-specific sequence as it amplified product from XX and XY, but not YY templates. That marker, SpoX, identified the first reported spinach X-specific sequence. It was used to make the connection between the androdioecious phenotype of XY individuals and the pure male phenotype of YY individuals. The sex reversal of the XY mutant to hermaphrodite is strong evidence that the sex chromosomes of dioecious spinach have a two-gene sex determination system. These results are crucial towards completing the sequence of the spinach X chromosome and generating YY material that can be used for sequencing the Y chromosome.

Introduction

How and why 6% of angiosperm species (Renner 2014) came to evolve and maintain dioecious breeding systems have been a curiosity since the time of Darwin (Darwin 1877). To acquire a deep understanding of how sex chromosomes reinforce dioecy, the specific sex determination genes and sex chromosome sequences must be characterized. In recent decades, the advent of genomics has shed much light on the composition of plant sex chromosomes (Ming and Moore 2007; Zhang et al. 2014), however, progress has been limited. The best studied plant sex chromosomes are those of papaya (*Papaya carica* L.) and white campion (*Silene latifolia* Poir.). Papaya has the first sequenced plant sex chromosomes, which required a painstaking BAC-by-BAC cloning strategy to separately obtain the Hermaphrodite and Male-Specific region of the Y chromosome (HSY and MSY) and their X chromosome counterpart (Gschwend et al. 2012; Wang et al. 2012; Vanburen et al. 2015) .

Though the evolution of the papaya sex chromosomes has been well defined, the specific genes responsible for sex determination have yet to be described (Wang et al. 2012). Much work has also been done to study the large MSY of white campion, but no sex determination genes have been identified there either (Slancarova et al. 2013). The major constraints to studying the MSY in these species are that the MSY, by definition, does not recombine with the X chromosome and contains distinct sequences from the X-specific region of the X chromosome. The lack of recombination between the X and Y in the non-recombining region makes genes on the MSY unmappable, and the difficulty of separating homologous X and Y sequences in whole genome assemblies makes the XY genotype difficult to sequence. These problems could potentially be worked around by researching species with early stage sex chromosomes, in which the YY genotype is viable.

Sex chromosomes evolve in stages (Ming et al. 2011). In the first stage, sex determination genes evolved in a linked region that is still recombining, a pair of prototype sex chromosomes. One sex determination gene suppresses carpel development, the other promotes stamen development. Stage 2 is defined after recombination is repressed between the two sex determination genes, but before the X and Y diverge to the extent that the Y loses genes essential for viability. Thus, species with Stage 2 sex chromosomes are YY viable. By expanding sex chromosome research into dioecious species with Stage 2 sex chromosomes, the MSY can be analyzed without the constraint of a non-recombining region in XY chromosomes.

YY viability is rare, and spinach is among those species with the YY genotype still being viable, indicating that the spinach sex chromosomes have only evolved to sStage 2 of sex chromosome evolution (Stevenson 1954). The small stature, short generation time, a large germplasm collection, and ability to be transformed makes spinach a desirable species for genomic research. These advantages and the fact that spinach is primarily dioecious have made spinach the subject of plant sex determination studies since the mid-twentieth century (Stevenson 1954; Janick and Stevenson 1955b; Janick et al. 1959; Mahoney et al. 1959; Ellis and Janick 1960; Iizuka and Janick 1962; Ramanna 1976; Akamatsu and Suzuki 1999). The approximately 1:1 sex ratio of dioecious spinach is maintained by a pair of sex chromosomes. The male was identified as the XY parent while females are XX (Janick and Stevenson 1955a, b). The sex chromosomes of spinach are typical of primitive plant sex chromosomes, where the non-recombining SDR is only a small portion of the entire sex chromosome (Ming et al. 2007; Onodera et al. 2011). The MSY has been physically mapped to chromosome 1 (Iizuka and Janick 1963; Akamatsu and Suzuki 1999; Ito 2000; Khattak et al. 2006; Onodera et al. 2011). At least 6

male-specific PCR markers have been found, suggesting suppressed recombination of the MSY (Akamatsu and Suzuki 1999; Lan et al. 2006; Deng et al. 2013).

Though spinach is mostly dioecious, there is a diversity of sexual phenotypes in this species. The most common exception to dioecy is monoecy **(Table 2.1)**, which occurs in varying frequencies in most accessions in the USDA germplasm. Monoecy (M) is heritable and has been roughly mapped to chromosome 1. The M allele is dominant and causes an XX individual to have some male flowers. One copy of the M allele is sufficient to cause an XX individual to develop some male flowers, while two copies of the M allele increases the ratio of male to female flowers on the monoecious individuals (Janick and Stevenson 1955a; Onodera et al. 2008). Heritable gynomonoecy **(Table 2.1)** has also been described in a few Japanese varieties (Onodera et al. 2008). In addition to heritable exceptions to dioecy, plantlets regenerated from spinach calluses by tissue culture develop some hermaphroditic flowers on regenerated XY plants. The array of sexual phenotypes found in spinach provides an excellent system for sex determination research. However, despite recent advances in sequencing technology, little progress has been made to sequence the spinach MSY, and none of the sex determination genes involved in either dioecy or monoecy have yet been identified. No progress has yet been made to sequence the X-specific region of the X chromosome or predict the size of the SDR.

Table 2.1 Sexual Phenotypes of Spinach

Dioecy	Individuals are unisexual with only male or female flowers
Monoecy	Having both unisexual female and male flowers on the same individual
Gynomonoecy	All individuals have both unisexual female and hermaphrodite flowers
Androdioecy	Individuals are either unisexual females or have both male and hermaphrodite flowers, here individual plants with male and hermaphrodite flowers are referred to as androdioecious

We used a field screen of the diverse USDA germplasm collection to identify a novel sexual phenotype, androdioecy **(Table 2.1)**, where XY individuals bear some hermaphrodite flowers. The androdioecious individuals are self-fertile and can produce seed for YY individuals. The YY individuals in this accession were pure male, lacking the hermaphroditic flowers found on XY individuals. An X-specific PCR-based marker was used in conjunction with controlled crosses to connect chromosomal genotype with sexual phenotype. That marker, SpoX, can be used by researchers or breeders to discern XY from YY spinach. SpoX also identified the first X-specific sequence found in spinach. Identifying a novel sexual phenotype which produces the YY genotype and the first sequences of the X-specific region of the X chromosomes are major steps towards resolving the genetic basis of sex determination in spinach.

Materials and Methods

Plant Growth Conditions

In spring 2013, 395 accessions of spinach from the USDA germplasm collection were grown in the field at the University of Illinois at Urbana-Champaign Vegetable farm, Champaign, Illinois. Notes were taken on any morphological traits which varied between accession from April to July (**Appendix A, Appendix B, Appendix C, Appendix D, Appendix**

E, Appendix F). One accession, PI 217425, was noted as having 'pistillate male' individuals. To verify and characterize the expression of the 'pistillate male' phenotype, flats of 18 pots were grown in the greenhouse. Sexual phenotype of the flowers and seeds in PI 217425 were imaged using a Mighty Scope 5.0M digital microscope by Aven to confirm that flowers on those plants were unisexual male, unisexual female, or hermaphrodite. Only one individual from the original seed stock was pure male. That one pure male was crossed with a female. Seven progenies of that cross were grown in a growth room in 16 hour days under ambient temperature. One male was selected to develop the inbred line. The parent was self-pollinated. Seeds were bulked and grown for 7 generations in a growth chamber.

Identifying X-specific sequence in an XX reference

Whole genome sequence data of two male and two female individuals were used to identify X-specific sequence in an XX reference. The genomic DNA of one male and one female from the Syrian cultivar Shami (PI 445783) and one male and female from USDA accession PI 664497 (Chinese) were isolated and separately sequenced. The sex of each of the four individuals was determined by observing floral morphology. Genomic DNA was extracted separately from leaves of each of the four individuals using the CTAB method. Tru-seq Fast Model PE 2*150 nt genomic sequencing libraries were prepared for each of the four individuals at half volume of the manufacturer's instruction. Sequences were generated on a HiSeq 2500. The female and male from Shami were each sequence at approximately 17X coverage with 63867790 female and 59769899 male raw reads. The female and male from PI 664497 were sequence at approximately 12X coverage with 11.38 Gbp with 51295985 female and 45175968 male raw reads.

Male and female short reads were used to identify X-specific sequence by depth of coverage analysis. Males have only one copy of the X chromosome, while females have two copies. Thus, when whole genome sequences of male and female individuals are aligned to an XX reference, the X-specific scaffolds will have lower depth of coverage in male alignments than in female alignments (**Figure 2.1**). This depth of coverage difference was used to find X-specific sequence. To find sex-specific copy number variants (CNVs), sequences of two males and two females were separately mapped to the XX draft genome using Burris-Wheeler Algorithm (Li and Durbin 2009; Dohm et al. 2014). Reads with a map quality score of less than 20 were removed. Duplicate reads were removed by Piccard tools program AddOrReplaceReadGroups (http://picard.sourceforge.net./). The program CNV-seq was used on the cleaned alignments of 123637689 female reads and 96471953 male reads to detect sex-specific CNVs (Xie and Tammi 2009).

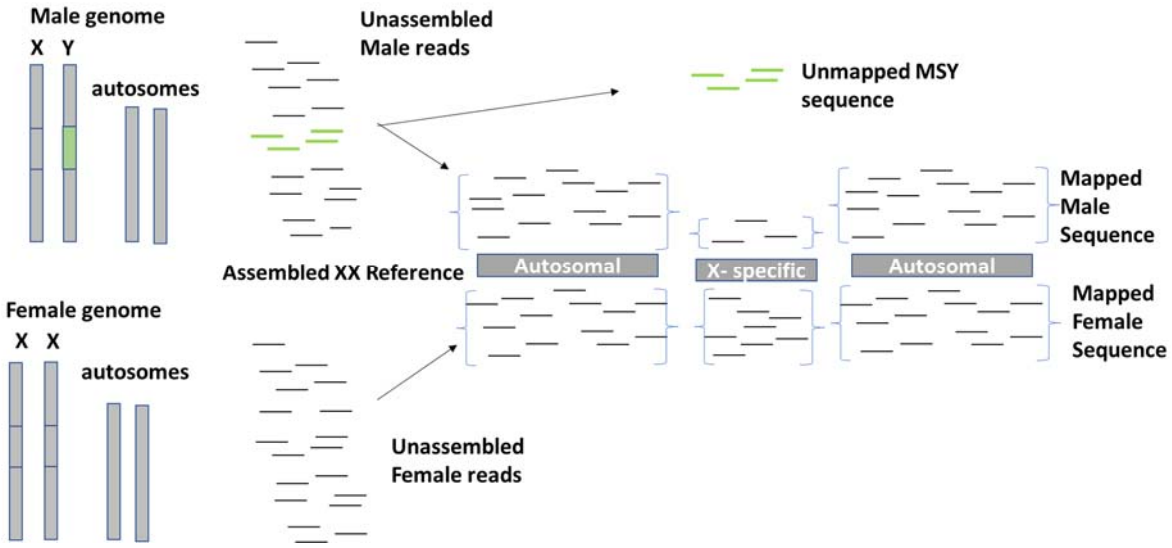

Figure 2.1 Depth of coverage analysis to find X-specific sequence in a female reference genome. The female genome contains X chromosome and autosomal sequences. The male genome contains the same X and autosomal sequences (redundant sequences in gray), but the male genome has only one copy of the X-specific region of the X chromosome. The male genome also contains MSY sequence (green), which is unique to the male genome. When female and male genome sequences are mapped to a female reference, the female and male alignment patterns will be identical for autosomal sequences. However, a different alignment pattern will be detected for the X-specific scaffolds where the two copies of the female genome align, but only one copy of the male does. The MSY reads do not map to the X. X-specific scaffolds will have half the coverage when aligned to by male sequences compared to female sequences.

To eliminate false positives when calling X-specific sequence, markers designed for 10 scaffolds most likely to be X-specific were tested on XX, XY, and YY genomic templates. Genomic DNA was extracted from a female (XX), an androdioecious (XY), and a pure male (YY) individual using the CTAB method. The one primer pair found to have a sex-specific banding pattern, SpoX, was used in multiplex with the published male-specific marker T11A (Akamatsu and Suzuki 1999) on XX, XY, and YY genomic template: SpoX F TAGCTGAAAGCGAAAACTGA, SpoX R AGACACAGAGCGCATAATGA. PCR was done following the manufacturer's instructions using the GoTaq kit, from Promega. PCR conditions were as follows 20 min 95°C, followed by 31 cycles of 30 seconds at 55°C, 80 seconds at 72°C,

and 15 seconds at 95°C. The 31 cycles were followed by 5 minutes at 72°C.0. Products were visualized after electrophoresis on a 1.5% agarose gel.

Results

Discovery of YY Spinach

YY genomic DNA is needed to validate X-specific markers. No method for generating YY spinach was available, so a screen was done of the entire USDA spinach germplasm collection to find a novel genetic resource which segregates YY spinach. The screen found a spinach accession PI 217425 (Cornell #9) in which individuals with panicles like those of males (having suppressed vegetative bracts and elongated, spike-like peduncles) bear flowers with carpels. Most flowers on these plants are male; however, some are hermaphrodite flowers, having both male and female reproductive organs on the same flower (**Figure 2.2, Figure 2.3**). The hermaphrodite flowers are self-fertile. They produce seed that looks noticeably different from the fruit of female flowers, which have green persistent sepals that incase the utricle **(Figure 2.4)**.

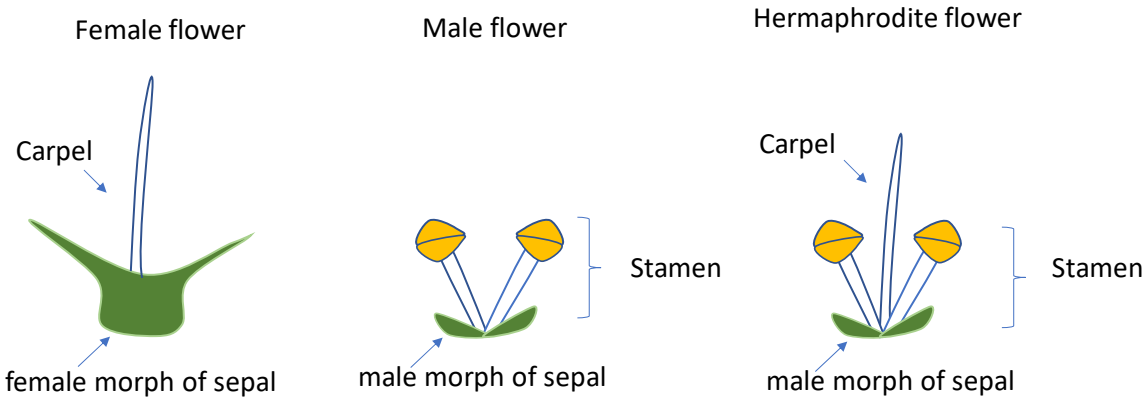

Figure 2.2 Cartoon of Female, Male, and Hermaphrodite Spinach Flowers. The unisexual female flowers of female and monoecious spinach are comprised of large persistent sepals which often develop a pair of spiny horns. There are no traces of petals or stamens, but a slender white carpel is present. The unisexual male flowers on male and monoecious spinach have small sepals, which do not persist after anthesis. Male flowers develop anthers and have no trace of petal or carpel primordia. The hermaphrodite flowers found in the androdioecious accession Cornell #9 are similar to male flowers in terms of the morphology of the sepal, but distinct because of the presence of a carpel.

Fertilized hermaphroditic flowers bear small brown seeds as both the green sepals and fruit are absent (**Figure 2.4**). Androdioecy is the best term for a breeding system comprised of unisexual females and males with some hermaphrodite flowers. This is in contrast to monoecy where individuals have separate unisexual male and female flowers (**Figure 2.2**). All but one of the individuals in the original lot of seed from the USDA were either pure female or had a mixture of male and hermaphrodite flowers. One individual was pure male. The one pure male individual was crossed with a female (XX). The seven offspring were all male, indicating the pure male was YY, not XY ($p = .0078$).

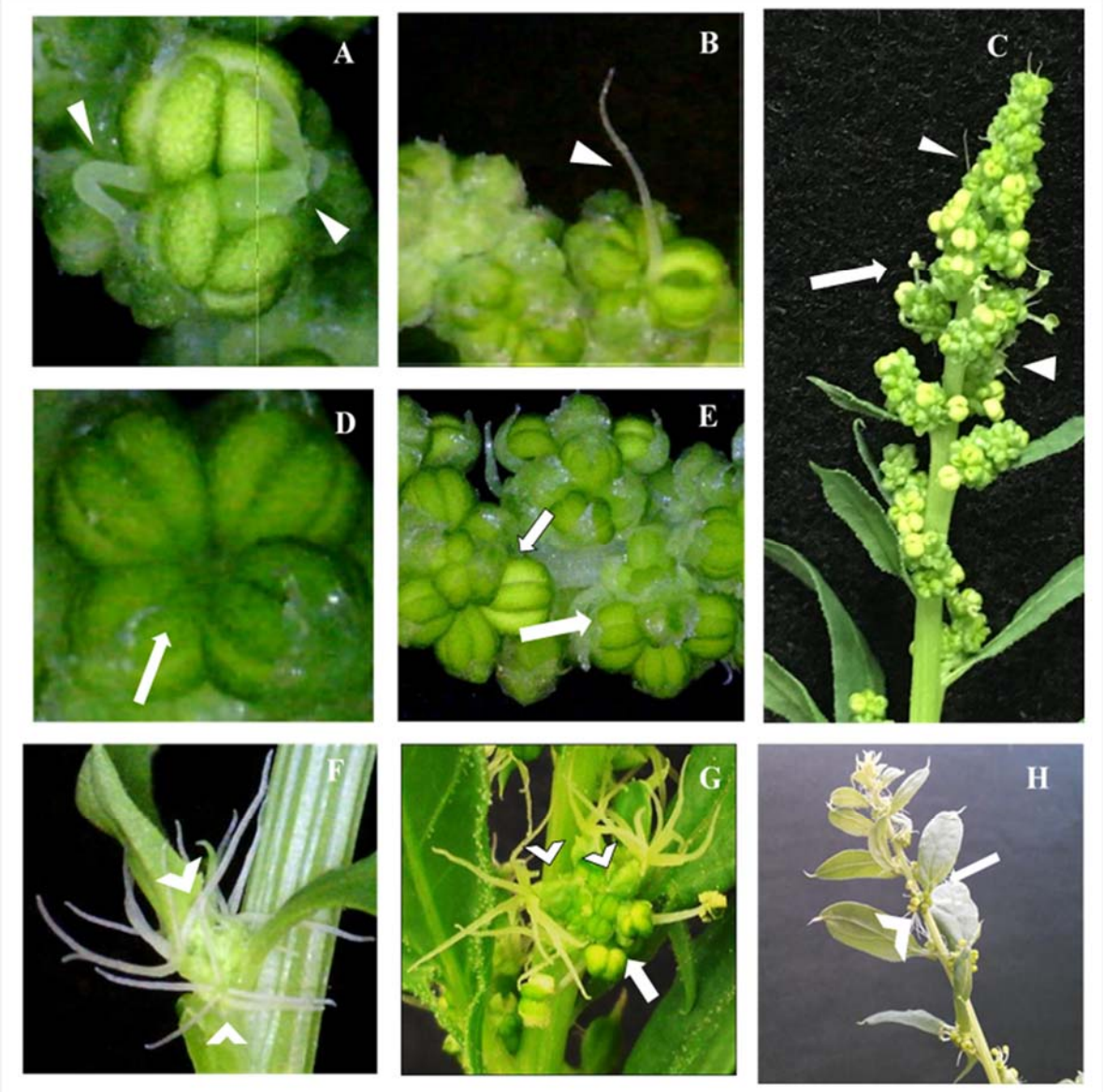

Figure 2.3 Comparison of androdioecious, male, female, and monoecious morphology. a-b. Hermaphroditic flowers of an androdioecious individual. c. Bolting androdioecious inflorescences. d. Male flowers. f. A female thyrse. g. A monoecious thyrse. H. A monoecious individual. White wedges point to the pistils on the hermaphroditic flowers. Male flowers are indicated by long arrows. Female flowers are indicated with chevrons.

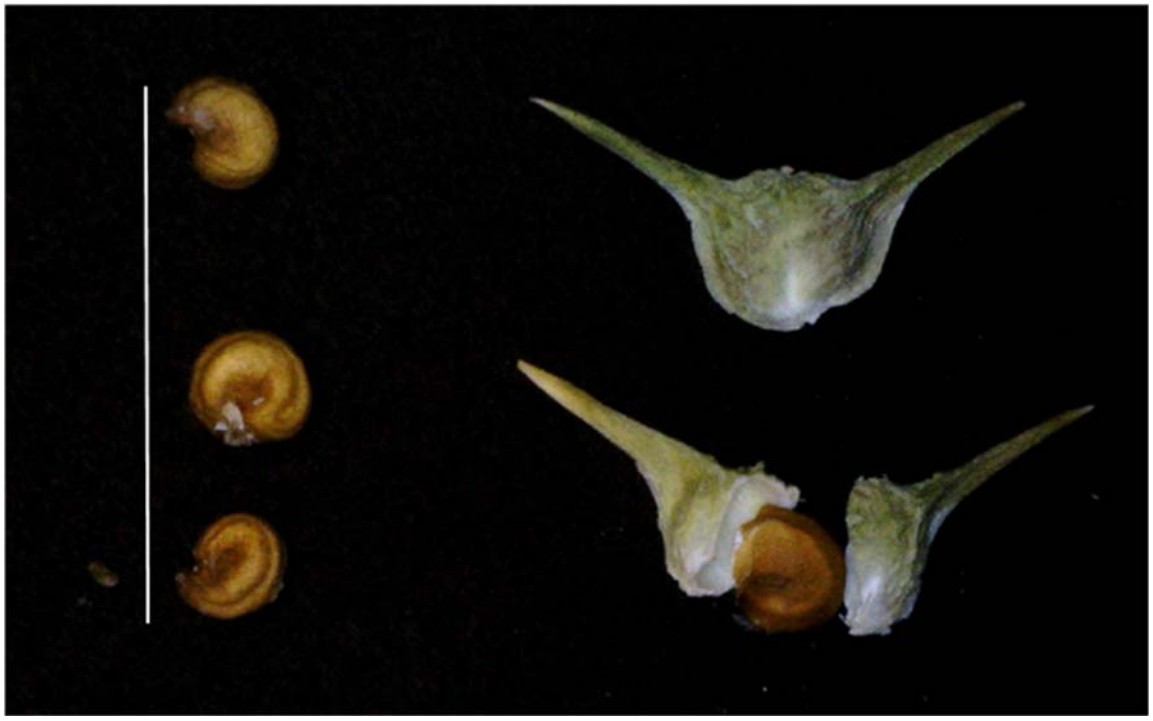

Figure 2.4 Comparison of the products of hermaphrodite and female flowers. Three brown seeds that developed on hermaphrodite flowers of XY individuals in accession Cornell #9 are laid in a 1 cm row. a. To the right are two fully formed green fruits which develop from female flowers. One of the utricles was manually broken open the reveal a brown seed, which is indistinguishable from the seeds the develop on hermaphrodite flowers. b.

Identification of X-specific sequence

Because androdioecious spinach segregate XX, XY and also YY individuals, YY genomic DNA template was available to aid the discovery of X-specific sequences in an XX reference genome. The female genome has two copies of the X chromosome, while the male genome has one. Thus X-specific sequence can be identified by comparing depth of coverage differences between male and female alignments from an XX draft genome (Dohm et al. 2014). Of the 104,270 scaffolds and contigs in the spinach draft genome 19 had significantly lower depth of coverage when aligned to by male than female sequences, indicating they are possibly X-specific sequence. However, most of those CNVs were presumed to be noise because 18 scaffolds in the reference had significantly lower depth of coverage when aligned to by female

than by male sequences. The roughly equivalent, small number of CNVs indicated this analysis method identifies false positives. To verify which candidate scaffolds represent X-specific sequence of the X chromosome, primers were designed for 10 of the 19 possible X-specific scaffolds with the most significant variation between male and female alignments.

All 10 primer pairs amplified a product of the expected band size in XX and XY templates, but only one failed to amplify product from YY template. The PCR marker, *Spinacia oleracea* X (SpoX), amplified a 198 bp product from genomic DNA extracted from XX and XY, but not YY. SpoX can be used in multiplex PCR with the previously published male-specific marker T11A. Together, these two pairs of primers are able to discriminate between XX, XY, and YY template (**Figure 2.5**). SpoX targets scaffold29742 in the spinach draft genome version 1.0.1 (Dohm et al. 2014). On this contig, a CNV was detected with a confidence of $p=1.55e^{-38}$ (Xie and Tammi 2009). Scaffold29742 is 17294 bp long, but variation in depth of coverage was only reported for a 9954 bp region in a window between 1423 and 12798. The *ab initio* gene predictor Augustus found no likely coding sequences on the scaffold (Stanke et al. 2006).

Penetrance of Androdioecy

While the androdioecious individuals in accession Cornell #9 were presumed to be XY, the single pure male found in accession Cornell #9 was found to be YY by both a controlled cross and the marker SpoX. To test if copy number of the Y chromosome causes that difference in sex expression, an inbred population was developed to observe the relationship between sex expression and chromosomal genotype.

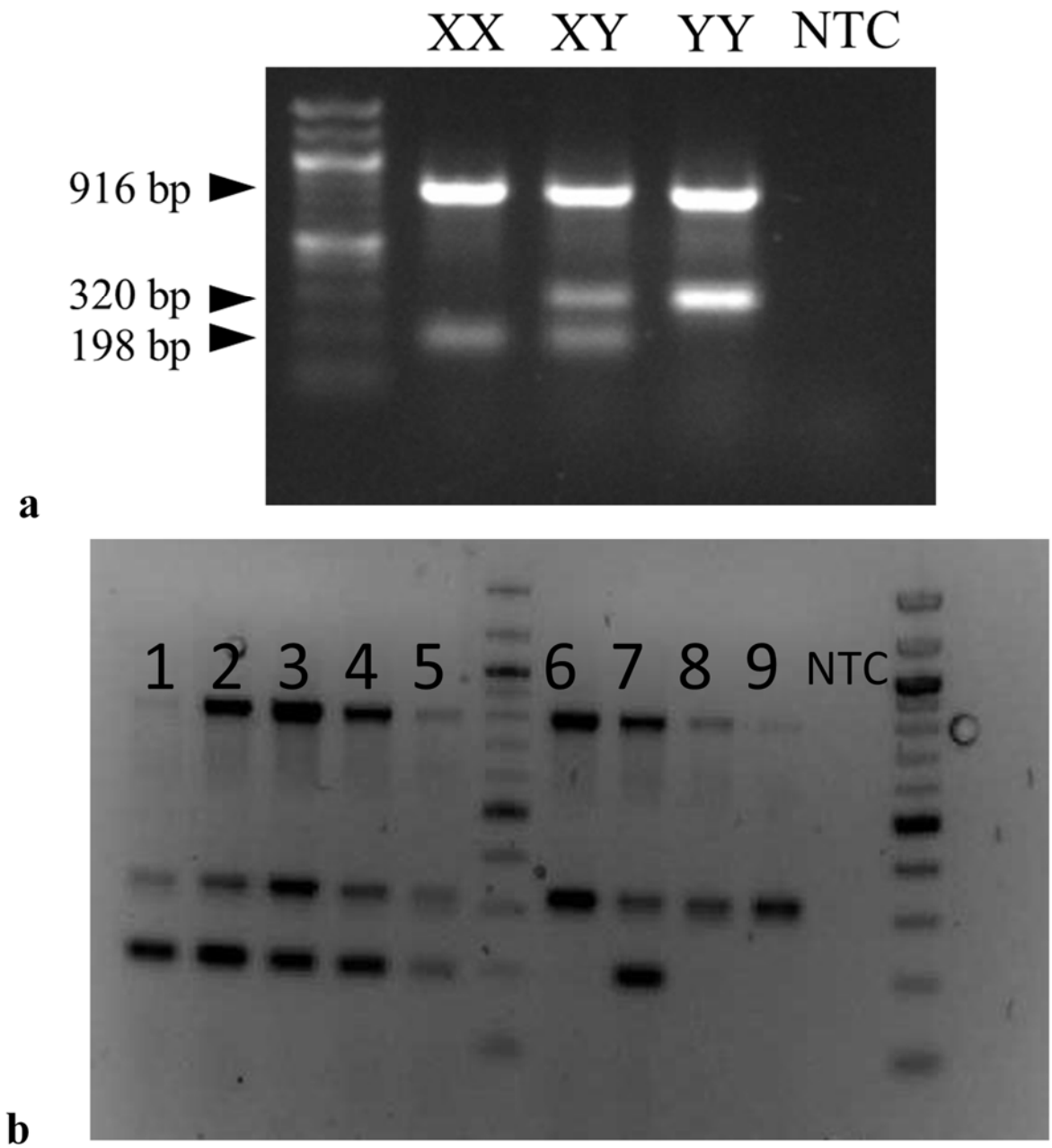

Figure 2.5. Sex-Specific amplification of SpoX and T11A. a. PCR results of SpoX and T11A amplifying from XX, XY, and YY genomic templates with a no template control (NTC). T11A amplifies a 916 bp band from an autosomal target from DNA of each sex type, but only the 320 bp male-specific band from XY and YY template. SpoX amplifies a 198 bp product from XX and XY templates but nothing from YY template. B. All staminate individuals in generation seven of inbreeding Cornell #9 were tested for presence of SpoX. Individuals 1-5 and 7 had hermaphroditic flowers, even though 7 had only a single hermaphroditic flower. Individuals 6, 8, and 9 were male. An NTC was included.

Seed from a single XY plant was used to generate an inbred line which segregates XX, XY, and YY individuals. Seeds for successive generations were collected from self-pollinated androdioecious parents. All generations segregated female, androdioecious, and male individuals. At generations 5 and 7, androdioecious and male individuals were tested with the sex-specific markers T11A (Y-specific) and SpoX (X-specific) to confirm genotype as XY or YY (**Table 2.2**). Females were presumed to be XX. Generation 5 comprised of 7 female, 15 androdioecious, and 8 male individuals. T11A amplified autosomal and Y-specific product from each of the 15 androdioecious and 8 male templates. SpoX amplified product from 15 androdioecious and 7 female templates but not from any of the 8 YY templates. At generation 7, 6 androdioecious and 3 pure male individuals were tested (**Figure 2.5**). T11A amplified both product from all 9. SpoX amplified product from only the 6 androdioecious individuals, including one with a single hermaphroditic flower.

Table 2.2 SpoX present on XY but not YY template DNA

Generation 5	Hermaphrodite	Male	Generation 7	Hermaphrodite	Male
SpoX present	15	0	SpoX present	6	0
SpoX absent	0	8	SpoX absent	0	3

Discussion

The field screen of the diverse USDA spinach germplasm collection led to the discovery of an accession of androdioecious spinach. This discovery was both informative to the genetic content of the spinach MSY and provides a resource to validate X-specific markers. It was found that XY individuals have hermaphroditic flowers, while YY individuals are pure male. These YY individuals were also useful for validating the X-specific marker SpoX, which identified the first contig of the X-specific region in spinach X chromosome.

Because dioecy has evolved independently in clades across the plant kingdom, the sex determination genes and mechanisms that control unisexuality may vary between dioecious species. The best studied plant sex chromosomes control dioecy with two sex determination genes on the MSY. However, the two-gene sex determination system (predicted by the gynodioecy pathway) may not be the only way that a plant species can maintain dioecy (Spigler and Ashman 2012). The monoecy-paradioecy pathway suggests genes influencing the ratio of unisexual male and female flowers in a monoecious species could mutate to generate pure male or pure female individuals (Lloyd and Webb 1977; Lloyd 1980). Were such a gene to mutate to create a dominant pure male allele (Y) and a recessive pure female allele (X) a species could develop a single gene sex determination system with the male as a heterogametic male parent. The recent discovery that the same gene controls sex determination in both dioecious and monoecious species of *Diospyros* (persimmon) led to the speculation that sex determination in persimmon is controlled by a single, "SRY-like" gene (Akagi et al. 2014). It is possible that persimmon and other species may regulate dioecy without the standard two-gene system.

Spinach has also been speculated to employ the same sex determination system to control both dioecy and monoecy. As in persimmon (Akagi et al. 2014), the male flowers in male and monoecious individuals resemble each other, as do the female flowers in female and monoecious individuals (**Figure 2.1, Figure 2.2**). Controlled crosses between monoecious and dioecious populations do not segregate hermaphrodites, indicating sex determination genes on both the M locus and MSY may have the same molecular targets (Janick and Stevenson 1955a; Onodera et al. 2011). Also, dioecious individuals of those species can be forced to reverse sex from male to female or vice versa from pruning or hormone treatment (Chailakhyan 1979; Chailakhyan and Khryanin 1979). These three properties led to the speculation that dioecy and monoecy may be

controlled by a single switch sex determination system that regulates canalized development of flowers to be male or female.

The discovery of sex reversed XY individuals with hermaphrodite flowers, instead of female flowers, supports the existence of a two gene sex determination system on the MSY (Charlesworth and Charlesworth 1978). As predicted by the gynodioecy pathways, one gene promotes stamen development and another gene represses carpel development. The haplotype of the Y chromosome found in Cornell #9 must have a partially dominant allele of the gene that controls carpel suppression. The intermediate phenotype of the XY and YY genotypes indicates the mutation of the carpel suppression gene is not a complete loss of function but reduces gene expression. This reduction is likely by more than the half of typical expression considering that artificially cultured polyploids of the XXY and XXXY genotypes are pure male (Mahoney et al. 1959).

A single gene sex determination system would not require the repressed recombination between the X and Y that was identified by the depth of coverage analysis which found X-specific sequence. Several male-specific markers were previously identified, indicating a repression of recombination on the MSY. Arguably, those markers could have identified variants tightly linked to the 'pure male allele'. Using depth of coverage analysis to identify a 9954 bp X-specific region is strong evidence that there is a physical divergence between X and Y-specific regions of the spinach X and Y sex chromosomes. Although the X and Y chromosomes are indistinguishable by size, the combined evidence of MSY marker data and the identification of an X-specific marker makes the existence of a non-recombining region between the spinach X and Y chromosomes difficult to refute.

Although there are no sex determination genes on the X-specific region of the X chromosome, the discovery of X-specific sequence is a major step toward characterizing and comparing the chromosomal architecture of the X and Y chromosomes in spinach. This X-specific sequence is the longest spinach sex chromosome sequence yet reported. The identification of 17294 bp of X-specific sequence shown here is the first step towards obtaining the complete sequence of X chromosome. Though the size of the non-recombining regions of the sex chromosomes are unknown, we expect them both to be larger than the 17294 bp scaffold identified by SpoX. A similar analysis done in date palm failed to find any of the sex-specific sequence of its primitive sex chromosomes (Younis et al. 2008; Al-Dous et al. 2011; Cherif et al. 2013). The success of CNV-seq here implies the spinach X and Y may be more divergent than the date palm sex chromosomes, but only so divergent that one scaffold was detected. Minimal but detectable divergence between X and Y sequences is expected for Stage 2 sex chromosomes.

Conclusions

Identification of X-specific sequence and YY individuals completes the criteria that spinach has Stage 2 sex chromosomes. Those sex chromosomes are expected to use a two-gene sex determination system to maintain dioecy. The discovery of an X-specific marker is the first step towards sequencing the X-specific region of the sex chromosome. This marker can also be used to identify YY individuals before flowering. Discovery of the SpoX marker and androdioecious spinach paved the way for studies that rely on YY genotype for sequencing material or to generate recombined Y chromosomes.

CHAPTER 3: IDENTIFICATION AND CHARACTERIZATION OF SEX DETERMINING REGION IN NASCENT SEX CHROMOSOMES OF *SPINACIA OLERACEA* L. (SPINACH)

Abstract

Spinach is a primarily dioecious species having male heterogametic XY chromosomes. The YY genotype (so called supermale) is viable, indicating the sex chromosomes are at a very early stage of sex chromosome evolution by having a small non-recombining sex determining region. *De novo* assemblies of the XX (female) and YY (supermale) genomes provided an unprecedented opportunity to identify X- and Y-specific sequences in the non-recombining region of the sex chromosomes. Using the X-specific marker SpoX, a 1.14 Mbp of X-specific contig was identified and validated. This contig was annotated and contained 31 genes, of which seven were found not to be repetitive. Each of the seven X genes were used to search for homologous sequences in the YY genome, and seven homologous Y contigs were identified, totaling 427 kb of novel Y chromosome sequence. Those Y contigs were composed of k-mers absent from the female genome. The percentage that Y contigs were uncovered by female k-mers decreased dramatically from 9.5% to 0.6% between positions 65.8 Mbp and 66.7 Mbp on the X chromosome, indicating that one of the boundaries in the non-recombining region of the sex chromosomes was reached. Further detailed analysis of the seven Y-specific contigs may reveal Y-specific genes and candidate genes for sex determination. The availability of the XX female and YY supermale genomes expedite sex chromosome research in spinach.

Introduction

Approximately 6% of angiosperms are dioecious (Renner 2014). By studying the sequence and structure of the sex chromosomes that control dioecy, light can be shed on how plant sex chromosomes have evolved from the autosomes of hermaphroditic ancestors (Charlesworth and Charlesworth 1978). Research of plant sex chromosomes has been guided by proposed evolutionary pathways that predict changes in chromosomal architecture and the mutations that give rise to sex chromosomes. The most accepted evolutionary pathway is the gynodioecy pathway (Spigler and Ashman 2012). This pathway describes the evolution of two sex determination genes and a chromosomal rearrangement that results in an XY breeding system. One sex determination gene promotes the growth of the stamens: this gene loses function on the X chromosome, thus making XX plants male sterile (female) while XY plants can still produce pollen. The other sex determination gene suppresses carpel development, making flowers on XY plants female sterile. The chromosomal rearrangement represses recombination between the SDR of the X and Y, resulting in two coupling phase sex determination genes in the male specific region of the Y chromosome (MSY) (Charlesworth and Charlesworth 1978).

After the initial repression of recombination, the size of the MSY increases because recombination is repressed, resulting in accumulation of retrotransposon insertions. The repression of recombination caused by the chromosomal rearrangement extends beyond the physical boundary of the rearrangement into the flanking psuedoautosomal regions (PAR). The non-recombining region (NRR) is subject to Muller's Ratchet: the tendency to accumulate deleterious mutations. Over evolutionary time, sequence divergence in the NRR becomes great enough to interfere with homologous pairing during meiosis, thus expanding the region where recombination is repressed. This results in divergent X and Y sequences near the physical

boundary of the NRR, and a gradual drop off of sequence divergence towards the genetic and then molecular boundaries of the NRR.

While the sequence divergence of homologous X and Y sequences is useful for pinpointing the positions of the SDR, it also presents a problem when sequencing sex chromosomes. Most contemporary sequencing and assembly strategies collapse minimally divergent X and Y sequences into chimeric contigs. Here we present novel sequences from the SDR of the spinach X and Y chromosomes, which were identified from two assemblies generated from separately sequenced XX (female) and YY (supermale) spinach DNA. From the female assembly, X-specific sequences were found flanking the previously identified X-specific marker, SpoX. From the supermale assembly, Y-specific sequences were identified both by homology to X chromosome genes and by percent that the homologous Y contigs have diverged from female sequences. These combined methods were able to identify 7 paired genes on the X and Y chromosomes. Analyzing changes in the percent that MSY contigs are uncovered by female k-mers indicated one of the PAR boundaries is within the 1.14 Mb area flanking the marker SpoX in the X chromosome.

Materials and Methods

Genomic resources

Both the female and supermale assemblies were generated by *de novo* PacBio sequencing. The female assembly was generated by the Spinach Genome Consortium, led by Allen Van Deynze at UC Davis. That assembly is 911 Mb of 989 Mb genome, and has been oriented into chromosome level pseudomolecules. Gene predictions were made using the MAKER pipeline (Campbell et al. 2014). The supermale assembly was generated for this

project. Genomic DNA from the pooled leaf tissue of 16 YY individuals was isolated using the SMRTbell DNA extraction protocol for *Arabidopsis*. Sequencing libraries were constructed from the DNA, which were sequenced at 63X depth. Sequence data was assembled using the program Canu. Final assembly was 823 Mb of the 989 Mb genome. The 22658 contigs have an N50 of 42 kb and an N90 of 19 kb.

Using Male-specific primer to identify MSY contigs

The sequence and reverse complement of 14 male-specific PCR markers found on the NCBI database were searched for exact matches into the supermale assembly. Sequences with exact matches to male-specific primers were queried by BLAST into databases of both the female and supermale genome (Altschul et al. 1990). Contigs with high BLAST results in the supermale, but not female assembly, were considered MSY sequence.

Identifying contigs with male-specific sequence

To find candidate MSY sequences, the program YGS.pl was used to identify contigs in the masked supermale assembly with with sequences absent in the female genome (Carvalho and Clark 2013) (**Figure 3.1**). The YGS method uses jellyfish to index all 15-mers in unassembled short reads of both female and male paired-end libraries. Those k-mers are then covered onto the supermale assembly. The program indicates the percent to which any contig is uncovered by female sequences, while still covered by male sequences. Contigs in the supermale assembly which are largely uncovered by female sequences are candidates for MSY sequence.

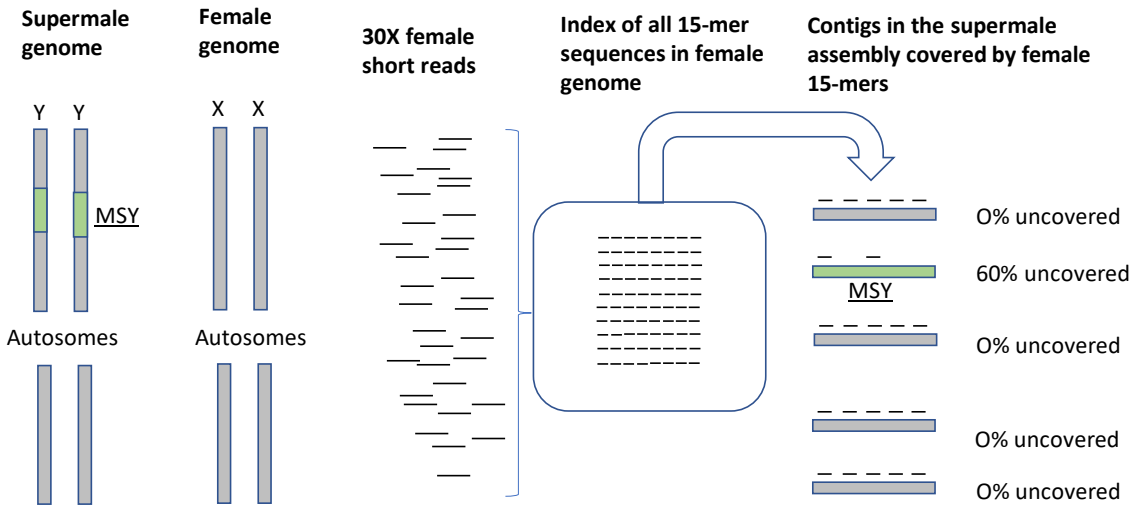

Figure 3.1 The YGS method applied to a supermale assembly. The YGS method identifies sex-specific contigs in an assembled male genome using unassembled female short reads. This works because the MSY (shown in green) is composed of sequences unique to the male genome when compared to the female genome. Here female short reads were used to identify contigs in the supermale assembly that are not covered by k-mers in the female genome. The program jellyfish made an index of all 15 base pair k-mers in the sequence of unassembled female short reads. Those female k-mers were covered onto the contigs of the supermale assembly. Most of the contigs in the supermale assembly are autosomal, which will be covered by the k-mers of the identical autosomal sequence of females, however; the contigs representing MSY sequence will be partially uncovered by female k-mers in portions where X and Y chromosome sequences have diverged. Reduction in coverage by female k-mers is one of the criteria used for determining a contig is MSY sequence.

Results

Locating X-specific sequence on the X chromosome

The contigs of the female reference genome were scaffolded into chromosome level pseudomolecules by our collaborators at UC Davis. The 188 bp sequence of the X-specific marker SpoX (Chapter 2) was searched in the female genome using BLAST (Altschul et al. 1990). The single BLAST hit was found in the female assembly. That hit was to the second longest pseudomolecule in the female assembly, which was designated as chromosome 1, the sex chromosome. The 188 bp sequence aligns to position 66019679-66019807 on chromosome 1. SpoX has no BLAST hits when queried into the supermale assembly.

Analysis of X sequence

Although the X chromosome is not predicted to have any genes involved in sex determination, analyzing the genes on the X chromosome corresponding to the MSY is necessary for comprehensive comparison of the sex chromosomes. X-specific genes were annotated on the contig of the female assembly targeted by the X-specific marker SpoX (Table 1). 31 genes from this contig were predicted by the MAKER pipeline. Of the 31 genes, 9 genes were annotated as functionally characterized genes. Of the 31, genes 6 and 7 were a tandemly duplicated pair of genes 8 and 9, and genes 21 and 22 were tandem duplicates. Eighteen genes were annotated as hypothetical protein, and among them, gene 30 and 31 were tandemly duplicated. Gene 26 was annotated as partial coding sequences of the Ty1-copia retrotransposon pol gene for reverse transcriptase. The remaining three genes had no annotation.

Table 3.1 List of genes annotated in the X-specific contig.

Gene on X	Hits in YY	Annotated gene with best alignment	NCBI Accession
1	13	PREDICTED: Beta vulgaris subsp. vulgaris beta-fructofuranosidase, insoluble isoenzyme CWINV3 (BIN46), mRNA	XM_010671083.2
2	8	hypothetical protein SOVF_201590 [Spinacia oleracea]	KNA04231.1
3	43	hypothetical protein SOVF_001400 [Spinacia oleracea]	KNA25970.-1
4	238	no annotation	
5	32	hypothetical protein SOVF_151650 isoform B [Spinacia oleracea]	KNA09649.1
6	24	*hypothetical protein SOVF_151650 isoform A [Spinacia oleracea]*	KNA09648.0
7	32	*hypothetical protein SOVF_151650 isoform B [Spinacia oleracea]*	KNA09649.1
8	32	*hypothetical protein SOVF_151650 isoform A [Spinacia oleracea]*	KNA09648.0
9	22	*hypothetical protein SOVF_151650 isoform B [Spinacia oleracea]*	KNA09649.1
10	10	PREDICTED: Beta vulgaris subsp. vulgaris DEAD-box ATP-dependent RNA helicase 8-like (LOC104892301), transcript variant X2, mRNA	XM_010678190.2

Table 3.1 (Cont.)

11	10	PREDICTED: monothiol glutaredoxin-S15, mitochondrial [Beta vulgaris subsp. vulgaris]	XP_010676561.1
12	500	PREDICTED: telomere length regulation protein TEL2 homolog [Beta vulgaris subsp. vulgaris]	XM_010697884.2
13	169	no annotation	
14	4	PREDICTED: Beta vulgaris subsp. vulgaris ABC transporter G family member 39-like (LOC104898459), mRNA	XM_010685545.2
15	4	*PREDICTED: Beta vulgaris subsp. vulgaris ABC transporter G family member 39-like (LOC104898459), mRNA*	XM_010685545.2
16	241	hypothetical protein SOVF_132170 [Spinacia oleracea]	KNA11746.1
17	189	*hypothetical protein SOVF_132170 [Spinacia oleracea]*	KNA11746.1
18	1	PREDICTED: phosphatidate phosphatase PAH2 isoform X1 [Beta vulgaris subsp. Vulgaris]	XP_010692790.1
19	199	hypothetical protein SOVF_130590 [Spinacia oleracea]	KNA11916.1
20	35	hypothetical protein SOVF_160170 [Spinacia oleracea]	KNA08713.1

Table 3.1 (Cont.)

21	500	hypothetical protein SOVF_010560 [Spinacia oleracea]	KNA25006.1
22	500	*hypothetical protein SOVF_010560 [Spinacia oleracea]*	KNA25006.1
23	404	hypothetical protein SOVF_108740 [Spinacia oleracea]	KNA14294.1
24	1	PREDICTED: B-cell receptor-associated protein 31 [Beta vulgaris subsp. vulgaris]	XP_010689493.1
25	240	hypothetical protein SOVF_089380 [Spinacia oleracea]	KNA16411.1
26	500	Spinacia oleracea Ty1-copia type retrotransposon pol gene for reverse transcriptase, partial cds	D12840.1
27	139	hypothetical protein SOVF_101190 [Spinacia oleracea]	KNA15118.1
28	207	PREDICTED: uncharacterized protein LOC104889316 [Beta vulgaris subsp. vulgaris]	XP_010672813.1
29	500	no annotation	
30	55	hypothetical protein SOVF_132170 [Spinacia oleracea]	KNA11746.1
31	53	*hypothetical protein SOVF_132170 [Spinacia oleracea]*	KNA11746.1

Identifying sequence in the supermale assembly absent in the female genome

Although many of the genes on the X and Y are homologous, the exact sequences of the MSY in the supermale assembly are absent from the female genome. To find male-specific sequences, contigs in the supermale assembly with incomplete coverage of k-mers from the female genome were found. The program YGS indexed all female k-mers of unassembled female short reads to generate a percentage to which each contig in the supermale assembly is partly uncovered by female sequences. Most contigs in the supermale assembly were completely covered by female k-mers; the median was 0.0% uncovered. However, of the 22575 contigs, 1081 are at least 5% uncovered by female sequences. 638 of contigs are at least 10% uncovered by female sequences and 308 contigs are at least 20% uncovered. The percent a contig is uncovered by female sequences is not sufficient to identify male-specific sequence, but was used in combination with other analysis techniques to identify MSY sequence.

One resource used to find male-specific sequences were published male-specific markers. 14 primers for male-specific markers found on the NCBI database were used to identify MSY contigs in the supermale assembly. This method identified two sequences absent in the female assembly. The sequence of primer E15150 exact matched to contig 00016089 and the sequence of primers E15149 and E15148 matched to contig00018122. Those sequences had significant BLAST hits when queried into the supermale genome, but not in the female genome. Those contigs had uncovered values of 28.6% and 33.6%, respectively. Of the 22574 contigs ranked by percent uncovered by female k-mers, those two contigs have the 104^{th} and 98^{th} highest percent uncovered by female sequences.

The seven X-specific genes that were not repetitive sequence were searched by BLAST into the supermale assembly to identify Y chromosome homologs (**Table 3.2**). Of the seven

genes, five had multiple hits (4-13 hits) and two had single hits. The X-specific PAH2 gene and the B-receptor only hit to their Y homologs. For the genes with multiple hits, percentage that the contigs were uncovered by female sequences was used to determine which of the multiple hits was the Y homolog, and the two single hit genes were tested for percentage of male-specific sequence as well. Of the seven genes, six had only one hit with more than 1.0% male-specific sequence, and the other one, Cell Wall Invertase (CWIN) that is duplicated on the X twice, had two hits over 1.0% (Table 3.2). Those two hits corresponded to both Y homologs of CWIN.

The percent that homologous Y contigs were uncovered by female sequences decreases in correlation to the position of genes on the X chromosome. This relationship was plotted using physical position of the X chromosome as the X axis and the percentage of the male-specific sequence as the Y axis. The highest percentage of male-specific sequence was CWIN, close to 10%, and gradually dropped to <1% for PAH2 and B-receptor (**Figure 3.2**). It appears that one border of the NRR was reached.

Table 3.2 Paired Genes in order of position on the X.

Gene	Position on X	Hits in YY	MSY Contig	Uncovered by Female K-mers	Size bp
Hypothetical Spinach Protein	65809609-65809803 65815745-65815939	8	tig00015263	9.50%	94195
CWIN	65810291-65810693 65816429-65816831	13	tig00015263 tig00018939	9.50% 6.70%	94195 19439
Dead Box ATP	65936873-65936679	10	tig00023014	7.20%	16083
SpoX	66019679-66019807	0			
Monothiol glutaredoxin-S15 Mitochondrial	66050602-66051204	10	tig00013251	6.40%	37265
ABC Transporter	66282718-66281815	4	tig00009352	2.20%	28394
PAH2	66455519-66453569	1	tig00032412	0.70%	87005
B-receptor	66740417-66740417	1	tig00030702	0.60%	51144

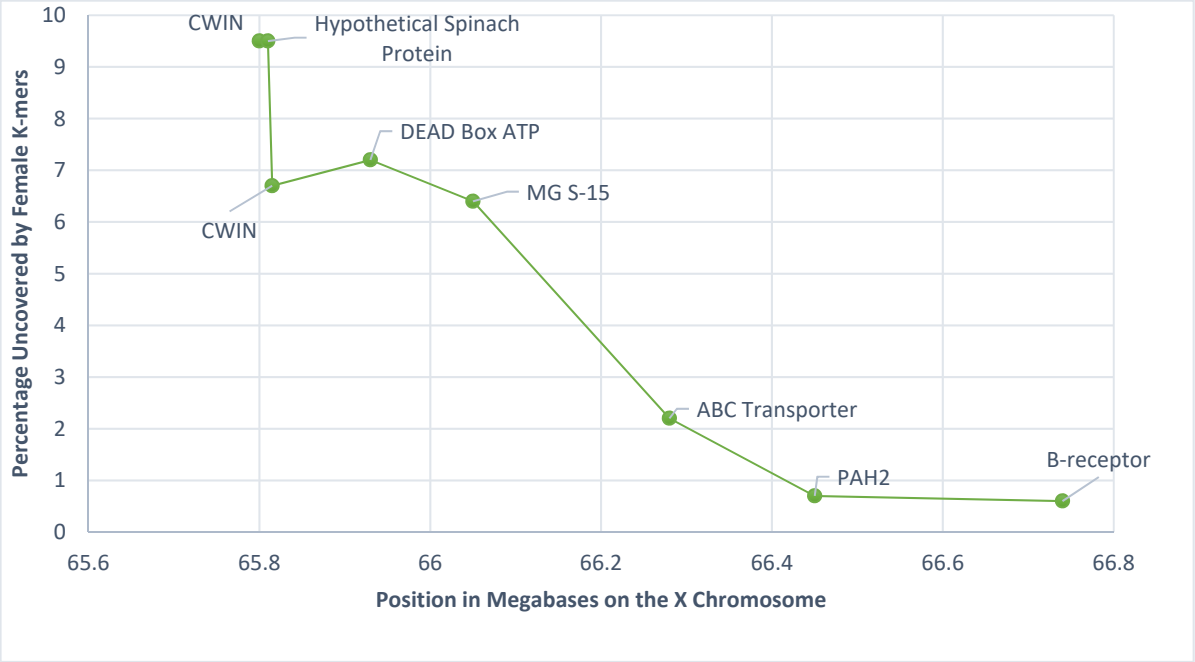

Figure 3.2 Percentage Y contigs are uncovered by female sequences compared to position of homologs on the X. Each of the MSY contigs with sequence homologous to genes on the X are graphed according to the percentage that they are uncovered by female sequences. Each contig is labeled by the gene on homologous X sequence.

Discussion

Each of the seven genes that are not repetitive in the X-specific contig matched to a Y homolog. Abundant homologous X and Y paired genes are expected for Stage 2 sex chromosomes with YY genotype viable because the X and Y initially have the same genetic content as the ancestral autosome. However, after recombination is repressed in the SDR, Muller's Ratchet causes sequences on the MSY to degenerate because deleterious mutations are not purged by recombination (Engelstädter 2008). As sequences from the Y diverge from the X, the region of repressed recombination increases. The gradual expansion of one of the NRR boundaries is shown by the decrease of sequence identity between the paired X and Y sequences **(Figure 3.2)**.

The contigs with Y copies of the Hypothetical spinach protein, CWIN, DEAD box, and MG-S15 genes correspond to the genes on the X chromosome between positions 65.8 and 66.0

Mbp. Those MSY contigs are 6.2-9.5% uncovered by female sequences in descending order. Successively, ABC transporter, PAH2, and B-receptor genes on the X pair with MSY contigs with 2.2%, .7%, and .6% uncovered by female sequences. The trend of increasing coverage of MSY contigs by female sequences reflects the gradual expansion of the genetic border of the NRR into the pseudo-autosomal region (PAR). The X-specific marker SpoX is at a position near 66.0 Mbp in the non-recombining region, which is consistent with previous results indicating SpoX targets the X-specific region of the X chromosome (Chapter 2).

The *de novo* assembly of the spinach transcriptome found 72,151 unigenes in the 980 Mbp spinach genome, meaning there should be on average one gene every 13.5 kb (Xu et al. 2015). Thus, a region the size of the 1.14 Mb X-specific contig should have at least 84 genes on average. The X-specific contig contains 31 gene predictions and only seven genes sequences which appeared to be functional genes when manually annotated. The gene paucity seen here is a hallmark of a pericentric non-recombining region in sex chromosome, as shown in papaya (Liu et al 2004; Wang et al. 2012; Gschwend et al. 2012). Florescent *in situ* hybridization (FISH) of probes designed for male-specific sequences mapped to or near the centromeres in spinach (Chuanliang Deng, personal communication), corroborating the observation of gene paucity in the NRR. Also, the recombination rate around the SDR between the two X chromosomes in female spinach was found to be reduced, as would be expected near a centromere (Yamamoto et al. 2014). The NRR in papaya and *Vasconcellea parviflora* are also both near the centromere of the sex chromosome (Gschwend et al. 2012; Iovene et al. 2015).

Conclusion

By comparing separately sequenced female and supermale assemblies, 1.14 Mbp of X-specific sequence and 7 homologous MSY sequences, were discovered. The decrease in percent of male-specific k-mer content of the MSY sequences indicates the approximate position of one of the NRR boundaries, which includes the marker SpoX in the X-specific region of the X. The low gene density of the X sequences supports the hypothesis that the spinach SDR is near the centromere of the sex chromosome.

REFERENCES

Akagi T, Henry IM, Tao R, Comai L (2014) A Y-chromosome–encoded small RNA acts as a sex determinant in persimmons. Science (80-) 346:646–650.

Akamatsu T, Suzuki T (1999) Method for identifying the sex of spinach by DNA markers.

Al-Dous EK, George B, Al-Mahmoud ME, et al (2011) De novo genome sequencing and comparative genomics of date palm (Phoenix dactylifera). Nat Biotechnol 29:521–527. doi: 10.1038/nbt.1860

Al-Khayri JM, Huang FH, Morelock TE, Busharar TA (1992) In vitro seed production from sex-modified male spinach plants regenerated from callus cultures. Sci Hortic (Amsterdam) 52:277–282. doi: 10.1016/0304-4238(92)90029-C

Altschul SF, Gish W, Miller W, et al (1990) Altschul et al.. 1990. Basic Local Alignment Search Tool.pdf. J. Mol. Biol. 215:403–410.

Campbell MS, Holt C, Moore B, Yandell M (2014) Genome Annotation and Curation Using MAKER and MAKER-P. Curr Protoc Bioinforma 2014:4.11.1-4.11.39. doi: 10.1002/0471250953.bi0411s48

Carvalho A, Clark AG (2013) Efficient identification of y chromosome sequences in the human and drosophila genomes. Genome Res 23:1894–1907. doi: 10.1101/gr.156034.113

Chailakhyan M (1979) Genetic and hormonal regulation of growth, flowering, and sex expression in plants. Am J Bot 66:717–736.

Chailakhyan, Khryanin V (1979) The role of leaves in sex expression in hemp and spinach. Planta 144:205–207. doi: 10.1007/BF00387272

Charlesworth B, Charlesworth D (1978) A Model for the Evolution of Dioecy and Gynodioecy. Am. Nat. 112:975.

Cherif E, Zehdi S, Castillo K, et al (2013) Male-specific DNA markers provide genetic evidence of an XY chromosome system, a recombination arrest and allow the tracing of paternal lineages in date palm. New Phytol 197:409–415. doi: 10.1111/nph.12069

Darwin C (1877) The different forms of flowers on plants of the same species. John Murray

Deng C, Qin R, Cao Y, et al (2013) Microdissection and painting of the Y chromosome in spinach (Spinacia oleracea). J Plant Res 126:549–556. doi: 10.1007/s10265-013-0549-3

Deng C, Qin R, Gao J, et al (2012) Identification of sex chromosome of spinach by physical mapping of 45s rDNAs by FISH. Caryologia 65:322–327.

Dohm JC, Minoche AE, Holtgräwe D, et al (2014) The genome of the recently domesticated crop plant sugar beet (Beta vulgaris). Nature 505:546–9. doi: 10.1038/nature12817

Ellis JR, Janick J (1960) The Chromosomes of Spinacia oleracea. Am. J. Bot. 47:210.

Engelstädter J (2008) Muller's ratchet and the degeneration of Y chromosomes: A simulation study. Genetics 180:957–967. doi: 10.1534/genetics.108.092379

Freeman DC, Harper KT, Charnov EL (1980) Sex Change in Plants: Old and New Observations and New Hypotheses. Oecologia 232:222–232. doi: 10.1007/bf00346825

Geber M, Dawson TE, Delph L (2012) Gender and sexual dimorphism in flowering plants. Springer Science & Business Media

Gschwend AR, Weingartner LA, Moore RC, Ming R (2012) The sex-specific region of sex chromosomes in animals and plants. Chromosom Res 20:57–69.

Iizuka M, Janick J (1962) Cytogenetic analysis of sex determination in Spinacia oleracea. Genetics 47:1225.

Iizuka M, Janick J (1963) Sex chromosome translocations in Spinacia oleracea. Genetics 48:273.

Iovene M, Yu Q, Ming R, Jiang J (2015) Evidence for emergence of sex-determining gene(s) in

a centromeric region in vasconcellea parviflora. Genetics 199:413–421. doi: 10.1534/genetics.114.173021

Ito M (2000) Characterization of Spinach Chromosomes by Condensation Patterns and Physical Mapping of 5 s and 45s rDNAs by FISH. Sci Technol 125:59–62.

Janick J, Mahoney DL, Pfahler PL (1959) The trisomics of Spinacia oleracea. J Hered 50:47–50.

Janick J, Stevenson EC (1955a) Genetics of the monoecious character in spinach. Genetics 40:429.

Janick J, Stevenson EC (1955b) The effects of polyploidy on sex expression in spinach. J Hered 46:151–156.

Khattak JZK, Torp AM, Andersen SB (2006) A genetic linkage map of Spinacia oleracea and localization of a sex determination locus. Euphytica 148:311–318. doi: 10.1007/s10681-005-9031-1

Komai F, Shikazono N, Tanaka A (2003) Sexual modification of female spinach seeds (Spinacia oleracea L.) by irradiation with ion particles. Plant Cell Rep 21:713–717. doi: 10.1007/s00299-003-0592-y

Lan T, Zhang S, Liu B, et al (2006) Differentiating sex chromosomes of the dioecious Spinacia oleracea L. (spinach) by FISH of 45S rDNA. Cytogenet Genome Res 114:175–177. doi: 10.1159/000093335

Li H, Durbin R (2009) Fast and accurate short read alignment with Burrows-Wheeler transform. Bioinformatics 25:1754–1760. doi: 10.1093/bioinformatics/btp324

Lloyd DG (1980) The distributions of gender in four angiosperm species illustrating two evolutionary pathways to dioecy. Evolution (N Y) 123–134.

Lloyd DG, Webb CJ (1977) Secondary sex characters in plants. Bot Rev 43:177–216. doi:

10.1007/BF02860717

Mahoney DL, Janick J, Stevenson EC (1959) Sex determination in diploid-triploid crosses of Spinacia oleracea. Am J Bot 372–375.

Ming R, Bendahmane A, Renner SS (2011) Sex chromosomes in land plants. Annu Rev Plant Biol 62:485–514. doi: 10.1146/annurev-arplant-042110-103914

Ming R, Moore PH (2007) Genomics of sex chromosomes. Curr. Opin. Plant Biol. 10:123–130.

Ming R, Wang JP, Moore PH, Paterson AH (2007) Sex chromosomes in flowering plants. Am J Bot 94:141–150.

Onodera Y, Yonaha I, Masumo H, et al (2011) Mapping of the genes for dioecism and monoecism in Spinacia oleracea L.: Evidence that both genes are closely linked. Plant Cell Rep 30:965–971. doi: 10.1007/s00299-010-0998-2

Onodera Y, Yonaha I, Niikura S, et al (2008) Monoecy and gynomonoecy in Spinacia oleracea L.: Morphological and genetic analyses. Sci Hortic (Amsterdam) 118:266–269. doi: 10.1016/j.scienta.2008.06.008

Pfent C, Pobursky KJ, Sather DN, Golenberg EM (2005) Characterization of SpAPETALA3 and SpPISTILLATA, B class floral identity genes in Spinacia oleracea, and their relationship to sexual dimorphism. Dev Genes Evol 215:132–142. doi: 10.1007/s00427-004-0459-4

Ramanna MS (1976) Are there heteromorphic sex chromosomes in spinach (Spinacia oleracea L.)? Euphytica 25:277–284.

Renner SS (2014) The relative and absolute frequencies of angiosperm sexual systems: Dioecy, monoecy, gynodioecy, and an updated online database. Am J Bot 101:1588–1596.

Sather DN, Jovanovic M, Golenberg EM (2010) Functional analysis of B and C class floral organ genes in spinach demonstrates their role in sexual dimorphism. BMC Plant Biol

10:46.

Sather DN, York A, Pobursky KJ, Golenberg EM (2005) Sequence evolution and sex-specific expression patterns of the C class floral identity gene, SpAGAMOUS, in dioecious Spinacia oleracea L. Planta 222:284–292.

Sherry RA, Eckard KJ, Lord EM (1993) Flower development in dioecious Spinacia oleracea (Chenopodiaceae). Am J Bot 283–291.

Slancarova V, Zdanska J, Janousek B, et al (2013) Evolution of sex determination systems with heterogametic males and females in silene. Evolution (N Y) 67:3669–3677. doi: 10.1111/evo.12223

Spigler RB, Ashman TL (2012) Gynodioecy to dioecy: Are we there yet? Ann. Bot. 109:531–543.

Stanke M, Keller O, Gunduz I, et al (2006) AUGUSTUS: ab initio prediction of alternative transcripts. Nucleic Acids Res 34:W435–W439.

Stevenson EC (1954) A genetic study of the heterogametic nature of the staminate plant in spinach (Spinacia oleracea L.). In: Proc. Amer. Soc. Hort. Sci. pp 444–446.

Vanburen R, Zeng F, Chen C, et al (2015) Origin and domestication of papaya Y h chromosome. Genome Res 524–533. doi: 10.1101/gr.183905.114.9

Wang J, Na J-K, Yu Q, et al (2012) Sequencing papaya X and Yh chromosomes reveals molecular basis of incipient sex chromosome evolution. Proc. Natl. Acad. Sci. 109:13710–13715.

Xie C, Tammi MT (2009) CNV-seq, a new method to detect copy number variation using high-throughput sequencing. BMC Bioinformatics 10:80. doi: 10.1186/1471-2105-10-80

Xu C, Jiao C, Zheng Y, et al (2015) De novo and comparative transcriptome analysis of

cultivated and wild spinach. Sci Rep 5:17706. doi: 10.1038/srep17706

Yamamoto K, Oda Y, Haseda A, et al (2014) Molecular evidence that the genes for dioecism and monoecism in Spinacia oleracea L. are located at different loci in a chromosomal region. Heredity (Edinb) 112:317–324. doi: 10.1038/hdy.2013.112

Younis RAA, Ismail OM, Soliman SS (2008) Identification of sex-specific DNA markers for date palm (Phoenix dactylifera L.) using RAPD and ISSR techniques. Res J Agric Biol Sci 4:278–284.

Yu Q, Hou S, Hobza R, et al (2007) Chromosomal location and gene paucity of the male specific region on papaya Y chromosome. Mol Genet Genomics 278:177–185. doi: 10.1007/s00438-007-0243-z

Zhang J, Boualem A, Bendahmane A, Ming R (2014) Genomics of sex determination. Curr. Opin. Plant Biol. 18:110–116.

APPENDIX A: USDA ACCESSIONS WITH ABNORMAL SEXUAL PHENOTYPE

Table A.1: Accessions where non-dioecious breeding systems were most noted.

Field number	USDA Inventory number	Accession Name	Origin	Breeding system
112	PI 173972	Palak	India	staminate female
120	PI 174960	Palak	India	10% staminate female
129	PI 175927	Cornell ID #55	Turkey, Yozgat	50% staminate female
198	PI 209645	No. 2	Iran, Fars	10% staminate female palak
206	PI 217425	Cornell ID #9	South Korea	weedy, pistillate male
210	PI 220686	Palek	Afghanistan	50% staminate female

APPENDIX B: USDA ACCESSIONS WITH EXAGGERATED SEXUAL DIMORPHISM

Table B.1: Accessions where males and females had the most conspicuous morphological differences

Field number	USDA Inventory number	Accession Name	Origin	Trait of interested
5	Ames 20168	Cornell #272	China	long peduncle
107	PI 173128	Cornell ID #84	Turkey, Maras	long peduncles
108	PI 173129	Cornell ID #84	Turkey, Maras	long peduncles
115	PI 174385	Cornell ID # 92	Turkey, Diyarbakir	long peduncle, weedy male
134	PI 175932	Harlan 9148	Turkey, Kayseri	long peduncle
172	PI 181809	Cornell ID #46	Syria	long peduncle
180	PI 200882	Cornell ID #50	Afghanistan	short peduncle
196	PI 207518	Cornell ID #30	Afghanistan	extremely dimorphic secondary sexual traits
200	PI 209647	No. 4	Iran, Fars	exaggerated male features
299	PI 419218	Cornell #164	Hong Kong	early flowering, long peduncle
307	PI 445783	Shami	Syria	longest peduncle
314	PI 499372	Ispolonskijj	Soviet Union	extremely sexually dimorphic

APPENDIX C: USDA ACCESSIONS FIRST TO FLOWER

Table C.1: Accessions where multiple plants bolted extremely early

Field number	USDA Inventory number	Accession Name	Origin	Relative Flowering Time
78	PI 169676	Dikensiz	Turkey	earliest flowering
222	PI 229792	Espinage	Iran	early flowering
257	PI 296393	Cornell ID #180	Iran	early flowering, weedy palak
262	PI 339546	101-5	Egypt	early flowering
334	PI 604783	SPI 13/79	Afganistan	early senescence
358	PI 648942	119A0245	China, Beijing	early flowering

APPENDIX D: USDA ACCESSIONS LATEST TO FLOWER

Table D.1: Accessions which bolted extremely late

Field number	USDA Inventory number	Accession Name	Origin	Relative Flowering time
28	NSL 6693	Old Dominion	Michigan	latest flowering
29	NSL 6782	Hollandia	Netherlands	latest flowering/ savoy leaves
56	PI 164965	Cornell ID #242	Turkey	latest flowering
219	PI 227230	Jiromaru	Japan	latest flowering
241	PI 274050	Hiemalis	Germany	latest flowering
337	PI 604786	SPI 62/78	Nepal	latest flowering
370	PI 648954	102x99	Maryland	latest flowering
373	PI 648955	76x71	Maryland	latest flowering

APPENDIX E: USDA ACCESSIONS EXHIBITED STRESS

Table E.1: Accessions which were yellowed and stressed

Field number	USDA Inventory number	Accession Name	Origin	Stress Phenotype
250	PI 274061	Viking	UK, England	stressed bolting
267	PI 358248	Ohdriski	Serbia	stressed, intermediate flowering
290	PI 379547	Veleski	Serbia	stressed, early flowering
323	PI 531453	Mohai	Hungary	stressed, miniature
367	PI 648951	257x251	Maryland	stressed
368	PI 648952	134x129	Maryland	stressed

APPENDIX F: USDA ACCESSIONS WITH MISCELLANEOUS NOTABLE TRAITS

Table F.1: Other notable accessions that do not resembled 'typical' spinach

Field number	USDA Inventory number	Accession Name	Origin	Distinction
30	NSL 22003	Badger Savoy	California	large plant
132	PI 175930	Cornell ID #61	Turkey, Kirschir	highly lobed leaves
166	PI 179594	445	Belguim	large diameter of rosettes, squat plant
199	PI 209646	No. 3	Turkey	green stem, palak like
203	PI 212120	Cornell ID #6	Afganistan	early seed development
209	PI 220121	Palek	Afganistan	very small in stature
226	PI 256079	Cornell #209	Afganistan	much lateral growth
333	PI 604782	SPI 12/79	Afganistan, Jowstan	not suppressing bracts
389	PI 604792	SPI 155/83	Germany	highly lobed leaves