GENOMIC DISSECTION OF STRUCTURAL VARIATIONS AND EVOLUTION FORCES OF THE SEX CHROMOSOMES IN SPINACH (*SPINACIA OLERACEA*)

BY

LI'ANG YU

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, 2020

Urbana, Illinois

Doctoral Committee:

Professor Ray Ming, Chair
Professor Gustavo Caetano-Anollés
Assistant Professor Tiffany Jamann
Assistant Professor Amy Marshall-Colon

Abstract

Sex chromosomes evolved independently and repeatedly animals and plants. Although considerable differences existed among diverse lineages in terms of sexual dimorphism and genetic basis of sex determination, the underlining principles and evolutionary forces that lead to the formation and progression of sex chromosomes are the same in plants and animals. The well-established theories and observations indicate that sex chromosomes in birds and mammals are ancient, but sex chromosomes in fish, amphibians, and plants are more recent. Studying those early-stage sex chromosomes could shed light on the inception and forces that shape sex chromosome evolution. We selected garden spinach, an annual leafy dioecious vegetable with a pair of nascent sex chromosomes to study sex chromosome evolution.

Based on a theory of sex chromosome evolution, sex chromosomes evolved from autosomes by genetic mutations relate to the female and male reproductive organ development. Suppression of recombination at the sex determination locus or loci is the pivotal point of sex chromosome evolution. As consequences of ceased recombination at the sex-determining region, accumulation of repetitive sequences, chromosomal rearrangements, and gene mutations occurred, accompanied with a strong selection due to a small effective population size of genes on the Y chromosome compared with autosomes. Genomic localization of the non-recombining region is critical for studying spinach sex chromosome evolution.

To identify the region associated with sex determination in spinach, I developed DNA markers to test 40 spinach accession from different countries. I found that a small 1.78 Mb region is featured by strong linkage with sex type and it is conserved across 40 different spinach accessions by genotyping plants with 12 DNA markers. However, a large-scale test by genome-wide polymorphic markers is needed to map the position of the region and estimate the size and numbers of genes within this sex co-segregation genomic region.

We re-sequenced the genomes of progenies from a segregating pseudo-test cross mapping population and used genome-wide DNA markers to construct two genetic maps, one for each parent, to identify this region. The two high-density linkage maps anchored a total of 40,324 SNPs for female map and 256,636 SNPs for male map with highly consensus orders with each other. The sex determination region on the Y chromosome was mapped to the largest linkage group at 45.18cM locus by male map, comprised 4,567 sex co-segregating SNPs. The two maps substantially improve resolution of the sex determination region, which also enabled a higher quality female genome assembly by anchoring contigs to a published reference genome. However, we need to generate sequences for both the complete genomic region of the Y chromosome and its X corresponded part to further understand the evolutionary forces related to sex chromosome evolution.

The viable YY spinach overcomes the difficulties of assembling heterozygous XY haplotypes. *De-novo* assembly facilitated by Hi-C and linkage maps of two new genomes (XX and YY) enabled us to define the male-specific region (MSY) and its X counterpart. Surprisingly, the comparison revealed a 39.26 Mb MSY and 38.92 Mb X counterpart, features by two inversions within the region along with gene loss and structural variations between paired genes from two sex types. A 6 Mb small region with a continuous pattern of gene divergence between X and Y chromosome further divide MSY into three evolutionary strata dated from 0.1-0.2 million years ago.

Together, these results expanded our knowledge on sex chromosome evolution in dioecious spinach. The knowledge and genomic resources generated can be applied to identify sex-determining genes to decipher sex determination and differentiation gene networks and improve efficiency and shorten cycles of spinach breeding.

Acknowledgements

The four-year doctoral training is just the beginning of my scientific research career while the list of people whom I would like to appreciate continuously grow. My doctoral research projects and dissertation would not be finished without their kindness, advice, and support.

Initially, I am grateful to my Ph.D. advisor Dr. Ray Ming for providing me the chance to study and work in plant biology department with other excellent researchers. The diligence, patience, and optimism from him largely inspired me to explore scientific questions audaciously and steadily. Dr. Ray Ming also gave me the freedom to pursue my research interests and encouraged me to involve other research projects from collaborators besides my Ph.D. dissertation. These experiences largely enhanced my research techniques, critical thinking, and communication skills. Then, I would also like to thank everyone in Ming lab for their help in the past four years and give special thanks to Dr. Fan Zhu, Dessiree Zerpa, Dr. Julie Ngyuen, and Xiaodan Zhang for sharing ideas and trouble-shootings for experiments. In addition, I appreciate the valuable suggestions and comments from my committee members: Dr. Gustavo Caetano-Anollés, Dr. Tiffany Jamann, and Dr. Amy Marshall-Colon regarding graduate coursework, my preliminary exam, and my Ph.D. defense.

I am also grateful to visiting scholars in Ming lab for their advice and support. Special thanks to Dr. Liangfa Ge and Dr. Youjin Deng for guiding me with many research techniques and my scientific writing skills, Dr. Gaihua Qin and Dr. Chuanliang Deng for inviting me to participate with their research projects. I would also like to thank Dr. William Wadlington for providing genetic sources and plant materials for my Ph.D. research projects, Dr. Xiaokai Ma from Fujian agricultural and forestry university, Dr. Allen Van Dyenze from the University of California-Davis, and Dr. Amanda Hulse-Kemp from USDA for collaborating spinach genome projects and technical support for data analysis. Special thanks to my friends from UIUC, Haoran Yu, Zhonghua Zheng, Yixiang Wang, Wei Wei, Hongxu Dong, Menghao Yu, and Jiayang Xie, for their encouragement for me to get through difficulties.

Finally, I appreciate the long-term encouragement and support from my parents and family members. I could not have defense my degree without you.

Table of Contents

Chapter I. Literature review

Abstract

Sex determination is of great interest in scientific research because of its pivotal role in sexual reproduction. As one of the most important research topics relates to sex determination, sex chromosomes are identified from a wide range of lineages in animals and plants with independent origin and trajectory of evolution. Previous studies revealed molecular mechanisms and genetic principles of sex chromosomes in different species, particularly in animals, and these discoveries provided informative references for studies in other different systems. As a common leafy dioecious plant species, limited knowledge existed regarding the genetic basis of dioecy in spinach. Also, biological mechanisms of spinach sex determination remain unclear due to complex spinach genome and limited genetic resources. Here, I reviewed the fundamental principles of sex determination and sex chromosome evolution, summarized the research progress of sex determination from higher plants, and discussed the feasibility as well as future perspectives of studying sex chromosome evolution in spinach.

Introduction

Sex determination in higher plants

Sex determination is one of the most fascinating topics for evolutionary biologists and geneticists. Compared with wide-spread sexual dimorphisms in animals, only 10% of the land plants are dioecious with separate male and female individuals while most species are hermaphrodites. Particularly, an estimation of 5% angiosperms are monoecious and 6% are dioecious. The sex-determination system is controlled by a verity of mechanisms includes chromosomal-sex determination, genic-sex determination, and sex determination by environmental factors (Zhang et al., 2014). Recent progress in plant genomics and biotechnology enabled the discoveries of candidate sex-determining genes in different families with no shared among families, reflecting diverse sex determination systems in plants. In general, sex is determined by sex-determining genes derived from genetic mutations related to stamen abortion or suppression of carpel development, which is frequently found among 75% of dioecious flowering plant families along with numerous sterility mutants (Renner & Ricklefs, 1995).

For sex determination in monoecious species, there was no direct evidence to justify the existence of sex chromosomes because of the lack of sexual dimorphism among individuals. In this system, one or more loci genetically control the sex determination and these genes could either linked or unlinked. In maize, the genic mutations and understanding of sex determination are well established by cloning of the genes and mutant gene-knockout experiments (Acosta et al., 2009; Chuck et al., 2007; Wu et al., 2007). There are six mutants that affect tassel (male flower) development and there are quite a few mutants that involved in gibberellic acid (GA) biosynthesis and perception and these mutants affect the ear (female flower) development. Most of these mutants either promote or suppress the pathways related to hormones. The sex determination of melon (*Cucumis melo* L.) is controlled by two unlinked genes *a* and *g* for andromonoecious and gynoecious respectively. The dominance and interactions of two genes result in other sexual phenotypes besides the monoecy (*A*-*G*-), includes andromonoecious (*aaG*-), gynoecious (*AAgg*) and hermaphrodite (*aagg*). The two genes are closely related to ethylene biosynthesis pathways that contributed to carpel or stamen development. (Boualem et al., 2008; Boualem et al., 2009; Poole & GRIMBALL, 1939). In another monoecious system cucumber (*Cucumis sativus* L.), the sexual polymorphisms are related to three loci, a partial dominant *F* locus controls the degree of femaleness, an *a* locus (recessive) contributed the increased degree of maleness and the other dominant *M* locus which encodes for ethylene perception and selectively controls abortion of the stamen. The interaction and combination of three genes contributed to the co-existence of hermaphrodite, monoecious, gynoecious cucumbers (Li et al., 2012; Perl-Treves, 1999; Trebitsh et al., 1997). These studies revealed different gene networks involved in flower development which finally led to the abortion of either female or male organ for sex determination.

In dioecious species, sex determination is often controlled by a pair of sex chromosomes, including XX-XO, XX-XY, and ZZ-ZW sex chromosome systems. Among 48 species from different genera, 44 dioecious species are controlled by sex chromosomes in male heterogametic XX-XY system and this system was found from both annual and perennial dioecious species. The genus-wide survey identified that all three species under genus *Spinacia*, all 13 species from genus *Pheonix*, and the only species in *Carica* are dioecious with XY sex chromosomes, indicated the sex-determining system might evolve independently before the speciation from these genera in different families. ZZ-ZW sex chromosomes are less prevalent and a pair of female heterogametic sex chromosomes were identified from wild strawberry and poplar (Cronk, 2005; Tennessen et al., 2016). Interestingly, the existence of both ZW and XY chromosomes were found in some closely related genera, such as genus *Populus* and *Salix* from the family Salicaceae (Geraldes et al., 2015; Zhou et al., 2020). Independent genomic analysis from two genera revealed a strong orthologous candidate gene as sex-determining genes for both species, which might be the effects of convergent evolution. The XO sex termination system was extremely rare and features by one species *Rumex acetosa*, whose sex type was determined by the ratio between X chromosome and autosomes (Ming et al., 2011b). Sex determination could be regulated by non-genetic factors, such as temperature, hormones, and epigenetics regulation. One direct evidence of epigenetic regulations was identified from sex reversal of male to andro-hermaphrodite flowers in dioecious white campion (*Silene latifolia*) by chemical treatments of seeds. The sex reversal is likely to be triggered by epigenetic intervention of carpel suppression gene(s).

The rise and evolution of sex chromosomes

Dioecy is the pre-condition of sex chromosomes. The “two-locus” model is one of the most prevalent models to explain the rise and evolution of sex chromosomes (Charlesworth & Charlesworth, 1978; Charlesworth & David, 2004). It was proposed that the sex chromosomes evolved from autosomes and the formation of sex chromosomes can be classified into six stages (Fig 1.4). In stage 1, two mutations are required to establish the stable dioecy from hermaphrodism, including one recessive mutation (loss of function) for male sterility and the other dominant mutation (gain of function) for female sterility. (Charlesworth & Charlesworth, 1978). The pair of homologous chromosomes where mutations occurred in this stage were termed as proto-sex chromosomes and recombination have not been suppressed as hermaphrodites or neuters remained in their progenies. In stage 2, suppression of recombination plays a substantial role in maintaining dioecy, featured by DNA methylation or chromosomal rearrangement (inversion, insertions, and deletions) which leading to mutations of Y-encoded genes from the surrounding region. (Jaarola et al., 1998; Janousek & Mrackova, 2010). Due to the limited scales of deleterious effects during degradation, the YY genotype is still viable in this stage. In stage 3, the gradual accumulated repetitive sequences and divergent genes in non-recombining regions of each homolog extended to a lager flanking region, leading to a considerable gene degradation, particularly on the Y chromosome. Those extensive mutations finally caused the lethality of YY genotypes. In stage 4, the Y-linked region continuously spread to the majority of chromosome and significant expansion of the Y chromosome was largely contributed to transposable elements insertions. In later two stages, the expanded Y chromosome accompanied with a higher chance of loss of gene function introduced the instability which results in a shrinkage of the Y chromosome with a morphologically smaller size compared with X counterpart. The Y chromosome in the last stage was completely lost due to large scale gene degradation and the new sex determination is based on numbers of autosome and X chromosome ratio.

The classification of sex chromosome evolution sets up important criteria to characterize the sex chromosomes from different species and information from a wide range of studies indicated that most sex chromosomes from animals are more ancient compared with most plants, fish, and insects (Bergero & Charlesworth, 2009; Westergaard, 1958). The fact also explains that sex chromosomes of animals have been widely investigated whereas not as many sex chromosomes have been investigated in plants. As a proxy for estimating the stages of sex chromosomes, the morphological characterization of sex chromosomes provided the evidence in terms of the degree of degradation from respective Y (W) sex chromosomes by comparing the size of each homologous pair. Accordingly, the sex chromosomes were classified as homomorphic sex chromosomes and heteromorphic sex chromosomes. The homomorphic sex chromosomes were found from most species with sex chromosomes in very early stages. There were less negligible size variations of X and Y or Z and W chromosome by observation of mitotic cells during metaphase. This is because these sex chromosomes initiated quite recently and there was an insufficient accumulation of repetitive sequences or deleterious mutations in the sex-determining region. Alternatively, heteromorphic sex chromosomes widely existed from species with ancient sex chromosomes, which can be classified as a larger Y(W) chromosome or a smaller Y(W) chromosome compared with their respective counterparts. The larger Y(W) chromosome is attributed to the accumulative insertions and smaller Y(W) chromosome is the outcome of gene loss leading to shrinkage at the very late stage of sex chromosome evolution.

Sex chromosomes in flowering plants

In the past few decades, a wide range of sex chromosomes was identified among land plants, including angiosperm, gymnosperm, and bryophytes. From these major classifications, the sex chromosomes have not been widely investigated which includes only 0.01%, 0.6%, and 0.02% of each analyzed based on either morphologically, genetically, or cytogenetically characterized sex chromosomes (Ming et al., 2011a). The 48 species across 20 families are found with evidence of sex chromosomes among the 146,000 species from 960 genera and 200 families of dioecious flowering plants (Renner & Ricklefs, 1995). Among these species, 17 species from 4 families share the heteromorphic sex chromosomes such as white campion (*Silene latifolia*), sorrel (*Rumex acetosa*), and date palm (*(Phoenix dactylifera*). The other 20 species from 12 families are in homomorphic fashion, includes papaya (*Cariaca Papaya*), strawberry (*Fragaria Virginiana*), spinach (*Spinacia Olracea*), and asparagus (*Asparagus officinalis L.*) (Ming et al., 2011a). Here we representatively selective species and made a respective summary of sex chromosomes research in each plant.

The wild strawberry (*Fragaria virginiana*) is a dioecious species with a pair of ZW sex chromosomes from family Rosaceae. Genetic mapping revealed the segregation of either hermaphrodites or neuters in a large F_2 mapping population and a 5.6 cM genetic distance was mapped between two sex-determining loci at linkage group 6 ((Spigler et al., 2010; Spigler et al., 2008). These results indicated the strawberry sex chromosomes are at the earliest stage of evolution, the stage without recombination suppression around the sex determination loci. Interestingly, genetic mapping of the sister species from *F. moschata* revealed a strict 1:1 sex type segregation ratio without the existence of hermaphrodites or neuters (Spigler et al., 2010). This might be explained by the independent sex chromosomes evolution after speciation under the *Fragaria* genus.

Asparagus (*Asparagus officinalis)* is a representative dioecious species from family Asparagaceae with a pair of XY nascent sex chromosomes. The existence of sex chromosomes was tested by mapping the sex determination loci to a small non-recombining region, along with evidence generated by trisomic analysis (Loptien, 1979). The viability of YY asparagus and homomorphic sex chromosomes indicates its sex chromosomes at stage 2. The recently sequenced genome of asparagus by sequencing a double-haploid (DH) from anther-culture induced asparagus yielded an assembly of sex chromosomes in two XX and YY genotypes. The non-recombining sex determination region was defined in a 750Kb region with only 13 functional genes. One Y-encoded gene was identified from *Arabidopsis* knock-out experiment, which served as a key gene related to stamen development encoded by the DUF247 domain. Another gain of function mutation Y-specific *SOFF* was identified as a carpel development suppressor (Harkess et al., 2017a). Identification of these two functional genes justified the "two-locus" model of sex determination.

As an example of tree species with sex chromosomes, sex determination in papaya (*Carica papaya*. L) is controlled in an XX-XY system. Besides the X and Y chromosome related to female and male flower development, a Y^h chromosome with a slight 1.2% of differences with the Y chromosome was identified which determines hermaphrodite flower development in papaya (Yu et al., 2008b). The SDR of papaya was genetically defined by 225 sex co-segregating AFLP markers which indicated the severe recombination suppression around sex determination loci (Ma et al., 2004). The further fine mapping, construction of physical maps, and sequencing of X and Y^h chromosome defined an 8.1Mb hermaphrodite specific Y chromosome (HSY) and a 3.5Mb female X counterpart, which provided a molecular basis for studying the genes within the region. Comparison of the pair-genes between X and Y^h chromosome identified three evolutionary strata which aligned to borders of two large scale inversion happened around 1.9-9.5 million years ago (MYA) and 0-6.9 MYA along with abundance retrotransposons insertion (Wang et al., 2012). This expanded Y chromosome and YY lethality indicated the sex chromosomes in papaya are at stage three.

Silene is the genus from Caryophyllaceae family and quite a few species under the genus were identified as dioecy plants. Moreover, both the homomorphic and heteromorphic sex chromosomes were identified from species within the genus. One dioecious clade includes *S. diclinis, S. dioica, S. heuffelii, S. latifolia,* and *S. marizii* are classified as species with homomorphic sex chromosomes; The sex chromosomes from another clade which includes *S. colpophylla, S. otites,* and *S. acaulis* are in heteromorphic fashion (Castillo et al., 2010). The well-studied *S. latifolia* have a greater size of the Y chromosome and at least three genes are responsible for sex determination: one gene encoded for carpel suppression and two genes for stamen development (Bergero et al., 2007a). Further identification of 11 X-linked genes suggested the degradation of missing alleles from the Y chromosome, mark the more advanced evolution stage of sex chromosomes (Jamilena et al., 2008; Matsunaga, 2006). Genomic analysis also indicated the inversion of SDR has been rarely identified but highly repetitive elements were found within the region, which suggested the insertions of repetitive sequences might play a more substantial role of recombination suppression during the evolution.

The sex determination of persimmon tree species from *Diospyros* genus under the family Ebanaceae is well studied. The morphological feature of flowers is featured by the fertile stamen along with the arrested carpels (degraded structure) as male, defective anther along with the well-developed carpels for female individuals (Duangjai et al., 2006). Substantial progress has been made by identification of sex-specific sequences based on primary genomic assembly from female and male individuals revealed several male-specific scaffolds, combining with transcriptome analysis, there were 62 differentially expressed genes between male and female persimmon. The downstream analysis identified one Y-specific determinant candidate *OGI* (Akagi et al., 2014). The Y-encoded *OGI* induce the small RNA interference which targets gene *MeGI* on an autosome, as a transcription factor regulating anther fertility and sex determination in persimmon. The discovery of the mutant is an exception as the 'two-gene model' of plant sex chromosome evolution and indicates the variation and plasticity of sex determination mechanisms in higher plants.

Spinach is an idea dioecious plant for sex determination studies

Garden spinach (*Spinacia oleracea* L.) is a common green leafy vegetable of *Spinacia* genus, under the subfamily Chenopodioideae, and family Amaranthaceae. The enriched nutrients, carotenoids, folate, vitamin C, calcium, and iron contribute to the increasing popularity among more than 60 countries and regions in the past decades with an annual production of 26.7 million tonnes worldwide (2018) based on the annual report from FAOSTAT (http://faostat3.fao.org/ home/index.html), the production of spinach increased almost ten folds in the past 40 years. Besides the domesticated spinach, the *Spinacea* genus consisted of two other wild relatives: *S. turkestanica* and *S. tetrandra* (Sneep, 1982). This relationship was further confirmed by transcriptome variants from 107 domesticated spinach accessions and 13 accessions from wild relatives, indicating the *S. turkestanica* is likely to be the progenitors of cultivated spinach.

Up to date, 2097 accessions under *Spinacia* genus were collected from word-wide germplasms which can be represented by 1959 accessions from *S. oleracea*, 89 from S. turkestanica , and 49 from *S. tetrandra* (Ribera et al., 2020). Among those collections, 353 accessions across 37 countries and regions of spinach were collected from U.S. national plant germplasm system (https://www.ars-grin.gov/npgs/). Central Asia, Northern American, and Europe are the major sources for germplasm collection. These collections can be classified as 339 accessions from *S. olracea,* six accessions from *S. turkestanica* and 13 accessions from *S. tetrandra.* Transcriptome diversity and population analysis suggested the spinach accessions derived from modern spinach breeding can be divided into three main groups, as *S. turkestanica* and *S. tetrandra* were clustered as a separate group, the US and European accession grouped as one cluster, most Chinese and eastern Asian cultivars grouped as the other group (Xu et al., 2017a). Combing the historical records and geographical evidence, domestication of spinach started about 2000 years ago (Rubatzky & Yamaguchi, 2012). The original vegetable spinach originated from Persia (Central Asian region) and introduced to Eastern Asia, Northern America, and some European regions. Traits for crop improvements include cold resistance, disease resistance, coloration, and yields.

Besides spinach, there are two important agronomic crops includes quinoa (*Chenopodium quinoa*) and sugar beet (*Beta vulgaris*) in the family Amaranthaceae. The genomic synteny analysis revealed an extensive conservation of orthologous genes between spinach and sugar beet, along with similar numbers of annotated protein-coding genes (27,421 genes from sugar beet and 25,495 genes from spinach). However, the spinach is marked as having a nearly 1/3 increased genome size, mostly from expanded intergenic regions, which could be partially explained by the burst of transposable elements happened ~1.5 MYA (Dohm et al., 2014a; Xu et al., 2017a). For genome evolution analysis, besides the whole-genome triplication event (gamma event) shared by all eudicots , there is no recent whole-genome duplication (WGD) event in spinach while a recent WGD happed in sugar beet, indicated an independent speciation event in different genera.

Spinach is dioecious with separate female and male individuals, as pure staminate flowers from male plants and pure pistillate flowers from female plants (BEMIS & WILSON, 1953; Ellis & Janick, 1960). Occasionally, the monoecious and andromonoecious spinach exist in some accessions (Janick & Stevenson, 1955b) (Table 1.1). The staminate flowers are morphologically distinguished by 4~5 stamens with yellow pollen at the end of filament during maturity, whereas pistillate flowers are identified by 4~5 white stigma and long sepals (Figure 1.3). Besides unisexual flowers from the same individuals. Monoecious spinach was also identified along with separated staminate flowers and pistillate flowers on the same plant (Janick & Stevenson, 1955b), which exhibited the different staminate/pistillate flower ratios among genotype (Figure 1.2). Andromonoecious spinach was identified in two spinach accessions includes ‘Cornell-NO.9’ and ‘Long standing Bloomingdale’ (Yamamoto et al., 2014). Some plants from these two accessions with XY genotype are bearing hermaphrodite flowers and male flowers on the same plant. For dioecious spinach, out-crossing is the only way of sexual reproduction. However, the monoecious and andromonoecious spinach can be cross-pollinated or self-pollinated. Seeds from self-pollinated andromonoecious spinach are morphologically different compared with seeds from other sex types, which are in a smaller size and poorly developed pericarp like a naked seed (Yamamoto et al., 2014).

The discovery of monoecious spinach can be traced back to 5-6 decades ago. Previous reports suggested the inbreeding monoecious spinach are bearing extremely low female flowers, whereas F_1 populations generated by true-breeding monoecy crossing with pure pistillate spinach revealed the increasing pistillate flowers from F_1 (Janick, 1955a). It was speculated that the expression of monoecious spinach is controlled by a partial dominant determinant located on the X chromosome of spinach, supported by the 3:1 segregation ratio of self-pollinated monoecious spinach (Janick & Stevenson, 1955a). The self-pollinated highly female-monoecious individual segregated only female and monoecious offspring indicated the gene related to monoecious spinach is in the X chromosome. Also, genetic mapping for monoecious (M) genes based on SCAR and ALFP makers suggested that the locus was 12cM distant from the X counterpart of sex-determining region in dioecious spinach. Further studies narrowed down the interval for *M* gene into 7.5cM along with nine flanking ALFP markers (Takahata et al., 2016), while no monoecism co-segregating maker has been identified due to limited numbers of genetic markers. Further fine mapping could potentially identify a candidate gene for monoecy in spinach in a larger mapping population.

Andromonoecious spinach was less common and only identified from a few spinach accessions. Self-pollinated XY andromonoecious spinach produces a small number of seeds attributed to the existence of hermaphrodite flowers with degraded pistils. Seeds generated by self -pollination have a low germination rate, attributed to the defection of hermaphrodite flowers. The YY individuals were identified in segregating progenies from self-pollinated hermaphrodite flowers and viability in YY seedlings is very low (Wadlington & Ming, 2018b).

The study of dioecism in spinach can be dated to the last few decades by crosses between male and female spinach. Such crosses yielded a 1:1 segregation ratio for the male to female among progenies and it was hypothesized that sex determination in spinach is related to a pair of X/Y chromosome (BEMIS & WILSON, 1953; Janick, 1954). The polyploid spinach experiments from XXXX and XXXY spinach genotype crossing yielded the same distribution as diploid spinach, matching the feature of segregation pattern of single gene genetics (Janick, 1955b; Janick, 1957). In addition, the chromosome containing the sex-determining loci was mapped to the largest chromosome by trisomic analysis, which is mapped from a cross between triploid and diploid spinach and a 2:1 female to male segregation of progenies from XXY genotype of the female parents (Ellis & Janick, 1960; Janick et al., 1959) (Fig 1.3a). The more recent cytogenetic experiments based on chromosomal painting provided additional molecular evidence of sex chromosomes (Fig 1.3b). For instance, the hybridization of a 45S rDNA with metaphase mitotic cells identified two signals around the centromeric region of one pair of chromosomes from female spinach while only one signal from male spinach (Deng et al., 2013b; Lan et al., 2006). This heterogeneous distribution of florescent signals indicated the presence of the X and Y chromosomes in spinach. Detection of sex-specific signals was performed by chromosomal microdissection of a sex-specific DNA maker T11A and chromosomal painting, which confirmed the largest chromosome pairs as sex chromosomes at the molecular level (Deng et al., 2013c).

The evidence of sex chromosomes in spinach was also characterized by development of sex co-segregating molecular markers and genetic mapping. There are two sex-linked DNA markers with strong linkage with spinach dimorphisms includes T11A and V20A, but it was still unclear about the genomic position of these two markers (Akamatsu & Suzuki, 1999). The first genetic map of spinach mapped a total 101 ALFP and 9 microsatellite DNA markers into seven linkage groups and one microsatellite marker (SO4) was mapped at 1.4cM distant with sex-determining region from linkage group 3 based on linkage analysis of 161 BC_1 population (Khattak et al., 2006). A recent high-density linkage map based on GBS (SLAF-seq) approach provided a higher resolution of spinach linkage map derived from a 148 BC_1 population, dividing into six linkage groups with 4088 SLAF makers been anchored. The sex-determining region for X/Y were mapped to a 17.48cM and 2.73cM interval on the largest linkage group (Qian et al., 2017). Six Y-linked DNA markers were developed from a male-specific region (MSY) sequences generated by the BAC-BAC approach, and the markers were verified in a large population. The approximate 500Kb sequences from six BAC libraries are highly repetitive with very few genes identified and extremely low recombination rate around the region (Kudoh et al., 2018b). The detection of the suppressed recombination region and highly repetitive BAC sequences provided the genetic evidence of sex chromosomes in spinach. Nevertheless, the positions of these BAC libraries are not been mapped and it is unknown about the genomic size of suppressed recombination near sex-determining loci.

The advancement of sequencing technologies and enhanced algorithms of bioinformatic tools for genome *De-novo* assembly provided good opportunities for studying large and complex plant genomes. Among family Chenopodioideae, the genomes of seven species under the seven genera of the family were sequenced, which provided valuable sources for comparative genomic analysis to address the problem of genome evolution and speciation (Dohm et al., 2014b; Jarvis et al., 2017; Patterson et al., 2019; Rodriguez Del Rio et al., 2019; Sunil et al., 2014; Wang et al., 2019). The first published spinach reference genome consisted of an approximate 498Mb genome in scaffolds level that is only around 50% of estimated genome size based on estimate from flow cytometry (Arumuganathan & Earle, 1991; Dohm et al., 2014b). The limited N50 also hindered completing a high-quality chromosomal level genome for evolutionary genomic analysis. A more recent published Sp75 (accession) genome substantially improves the assembly quality into 6 pseudomolecules by optical map correction and scaffolding by high- density genetic maps (Xu et al., 2017b). Genome analysis revealed that spinach genome expanded around 1.5 million years ago triggered by recent LTR-RTs burst while no whole-genome duplication detected within the genome. The recent insertions of DNA/RNA-mediated repetitive elements contributed to more than 70% of repetitive sequences of the genome and it makes the scaffolding of genome a challenging task. Due to the highly repetitive nature and limited reads from Illumina sequencing, there was only about 46.5% of the genome been anchored to chromosomal level and there was no evidence regarding the sex chromosomes and region associated with sex-determining loci.

Besides the information from DNA sequencing. The transcriptome level *De-novo* provide some insights for sex chromosomes and facilitate the identification of sex-linked genes from X and Y chromosomes. Recent transcriptomic-based gene discovery from male and female sib-crosses identified a 354 sex-linked SNPs among progenies. Amplification of those genes associated with sex linkage SNPs was further tested by linkage analysis in a large population (Okazaki et al., 2019). 12 pairs of X/Y genes were identified with suppressed recombination along with sex linkage pattern and co-segregating with sex-determining region.

Although these genomic resources benefit spinach breeding and discovery of new genes related to favorable agronomic traits, the sex chromosomes study in spinach is still limited due to its highly abundant repetitive sequences and large genome size. Construction of high-density genetic maps could effectively anchor fragmented scaffolds into chromosomes but the improved genome assembly by a higher sequencing depth and long-reads approach is still needed to resolve gaps between scaffolds. For sex determination studies, the identification and sequencing of SDR is the first step to include annotated genes in this region for further analyses.

Concluding remarks and future perspectives

The genetic mapping of spinach sex-determining loci and chromosomal painting play a substantial role to justify the presence of sex chromosome. However, the SDR of spinach has not been fully sequenced and genes underlying dioecy have not been identified. Particularly, several questions remain: Where is the position of this SDR? How large this region could be? How many genes were in the region? It is challenging to understand the mechanisms for sex determination in spinach without understanding these questions. Therefore, a further genome-wide study and experiments utilizing DNA sequences will be needed.

To fulfill the goal, we will apply a variety of plant genetic principles and methodologies which include genetic mapping, genome sequencing, and comparative genomic analysis to narrow down the region around the sex-determining region from the spinach Y chromosome and compare the genomic variation of its X counterpart. In the first experiment, we will perform our studies by making crosses of male and female spinach and developing the DNA markers to test plants across 40 spinach accession from different countries. We would exam the genomic variations between X/Y sequences where our sex co-segregating markers derived and detect suppressed recombination within the sex-linked region. In our second experiment, we will define the position and approximate size of SDR by genetic mapping from a pseudo-test cross population and characterize the sex linkage effects of this region. For the third experiment, we are aiming for conducting the genome assembly of the male and female genome to identify structural variations between X and Y chromosomes and evolutionary forces driving sex chromosomes evolution. By comparing sex-specific genes, we might be able to narrow down the candidates as sex determinants in the Y chromosome for future studies.

Tables and Figures

Table 1.1 Classification of spinach sex type

Type	Description
Dioecious	Individual is exclusively male or female by respective flower types
Monoecious	Individual with separate staminate and pistillate unisexual flowers
Hermaphrodite	Perfect flowers with both carpels and stamens
Andromonoecious	Individual plant with both staminate and hermaphroditic flowers

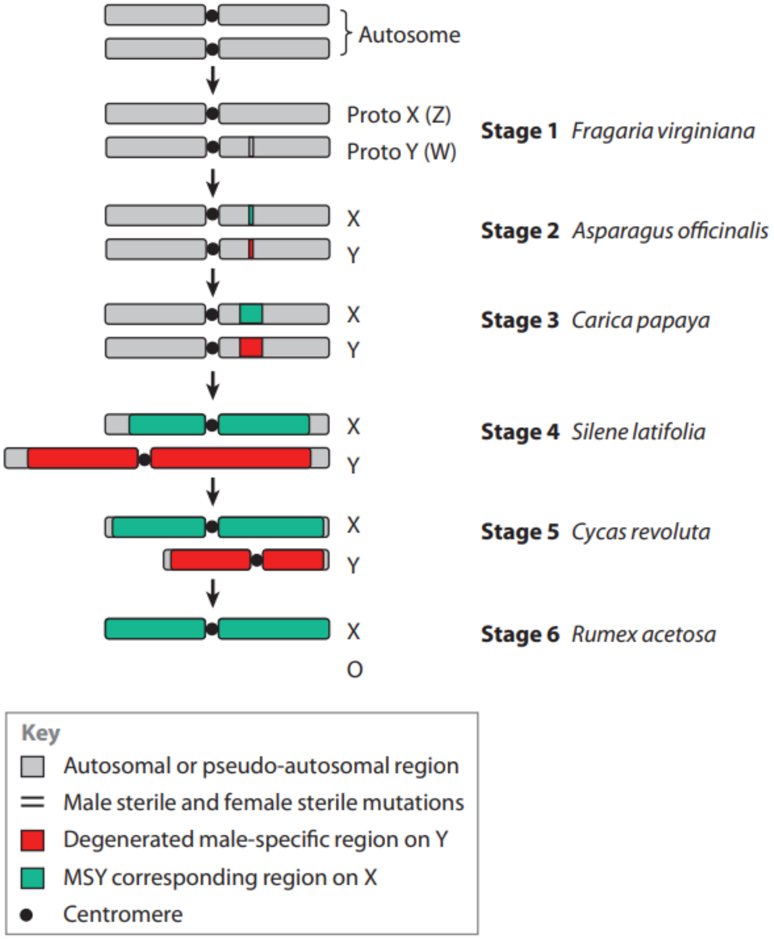
Autosome
Proto X (Z)
Proto Y (W)
Stage 1 Fragaria virginiana
X
Y
Stage 2 Asparagus officinalis
X
Y
Stage 3 Carica papaya
X
Y
Stage 4 Silene latifolia
X
Y
Stage 5 Cycas revoluta
X
O
Stage 6 Rumex acetosa
Key
Autosomal or pseudo-autosomal region
Male sterile and female sterile mutations
Degenerated male-specific region on Y
MSY corresponding region on X
Centromere

Fig 1.1 Summary of six stages of sex chromosome evolution

The six stages of sex chromosome evolution cited from Ming et al. in their publication Ming R, Bendahmane A, and Renner SS (2011) Sex chromosomes in land plants. *Annu Rev Plant Biol* 62:485-514.

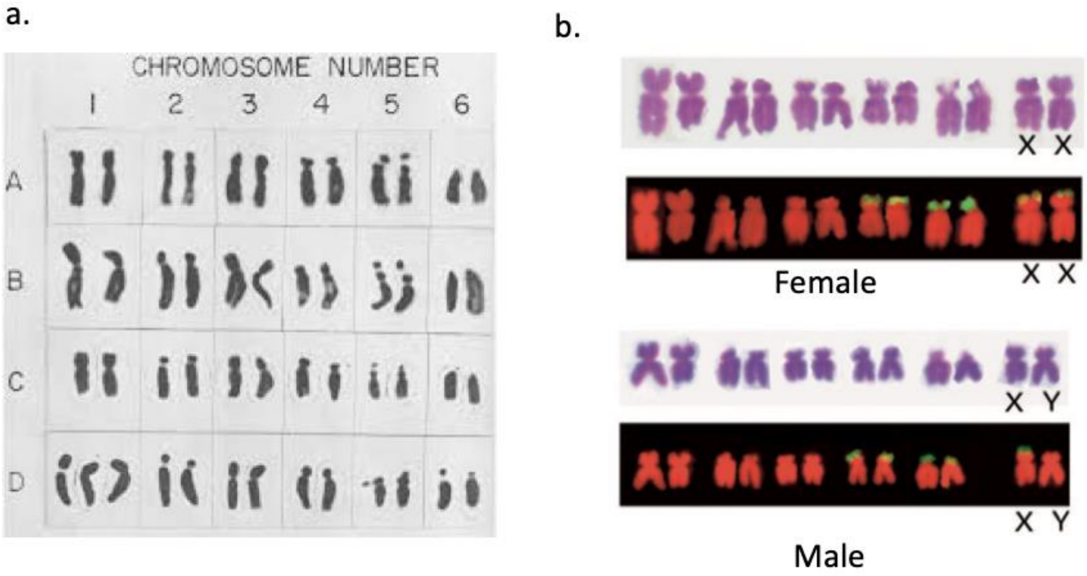

Fig 1.2 Cytogenetics analysis of sex chromosomes in dioecious spinach

a. Trisomic spinach for each chromosome derived from crosses between a diploid and triploid spinach. The D represent the trisomy of chromosome which carries the sex-determining loci, supported by an approximate 2:1 sex type segregation ratio of their progenies.

   Cited from Ellis, J. R., and Jules Janick. "THE CHROMOSOMES OF *SPLNACIA OLERACEA*." *American Journal of Botany* 47, no. 3 (1960): 210-214.

b. The chromosomal painting of female and male spinach derived from hybridization between a 45S rDNA probe. The two signals identified from one chromosome pair in female spinach and one signal from the same chromosome pair in male spinach indicated the existence of sex chromosome.

   Cited from Lan, T., S. Zhang, B. Liu, X. Li, R. Chen, and W. Song. "Differentiating sex chromosomes of the dioecious *Spinacia oleracea L.*(spinach) by FISH of 45S rDNA." *Cytogenetic and genome research* 114, no. 2 (2006): 175-177.

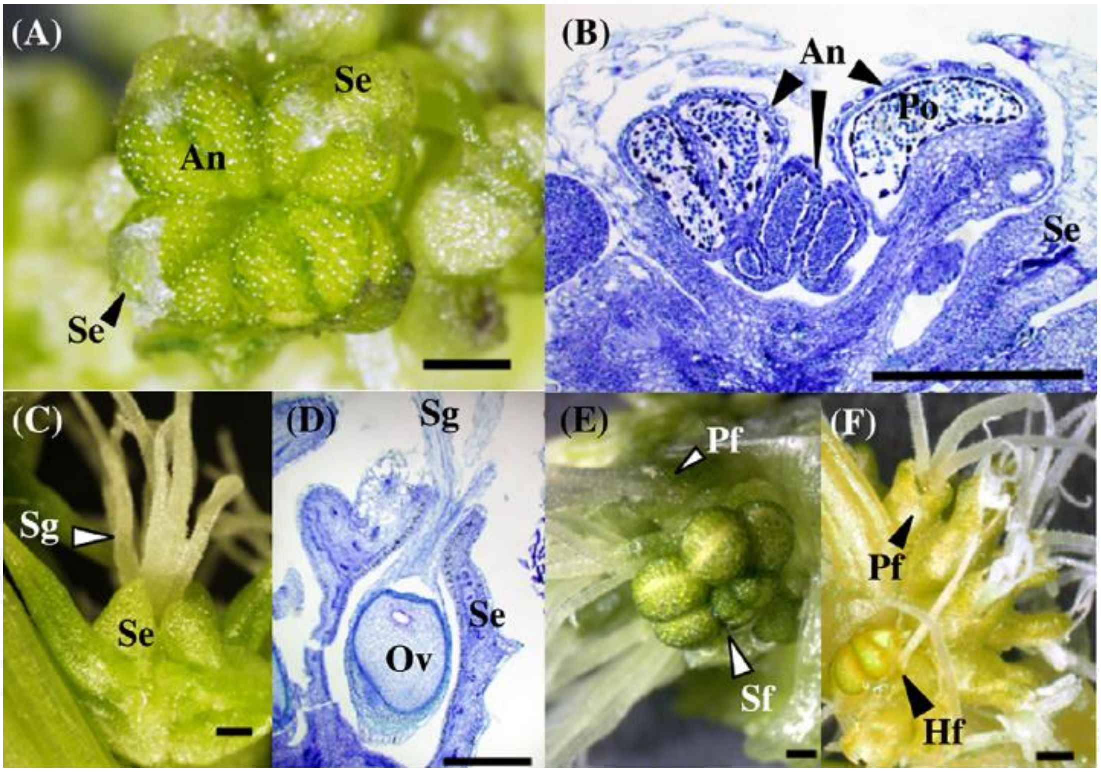

Figure 1.3 Flower structure of dioecious spinach

Floral morphology dissection of dioecious spinach Cited by Onodera, Yasuyuki, et al "Monoecy and gynomonoecy in Spinacia oleracea L.: Morphological and genetic analyses." *Scientia horticulturae* 118, no. 3 (2008): 266-269.

(A) An staminate flower borne on a staminate plant.

(B) Longitudinal section of a staminate flower.

(C) A pistillate flower formed on a pistillate plant. (D) Longitudinal section of a pistillate flower.

(E) Mixed pistillate and staminate flowers. (F) Mixed pistillate and hermaphroditic flowers borne in the same cluster.

Abbreviations: An, anther; Se, sepal; Po, pollen; Sg, stigma; Ov, ovule; Pf, pistillate flower; Sf, staminate flower; Hf, hermaphroditic flower

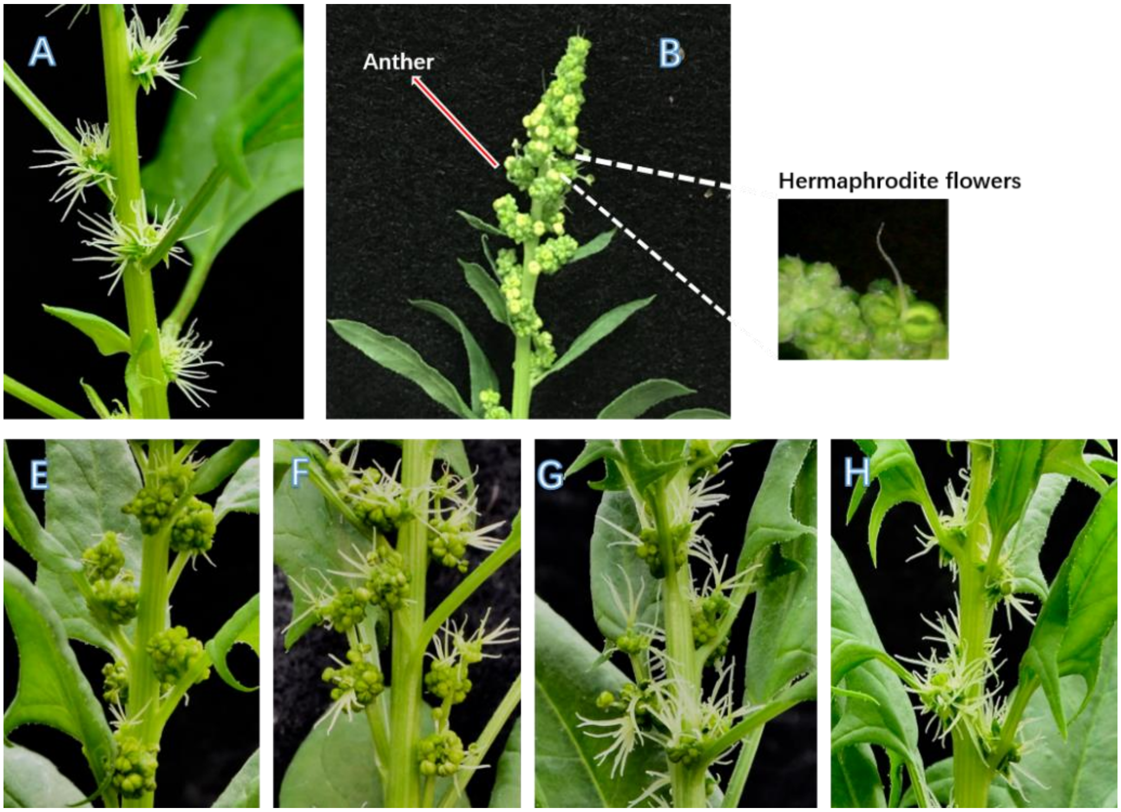

Figure 1.4 Monoecious spinach and andromonoecious spinach

A. Monoecious Spinach: staminate and pistillate flowers

B. Female spinach: Pure pistillate flowers

C. Androdiouecious spinach: hermaphrodite flowers and staminate flowers

D. Monoecious spinach with 5-10% pistillate flowers

E. Monoecious spinach with 20-40% pistillate flowers

F. Monoecious spinach with 50-70% pistillate flowers

G. Monoecious spinach with >90% pistillate flower

Chapter II. Identification of structural variations and polymorphisms of a sex-linked scaffold in dioecious spinach (*Spinacea olracea*)

Abstract

Spinach (*Spinacea olracea*) is a common dioecious leafy vegetable and dioecy is maintained by a pair of XY sex chromosomes. Limited studies were conducted to investigate the genomic landscape of the region near sex-determining loci. We screened the structure variations (SVs) between Y-linked contigs and a 1.78Mb X scaffold (Super_scaffold 66), which enabled the development of 12 sex co-segregating DNA markers. These markers were tested in 86 males and 95 females of a pseudo-test cross mapping population and further validated from 266 female and 245 male plants from 40 spinach accessions, which comprised of 692 individual plants with a strong sex co-segregation pattern. In addition, we found that Super_scaffold 66 was highly repetitive along with the popularity of LTR-RTs insertions and decreased microsatellite distribution compared with autosomes, which matches the extremely low gene density at one gene per 197.78 Kb compared with one gene per 39.22Kb of genome-wide average. Synteny analysis between Y contigs and Super_scaffold 66 revealed a 335kb accumulative Y contigs (non-continuous) and a 450Kb X counterpart along with SVs and wide-spread tandem duplications. Among nine genes from Super_scaffold 66, one ABC transporter gene revealed some noticeable SVs between the Y contig and its X counterpart, as an approximate 5Kb *Gypsy* LTR-RT insertion and exon-intron variation of its Y-linked neighboring gene. We proposed that Super_scaffold 66 is closely linked to sex-determining loci. The spread of 12 sex co-segregating markers from this 1.78Mb genomic region indicated the existence and expansion of sex determination region during progression of the Y chromosome.

Key words: spinach sex determination, synteny analysis, sex-linked markers

Introduction

As the most prevalent sexual reproductive mechanisms among angiosperms, hermaphrodism accounts for 89% angiosperms, whereas only 6% species are dioecious, and 5% are monoecious (Charlesworth, 2013b; Ming et al., 2011b; Renner, 2014a). The wide-distribution of dioecious plants among 75% angiosperm families indicated independent origins of dioecy from different lineages but only a minority of dioecious plants are studied in terms of evidence of sex chromosomes (Henry et al., 2018; Vyskot & Hobza, 2004). The best known “two-factor” model proposed that sex chromosomes evolved from autosomes by stages (Charlesworth & Charlesworth, 1978; Lewis, 1942). In stage one, one recessive mutation (loss-of-function) encoded for male sterility and the other gain-of-function mutation for carpel suppression initiated the transition to dioecy. There is low frequency of hermaphrodites and neuters in the population due to recombination between the two loci (Spigler et al., 2010; Spigler et al., 2008). In stage two, recombination was suppressed near two mutations associated with chromosomal rearrangements (inversion, insertions, and translocation) or DNA methylation (Jaarola et al., 1998; Janousek & Mrackova, 2010) while a small non-recombining region formed with YY genotype, which is viable. In stage three, the accumulative deleterious mutations, gene content variations of each homolog, and degeneration of Y-linked genes spread to neighboring regions, resulting YY genotype lethality. In the later few stages, a considerable gene loss and retrotransposon insertions shape the sex chromosome evolution, leading to the expansion then shrinkage of the Y chromosome.

In the past few decades, genome sequencing, sex-biased transcriptome studies, and comparative genomics provided valuable resources to study sex chromosomes in different dioecious plants and to test the “two-gene” sex determination hypothesis. In garden asparagus, sequencing of double haploid (DH) genotype for both “XX” and “YY” genotype led to the identification of a 750Kb male-specific region (MSY). The two Y-encoded genes named *SOFF* and *aspTDF*, act independently to suppress the carpel development or promote stamen development (Harkess et al., 2017a). In kiwifruit, the transcriptome of floral-biased expression pattern and genome sequencing revealed one Y-encoded *FrBy* gene with strong expression in tapetal cells relates to stamen development and another gene acting as a suppressor of feminization named *SyGl* (Akagi et al., 2018; Akagi et al., 2019). *Phoenix* genus-wide sequencing of female and male date palm from 14 species identified few male-specific sequences and further extended them by BAC clones or phased single molecules sequencing. Two male-specific genes (*CYP703* and *GPAT3*) were identified as candidates to be involved in male flower development while another LOG-like gene, which translocated from autosomes, might play a critical role of suppressing female flower development (Torres et al., 2018). Besides the "two-factor" model, an alternative mechanism was identified from sex determination in dioecious persimmon which is regulated in a "one-factor" from diploid persimmon. Particularly, one Y-specific transcription factor *OGI* which transcribed into a Y-encoded small RNA to target the *MeGI* and manipulate the sex determination by suppressing the feminization of *MeGI* (Akagi et al., 2014). These reports shed light on the diversified forces of sex chromosome evolution from a variety of plant families with different evolutionary stages and underlying sex determination genes. Thus, a detailed identification and dissection of genomic structure of sex determination region (SDR) on the Y chromosome from respective dioecious species will provided direct evidence regarding the presence of sex chromosomes and serve as a genomic resource to study sex-determining genes.

Spinach (*Spinacia olracea*) (*2n = 12*) is a dioecious annual plant with a short life cycle, a large germplasm collection, and an established genic transformation system (Zhang & Zeevaart, 1999). These advantages enabled the research of spinach sex determination since1950s. The sex determination in spinach was proposed to be controlled by one locus based on a 1:1 segregation pattern for male and female progenies from crosses between female and male plants (Iizuka & Janick, 1962; Janick & Stevenson, 1955a). The largest chromosome pair was identified as the sex chromosomes harboring sex-determining genes based on trisomic analysis derived from diploid-triploid crosses (Ellis & Janick, 1960; Janick et al., 1959). Recent chromosomal painting by either 45S rDNA probes or DNA sequence identified the largest pair of chromosomes with biased signals near centromeric region between male and female spinach (Deng et al., 2013b; Deng et al., 2013c; Lan et al., 2006). Besides the cytogenetics evidence, the sex-linked region of spinach was genetically mapped with few sex-linked DNA markers and partial Y-linked regions generated by five BAC-clones with an accumulative 500Kb region accompanying enriched repeat elements (Akamatsu et al., 1998; Khattak et al., 2006; Kudoh et al., 2018b; Onodera et al., 2011). However, there was still a lack of evidence regarding the degree of structural variations

(SVs) and sequence divergence between SDR and its X counterpart.

Viable YY genotype indicated a pair of nascent sex chromosomes in spinach, which is precious material for high-quality genome assembly because the genome assembly of two copies of Y chromosomes is more feasible compared with XY genotype assembly (Onodera et al., 2011; Wadlington & Ming, 2018a). Here, we sequenced the YY spinach genotype and assembled into a contig-level sequences using PacBio long-reads sequences. The draft sequences serve as a genomic basis to develop sex co-segregating polymorphic markers based on SVs between the Y chromosome and its X counterpart derived from sequence alignment. DNA from 692 plants were used to validate those DNA polymorphic markers by matching respective genotypes to their sex phenotypes. We further performed a genomic landscape survey regarding the distribution of genes and repetitive sequences. Our work provided evidence of SDR in spinach and SVs between the X and Y chromosome.

Methods and materials

Plant growth and DNA extraction

One spinach F_1 population was developed by a cross between male and female individuals from two different accessions, PI 217425 (Cornell-NO.9) and PI 347812 (Viroflay). Additional 40 spinach (*Spinacia Olracea*) accessions of 17 countries collected from USDA germplasms were planted for marker analysis. Plants above were maintained in 20-22°C environment in a greenhouse at the University of Illinois at Urbana-Champaign. Phenotyping for each plant was identified based on flower morphology. Fresh leaves of each F_1 individual were collected then treated within liquid nitrogen for preservation. For each spinach accession, 4~10 individuals of female and male leaves were pooled for DNA extraction with three technical replicates for each sex type. DNA extraction was followed by CTAB method with 1% Edward Buffer and DNA products were tested by 1.5% agarose gel and quantified by Nanodrop. The final DNA was diluted to 20ng/µL as DNA template for PCR.

Sequencing of YY spinach

The YY spinach was segregated from the self-pollinated andromonoecious spinach line PI 217425 (Cornell-NO.9), adopted from previous studies (Wadlington & Ming, 2018a).

Andromonoecious spinach individuals were kept in a 20°C growth chamber for pollination. DNA extraction was conducted by CTAB method described above and final DNA products were tested by two sex-linked markers: SpoX and T11A (Akamatsu & Suzuki, 1999; Wadlington & Ming, 2018a). The 16 selected YY spinach based on positive results of SpoX marker were pooled for genomic DNA extraction, using the SMRTbell DNA extraction protocol for *Arabidopsis* (Kim et al., 2014). High-quality sequencing libraries were constructed from the DNA then sequenced at 63X depth. Sequencing raw data were further assembled using the program CANU pipelines (Koren et al., 2017).

Identification of repetitive elements of X-linked scaffold(s)

The complete genomic sequence of three sex-linked markers, including an X-specific marker (SpoX), T11A, and V20A (Akamatsu & Suzuki, 1999; Wadlington & Ming, 2018a) was used to retrieve putative X-linked regions from the spinach reference genome (Xu et al., 2017a) through BLASTn with e-value = 1e-3, identities = 90, and hsps_cov = 90 as a cutoff. The genomic region from reference genome associated with hits were selected for further repetitive elements analysis. Initially, the complete genomic sequences of reference genome were used for building a repetitive sequence library by RECON and RepeatScout (http://www.repeatmasker.org/RepeatModeler/). The LTR_harvest program was further used to improve the detection of LTR-RTs. Results above were merged and filtered by retrieving sequencing for BLASTx against the REPbase peptide database (http://www.girinst.org/repbase/) to remove redundancy of sequences in libraries (cut-off: 80%). Finally, libraries above were used as input for RepeatMasker packages (http://www.repeatmasker.org/) to detect repeats from X-linked regions. We also identified the microsatellites through MISA (MIcroSatellite identification tools) (http://pgrc.ipk-gatersleben.de/misa/) program with following settings: a length greater than 30 nucleotides, motif lengths of 2 to 6 bp, and a minimum of 5 repeats The microsatellites identified were further classified into class I (>= 20bp) and class II (size >=12bp and <20bp) (Temnykh et al., 2001).To compare the variation of sex-linked regions and the rest of the genome, we also perform the same detection of repeats in the whole genome, published Y-linked BAC-BAC sequences (Kudoh et al., 2018b).

Synteny analysis of X-linked scaffold(s)

The local alignment between masked sequences of X-linked scaffold(s) and complete sequences of Y contigs from the draft assembly was conducted to identify colinear regions between X and Y sequences and anchor corresponded Y region to X-linked scaffold(s). The alignment was conducted by BLASTn (V2.3.4) software as default settings. Each collinear region was screened by sorting the position of subject hits and comparing its corresponded positions of query hits. The regions with at least three consecutive position hits pattern from query and subject were selected as a candidate synteny region. Also, distribution of synteny blocks was analyzed based on each candidate Y contig and Super_scaffold 66 through NUCmer and visualized by mummerplot from MUMmer software (Kurtz et al., 2004).

Gene content analysis of X-linked scaffold(s)

The genomic and protein sequences from the X-linked region were retrieved from SpinachBase (http://www.spinachbase.org/cgi-bin/spinach/tool/download.cgi) (Xu et al., 2017a). Protein and DNA sequences were submitted to NCBI conserved domain structure (CCD) database (https:// www.ncbi.nlm.nih.gov/Structure/cdd/wrpsb.cgi), NCBI non-redundant protein (nr) database to search conserved domain and orthologous hits respectively. To further understand the duplication of genes from X-linked region, we performed the global alignment of genomic sequences of X-linked genes against the genome assembly from XX and YY genotype respectively to search paralogous genes or homologous genes. The threshold for homologous genes were set as follow: e – value > 10e-3, identity > 95%, alignment coverage > 50%. The mismatches (SNP), gaps (INDELs) for each alignment were count. In addition, the single-copy genes from Super_scaffold 66 were aligned to corresponded Y contigs to identify structural variations between each homologous gene pair, the numbers of SNP, gaps (open) were summarized for further analysis.

Development of sex-linked markers and linkage analysis

Insertions and Deletions (INDELs) are informative SVs for PCR and gel electrophoresis. INDELs identified from synteny analysis and INDELs (>20bp) from global alignments were combined as candidate regions for marker development. Complete genomic sequences from those candidate regions were retrieved then carefully screened by pairwise local alignment tool as follow: Each correspond sequences from the reference genome and Y contigs were submitted to EMBOSS WATER alignment tool with ‘Smith and waterman algorithm’ (https://www.ebi.ac.uk /Tools/psa/emboss_water/nucleotide.html) (Rice et al., 2000). INDELs with more than 20bp and gaps (open) were selected as template region for primer design and those regions were screened through Primer-Blast (https://www.ncbi.nlm.nih.gov/ tools/primer-blast/). Parameter for primer design was selected as follow: primer length: 20-24bp; product size: 300-1000bp; Tm: 56~63C; database (reference: nr database): spinach genome (*Spinacia olracea TaxID:3562*). Each primer pair were tested by two parents’ DNA from two respective population through PCR and electrophoresis. 2× Gotaq master mix was used to conduct PCR with 10µL reaction system as protocol along with the workflow as follow: Initial Denaturation 95°C for 2 minutes, 40 cycles of the 3 consecutive steps are followed: Denaturation 95°C for 30 seconds, annealing (55-60°C) for 30 seconds with 40 cycles, Extension 72°C for 60 seconds. Then extended in 72°C for 7 minutes then soaked at 4°C. Finally, the PCR products were tested by 1.5% agarose gel with 1× TAE Buffer. The positive markers which featured by distinguished variations through agarose gel between male and female phenotypes were further tested by the F_1 population and 40 spinach accessions.

Results

Plant materials and genome sequencing

A cross between a female (PI 217425) and a male spinach (PI 347812) generated a total of 181 segregating pseudo-test spinach individuals with 86 males and 95 females, matched the expected 1:1 segregation ratio (Chi-square test: $P > 0.05$). Seeds of 40 spinach accessions from USDA germplasm generated 423 samples, comprised of 266 female and 245 male individuals (Table 2.1). The pooled samples of males and females for each accession with three replicates generated 240 sets DNA samples. 181 DNA samples were extracted individually from the pseudo-test population. For YY spinach sequencing, an approximately 53G raw reads was generated through PacBio sequencing and a total of 813,320,575bp YY genome assembly was generated through CANU pipelines (https://canu.readthedocs.io/en), which comprised of 22,658 small contigs, ranging from 1473bp – 825271bp with N50 at 42Kb and N90 at 19Kb.

Identification of a sex-linked scaffold by SpoX

The global sequence alignment of three sex co-segregating markers include SpoX, T11A, and V20A against the reference genome assembly of spinach generated thousands of hits while one unique hit with 100% identity and 100% coverage (e-value = 0). This is the only high-fidelity hit from the reference genome (Table 2.2). The hit located on a 1,780,654bp scaffold (Super_scaffold 66) with only nine predicted genes with gene density at 197.78Kb/gene, which is significantly lower than gene density from genome-wide average at 39.22Kb/gene (P = 1.06e-5). Gene distributed unevenly across the entire scaffold and there was no gene identified from the initial 1.4Mb consecutive region of Super_scaffold 66. The scarcity of genes attributed to the repetitive sequence (51.58%) and genome assembly gaps (38.05%) (Table 2.3). Among all types of repetitive sequences, 39.30% of repeats are *Gypsy* and *Copia* retrotransposons. The remaining repetitive sequences are mostly Type II DNA transposable elements and unclassified repeat units. There were only 0.14 % microsatellite units identified along with lower density distribution compared with whole-genome average but similar with Y-linked BAC-BAC sequences (Table 2.4). Thus, Super_scaffold 66 is closely surround sex-determining loci, because of the reduced gene density, wide-spread repetitive elements, and existence of X-specific marker 'SpoX' within this region.

Gene content of X-linked scaffold(s)

Among the nine genes from Super_scaffold 66, conserved domains were identified from eight of them and the remaining one gene might be pseudogene or annotation with low fidelity (Table 2.5). Six out of eight genes with a < 1e -10 E-value were used for gene duplication analysis. Detection of the duplications in reference genome classified three genes (*Spo14028*, *Spo14029*, and *Spo14032*) as unique genes and the other three genes as duplicated genes in XX genome (cutoff: identity > 80% and coverage > 50%). For the three duplicated genes, *Spo14027* is highly conserved with no SNP or INDEL while some noticeable SVs in *Spo14030* and *Spo14031* were identified when comparing with their paralogs in the genome (Table 2.6). Further, we performed the identification of homologs of three duplicated genes from YY genome. *Spo14030* and *Spo14031* are sharing decreased similarities among respective homologs, whereas *Spo14027* is more conserved (Table 2.7). Interestingly, *Spo14027* has almost two times more duplications in YY genotype, but the type of duplication cannot be determined due to unknow genomic position of those Y contigs.

Comparison between Y contigs and X-linked scaffold

Synteny analysis was conducted by screening sequence alignment hits from Super_scaffold 66 and Y contigs. Our manual screen revealed an accumulative 343,742bp syntenic regions consisted of nine Y contigs (Fig 2.1). Among those blocks, 13 gaps were greater than 2,000bp. Tandem duplications/ segmental duplications were widely distributed around the region. For instance, several tandem repeats greater than 1,000bp were identified from tig00028343, tig00013253, and tig00004922. The further micro-synteny analysis indicated the common distribution of adjacent or flanking segmental repeats and some regions were featured by divergent genomic structure between X and Y sequences, mainly contributed to INDELs and small-scale inversions (Fig 2.2). Comparison between Super_Scaffold 66 and Y contigs from gene structure level revealed incomplete gene structure of two Y-linked genes (Table 2.7), We found a 5Kb deletions between *Spo14028* and *Spo14029* on Y-linked regions along with partial gene conversions shared by two Y-linked genes (Fig 2.3). Also, a 4,938bp LTR-RT was identified within the intron region of a Y-linked gene compared with *Spo14028* from X counterpart. The structure of this LTR-RT was further dissected by separating the two LTRs, two TSDs, and encoded region of retro-elements. The 'TG' and 'CA' codons were identified from each terminal with a 99.31% similarity of two flanking LTRs. The internal part consisted of four intact domains includes reverse transcriptase, *Copia* family RNase, H integrase and *gag* integrase (Fig 2.4). The re-structured exon-intron relations of neighboring genes and retro-elements insertions could potentially disrupt the intact gene function.

Sex co-segregating markers and linkage analysis

DNA markers amplified from 12 pairs of primers co-segregated with sex phenotype of dioecious spinach (Fig 2.5), ranging from 174-654 bp from Y-specific regions (Table 2.8). The segregation pattern was confirmed from the 181 pseudo-test cross population (Viroflay × Cornell-NO.9). Six out of 12 markers are co-dominant markers which reveal different fragment size from 30bp to 700bp between Y sequences and X counterparts (Sp20, Sp21, Sp37, X07, X12, and X20), and the other six markers are Y-specific markers, which amplified from male individuals only (Sp28, Sp38, Sp39, X16, C20, and C25). Sex genotyping using these 12 markers on 266 female and 245 male plants from 40 accessions (120 sets DNA samples with 3 replicates for each) reached almost 100% co-segregation, indicating no recombination in this region across diverse germplasm. This suppressed recombination with sex co-segregation indicates that Super_scaffold 66 is part of the non-recombining region of the sex chromosomes in spinach.

Discussion

Recombination suppression could be identified in chromosomal regions in higher plants accompanying with the protection of haplotype for selective advantages, including sex determination (Löve, 1944), apomixes (Akiyama et al., 2004), and self-incompatibility (Casselman et al., 2000; Lukacsovich & Waldman, 1999; Wang et al., 2003). Therefore, identification of sex co-segregating markers is not only used for marker-assisted selection in breeding programs but also serves as a robust evidence to estimate the size of the region with suppressed recombination. Up to date, detection of sex co-segregating variants was performed in different dioecious species to estimate the degree of suppressed recombination surround sex-determining loci, such as 225 sex-linked AFLP markers spanning in 8.1 Mb of papaya MSY region, an approximate 100Kb non-recombining region from species under *Poplus* genus, and a 2.5Mb sex-linked interval from *Salix* genus (Dai et al., 2014; Ma et al., 2004; Pucholt et al., 2015). The 12 sex co-segregating DNA markers we developed distributed across the Super_scaffold 66 (1.78Mb). Hence, it is plausible to propose that Super_scaffold 66 closely resides in sex-determining region and the X counterpart of SDR is at least to be around 1.7Mb. However, the degree of ceased recombination could be beyond the Super_scaffold 66. Thus, a genome-wide detection of sex co-segregating makers is needed to estimate the extent of recombination suppression and genomic size of SDR in future studies.

As one of the primary steps to characterize SVs and estimate the scale of divergence between X and Y chromosomes, synteny analysis provides the framework for dissecting the genomic basis of SDR. Aligning Y contigs to Super_scaffold 66 revealed an uneven distribution of synteny blocks across Super_scaffold 66. Specifically, only nine contigs aligned to 24% genomic region of Super_scaffold 66 but the rest scaffold (1.3Mb) was poorly represented by synteny block analysis. Moreover, a 24% region of Super_scaffold 66 (approximate 450Kb) genomic region on Super_scaffold 66 corresponds to a reduced 335Kb Y corresponded part from nine Y contigs (Fig 2.1). Several reasons might explain these observations: (1) The Illumina short reads reference genome was not adequate for tackling highly repetitive regions, which results in the 38.05% of gaps from Super_scaffold 66 assembly and most gaps were in this 1.1Mb genomic region. (2) The 24% region on Super_scaffold 66 without matching Y contigs were highly repetitive to anchor corresponded contigs. (3) The SVs between X and Y chromosomes could be at a substantially higher frequency and the Y contigs from the assembly were not long enough to cover X counterpart attributed to the short reads length and the nature of repetitive sequences in SDR of Y chromosome.

Retrotransposon mediated genomic rearrangements occur more frequently after recombination ceased from regions surrounding the sex-determining loci. More than 11 chromosomal rearrangements each occurred in inversions 1 and 2 of the MSY in papaya (Wang et al., 2012). A small 258Kb insertion of MSY from Medaka fish and 706Kb insertions female-specific W chromosome (FSW) were detected (Kondo et al., 2006; Yin et al., 2008). Also, the X-autosome fusions were found from *Humulus japonicus*, several *Rumex* species, and *Viscum fischer* (Charlesworth, 2016). In our studies, the small tandem duplications and inverted repeats are ubiquitous along with high similarities for each duplication from Y contigs such as tig00013253, tig0008443, and tig00004962 (Fig 2.1). Further comparison of each synteny block indicated high frequencies of small INDELs, particularly as identified from tig00013251 and tig00028343, but large-scale inversion or insertions are not identified or could not be identified due to the fragmented nature of the reference genome and the Y sequences. Those SVs are in line with the expectation of a pair of nascent sex chromosomes with YY viability identified in some spinach accessions.

In dioecious plants, Y chromosomes could be larger than X chromosomes as an outcome of expanded MSY through a pronounced accumulation of repeat elements and mutations within the non-recombining region (Charlesworth et al., 1994; Na et al., 2014). This fact can be explained by the four-time (three-time) reduced effective population size of genes on the Y chromosome compared with autosomes (X chromosome), which is more susceptible for random mutations (Charlesworth & Charlesworth, 2000). The *S. latifolia* Y chromosome is 40% larger than the X chromosome due to accumulation of repetitive sequences, mostly retrotransposon insertions, as well as the 8.1Mb of HSY compared with the 3.5 Mb X counterpart in papaya (Matsunaga et al., 1999; Wang et al., 2012). As partial X counterpart of Y-linked region, Super_scaffold 66 revealed only 51.58% repetitive sequences by RepeatMasker, which could be explained by the modest assembly quality that approximates 38% of gaps remained in assembly. Thus, we could infer that Super_scaffold 66 is highly repetitive, and those remaining gaps are the reason that the corresponding sequence of the Y contigs fail to be matched to corresponded position. The actual sequence of the Y corresponded to part of Super_scaffold 66 is expected to be longer as discovered in other nascent sex chromosomes in plants as consequences of repetitive sequences insertions.

Dissecting gene features is crucial to understand the evolution of sex chromosomes by characterizing the gene duplications and divergence of X/Y genes near sex-determining loci (Yu et al., 2008a). The age of sex chromosomes could be studied by either estimating the rate of synonymous mutation among X and Y gene pairs based on molecular clocks or detecting the degree of gene structure variations from MSY. In our studies, full-length genomic sequences of nine genes from Super_scaffold 66 provided an opportunity to compare SVs of genes from corresponded Y contigs in terms of sequence identity and coverage. The estimate of divergence time based on synonymous rate of genes might be biased as there were only nine genes. However, the two neighboring genes from AAA superfamily domain-encoded genes revealed notable features on Y-linked region, as a recent genic insertion of LTR-RTs between introns and redistributed exon-intron structure from the other. Those interrupted Y-linked gene structures could potentially influence splicing or transcription and further impact gene function. Therefore, two Y-linked genes might be functionally diversified or even degraded compared with X alleles, which needs to be experimentally confirmed in subsequent studies.

Conclusion

A 1.78 Mb sex-linked scaffold from spinach genome was identified by developing 12 sex co-segregating DNA markers based on SVs between sequences from X and Y chromosome. Those markers were further validated in 86 males and 95 females of an F_1 mapping population and 266 female and 245 male plants from 40 spinach accessions. Genomic analysis revealed the enriched repetitive sequences, scarcity of genes, and exon-intron structural variations between X and Y alleles within this sex co-segregating scaffold, which could be the consequences of recombination suppression near sex-determining loci. We proposed this region as a part of X chromosome, which corresponded to SDR on the Y chromosome.

Tables and figures

Table 2.1 Spinach accessions from USDA for marker analysis

Accession Origin	Accession ID	Female No.	Male No.	Total No.
Turkey	PI171867	4	4	8
	PI177865	4	7	11
	PI177864	5	8	13
United States	NSL184398	7	8	15
	NSL184380	4	5	9
	NSL184379	7	6	13
Belgium	PI179596	9	8	17
	PI179595	6	10	16
India	PI165504	8	8	16
	PI174960	6	10	16
Afghanistan	PI165504	4	9	13
	PI677108	5	5	10
	PI604783	4	9	13
France	PI261789	7	5	12
	PI261788	4	6	10
Spain	PI262161	5	4	9
China	PI648945	6	4	10
	PI648946	9	8	17
	PI648947	8	6	14
	PI648948	8	6	14
	PI648949	6	8	14
Japan	PI604780	6	7	13
	PI604779	6	5	11
Greece	PI491262	8	7	15
	PI648936	5	7	11
Hungary	PI531453	8	8	16
	PI531455	4	6	10
Iran	PI227045	7	6	13
	PI227383	6	8	14
	PI229731	7	6	13
	PI229792	8	5	13
Georgia	PI647853	5	6	11
	PI647854	7	4	11
	PI647855	5	5	10
Egypt	PI319220	6	8	14
Sounth Korea	PI508504	5	7	12
	PI217425	5	6	11
Fomer Soviet	PI499372	9	7	16
Syria	PI445785	6	9	15
	PI181808	6	5	11

Table 2.2 Summary of sequence alignment of three sex co-segregating markers

Marker ID	Hits NO.	Hits identities	Hits Coverage	Hits position
T11A	3	91.23-97.83	0-77	SpoScf 01264
	3	95.95-97.09	5-64	SpoScf_0128
	3	96.70 – 98.01	0-76	Super_scaffold
V20A	68	79.50-98.05	0-65	chr1
	79	91.43-99.05	1-54	chr3
	654	80.35-96.42	0-76	Rest scaffolds
SpoX	1	100	100	Super_scaffold 66

Table 2.3 Repeat elements distribution in Super_scaffold 66

Classification	types	numbers	length	percentage
SINEs:		0	0 bp	0.00%
	ALUs	0	0 bp	0.00%
	MIRs	0	0 bp	0.00%
	LINEs:	34	21275 bp	1.21%
	LINE1	30	20563 bp	1.17%
	LINE2	0	0 bp	0.00%
	L3/CR1	0	0 bp	0.00%
LTR elements:	*Copia/Gypsy*	457	688317 bp	39.30%
	ERVL	0	0 bp	0.00%
	ERVL-MaLRs	0	0 bp	0.00%
	ERV_classI	3	274 bp	0.02%
	ERV_classII	0	0 bp	0.00%
DNA elements:		45	36866 bp	2.10%
	hAT-Charlie	0	0 bp	0.00%
	TcMar-Tigger	0	0 bp	0.00%
Unclassified:		347	203818 bp	11.64%

Table 2.4 Distribution of microsatellites from different sequences

Sequences	Size (Mb)	Class I		Class II		Total	
		Number	Density (Kb/per)	Number	Density (Kb/per)	Number	Density (Kb/per)
Super_scaffold 66	1.72	38	72.98	116	28.25	154	10.39
Y-linked region	0.45	4	112.50	14	32.14	18	25.00
spinach genome	997.63	41514	24.03	69958	14.26	111472	8.95

Table 2.5 Information of protein-encoded genes from Super_scaffold 66

Gene ID	Description	Conserved domain structure	E -value
Spo14032	Inosine-uridine preferring nucleoside hydrolase	Nucleoside hydrolases Superfamily	0e+00
Spo14031	putative lipid-binding protein AIR1	(HPS)-like subfamily	1.32e-16
Spo14027	DEAD-box ATP-dependent	DEAD-like helicases superfamily	7.12e-30
Spo14026	Uncharacterized proteins	NA	NA
Spo14025	hypothetical protein SOVF_164990	AdoMet_MTases super family	4.79e-04
Spo14030	hypothetical protein SOVF_070170	Thioredoxin_like super family	2.43e-15
Spo14029	pleiotropic drug resistance protein 2-like	AAA super family	5.59e-52
Spo14028	ABC transporter G family member	AAA super family	0e+00
Spo14023	Uncharacterized proteins	RVT_2 super family	4.50e-06

Table 2.6 Duplicated X-linked genes identified from reference genome

Gene ID	position	Identity	Coverage	Length	Mismatch	Gap	E-value
Spo14027	SpoScf_02174	100	100	2192	0	0	0
	SpoScf_00987	100	100	2192	0	0	0
	Super_scaffold_66	100	100	2192	0	0	0
	chr2	100	100	2192	0	0	0
Spo14030	chr2	92.52	100	588	25	6	0
	chr2	90.32	100	589	31	11	0
	Super_scaffold_66	100	100	577	0	0	0
Spo14031	Super_scaffold_66	100	100	336	0	0	2E-176
	SpoScf_51038	97.31	100	335	9	0	9E-161
	chr4	97.29	99	332	9	0	4E-159

Table 2.7 Duplicated Y-linked genes identified from YY genome

Gene ID	position	identity	coverage	length	mismatch	gap	e - value
Spo14027	tig00030865	100	100	2192	0	0	0
	tig00030864	100	100	2192	0	0	0
	tig00016404	100	100	2192	0	0	0
	tig00015783	100	100	2192	0	0	0
	tig00012982	100	100	2192	0	0	0
	tig00015782	99.91	100	2192	2	0	0
	tig00027787	99.59	100	2192	0	1	0
Spo14031	tig00021345	98.81	100	336	2	2	3.00E-169
	tig00034629	97.01	100	335	9	1	1.00E-158
	tig00025509	96.72	100	335	11	0	2.00E-157
	tig00002340	97.89	99	332	7	0	1.00E-162
Spo14030	tig00006202	92.52	100	588	25	6	0
	tig00012111	90.32	100	589	31	11	0
	tig00013251	98.61	100	577	8	0	0

Table 2.8 Primer sequences of 12 sex co-segregating DNA markers

ID	Forward primer	Reverse primer	Product (bp)	Tm
X07	CCAATTCCAATTTGCTATGGC	GAGATCATCTCATAGGTATGAG	174	59.0
X12	TGTCTGATCTTCCAACTTCCA	GGGAAGAGCACCAATTGTTTTT	283	58.5
X16	ATCTTTTCTACTCGCTCAGGA	GTAGAACTTACGTTCATGCTCT	859	54.6
X20	TGTCTGATCTTCCAACTTCCA	GGGAAGAGCACCAATTGTTTTT	260	56.0
C20	TGGTGTTCATTCAAAGAGGGA	TAGATTTTAGACTCGGGCGATA	389	56.5
C25	TGAGGCAGAATATCGATCGAT	TAACCGGGAAGGTGACATATTG	829	56.5
Sp20	GTTAAACCCCAACCTCCATGC	TTGGGTTCTGGTCGATTTACGG	265	55.5
Sp21	TCAGATAGGGTTGACGAGTAT	GCCTGCTAATTAACCCGTGC	244	56.5
Sp28	AACCCATCAACCTGACCTAGC	GATGGTATAGTGTGAAACGCAT	515	56.0
Sp37	CTGTTGCTACGTACGTGTGG	CCTGGGGTGTAGGGTCAAAA	195	56.0
Sp38	GGTTCTGTTTTTCTATCTGTCT	GAGGGTTTCCAAAAACCAAAA	654	56.0
Sp39	AACCCATTGTCGAACACCCA	AGAGTGCAAGTGGCCGAAAT	688	56.5

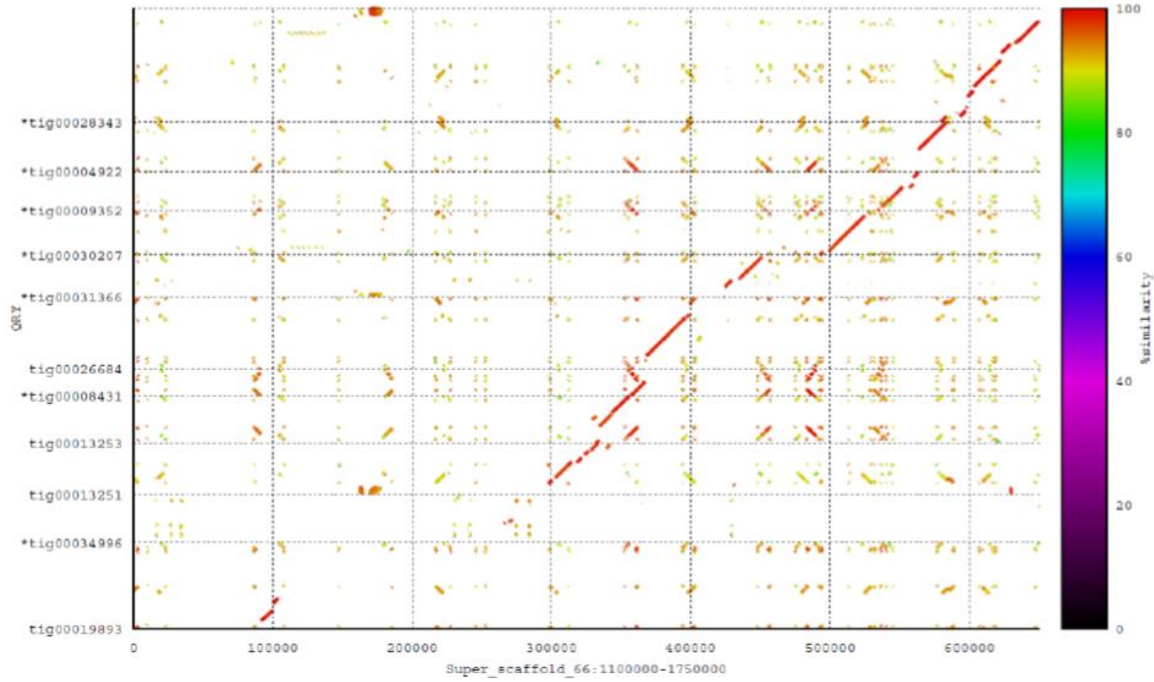

Fig 2.1 Matrix plot between Super_scaffold 66 and Y contigs

The matrix plot represents the synteny of partial Super_scaffold 66 (1.1Mb – 1.75Mb) and corresponded Y contigs, similarities were identified from each segment were corresponded to the color scale on right side.

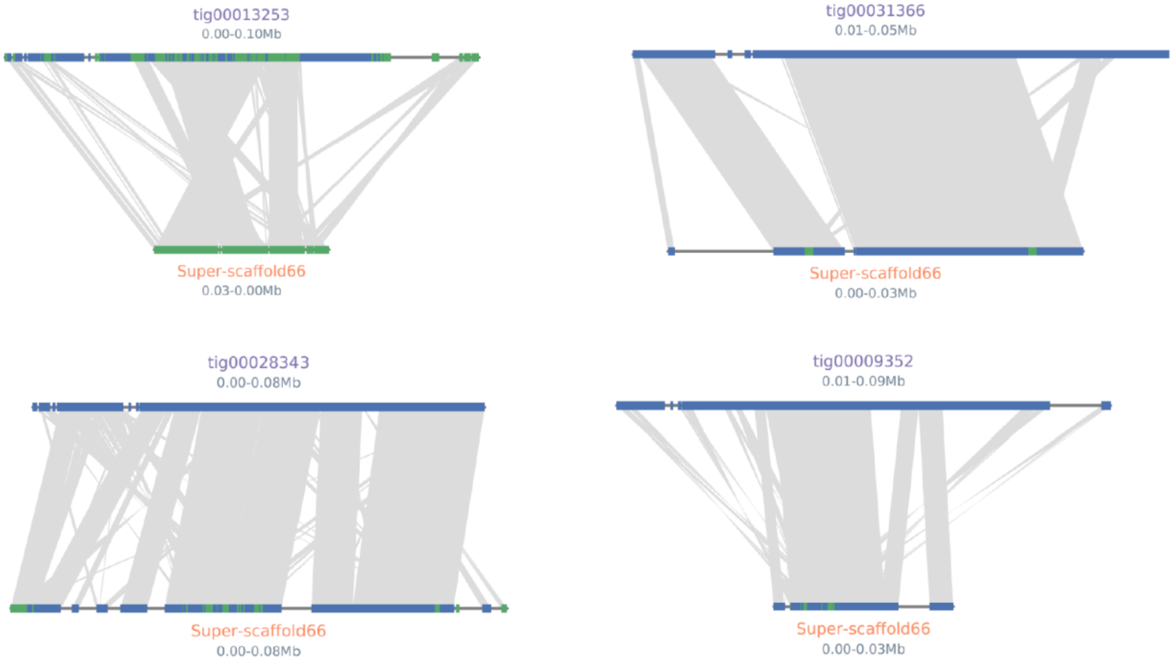

Fig 2.2 Micro-synteny analysis between Y contigs and Super_scaffold 66

Syntenic region between Super_Scaffold 66 and respective Y contigs from was illustrated by blue and green tabs, syntenic relations were connected by grey shades.

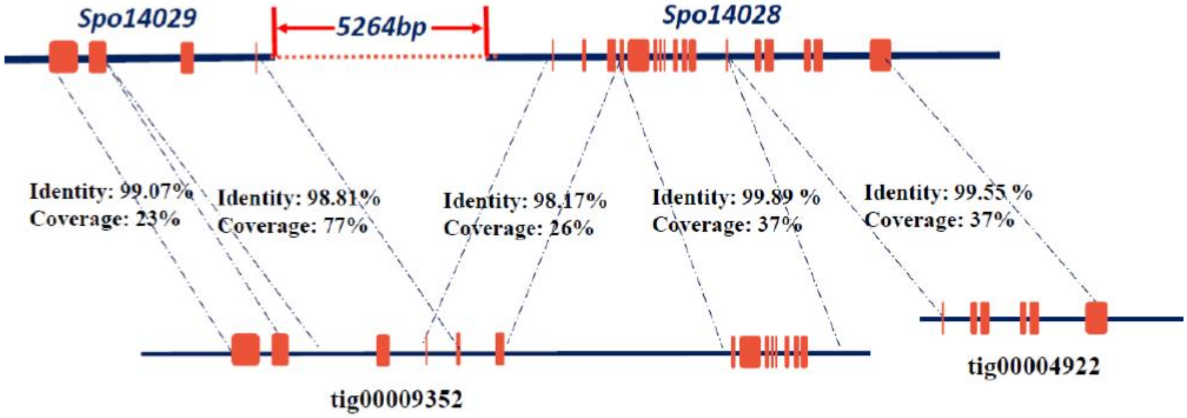

Fig 2.3 Genomic rearrangement of two X-linked genes from Super_scaffold 66

Exon-intron structure of two X-linked genes (*Spo14029* and *Spo14028*) and respective Y corresponded part was marked in orange boxes. An exon-intron rearrangement and a genomic insertion was identified on tig00009352.

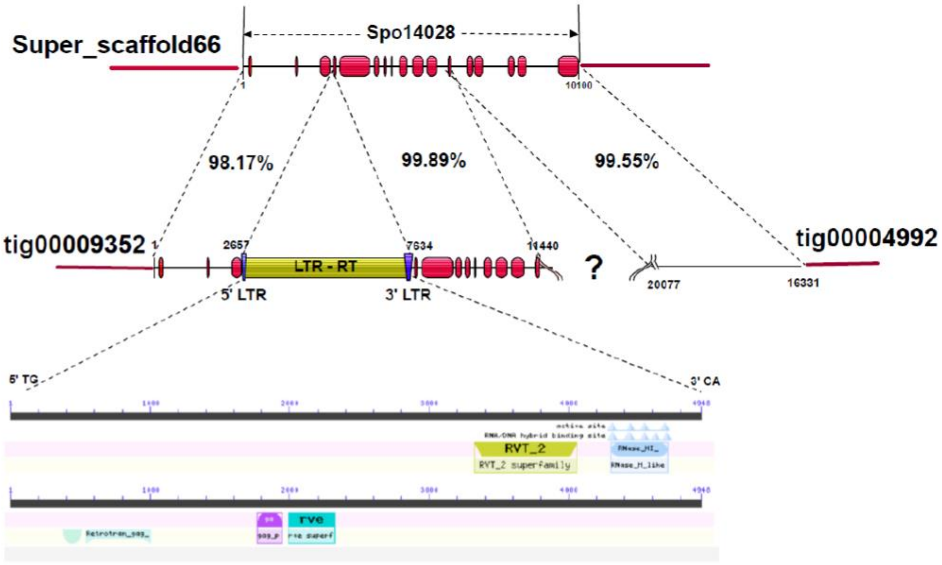

Fig 2.4 An LTR-RT insertion identified from Y contig

The 18 exons of gene *Spo14028* from spinach genome were plotted by red shaded boxes (length: 10100bp) and compared with tig00009352, identities of sequences were marked based on alignment from BLASTn. A genic LTR-RT insertion on Y corresponded part of *Spo14028* was marked, includes two LTRs, two TSDs, and four domains (*gag, RVT_2, RNaseH and rve*).

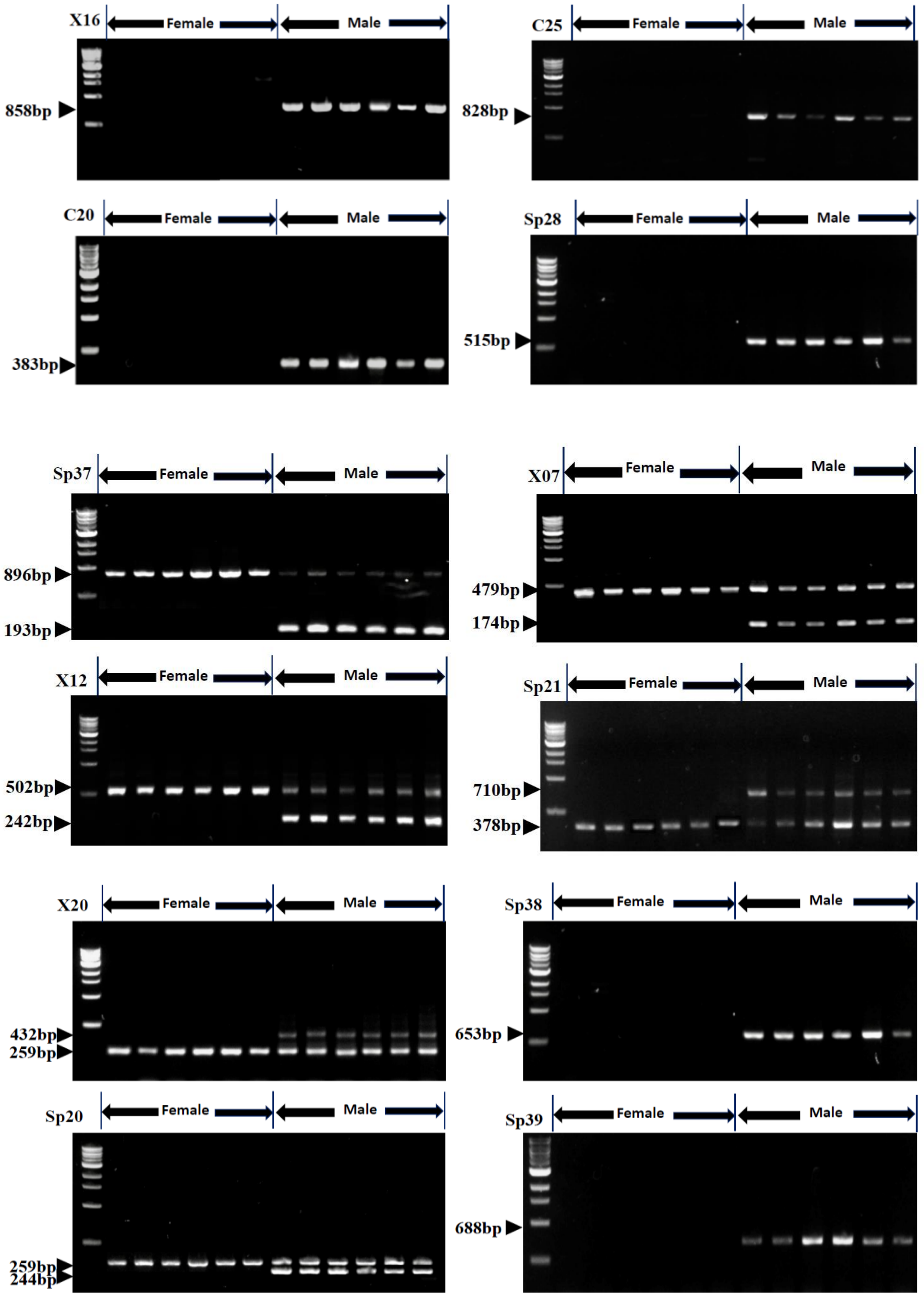

Fig 2.5 Sex co-segregating DNA marker from Super_scaffold 66

A total of 12 sex co-segregating markers were tested on 1.5% agarose gel with 6 male and 6 female spinach. The size of markers was measured by 1Kb DNA ladder.

Chapter III. Construction of high-density genetic maps defined sex determination region of the Y chromosome in spinach

Abstract

Spinach (*Spinacia olracea* L.,) is a dioecious leafy vegetable with a highly-repetitive genome of around 990 Mb, which is challenging for *De-novo* genome assembly. In our study, a segregating F_1 (double pseudo-testcross) population from ‘Viroflay’ × ‘Cornell-NO. 9’ was used for genetic mapping by re-sequencing genotyping. In the paternal ‘Cornell-NO.9’ map, 212,414 SNPs were mapped, and the total linkage distance was 476.83cM; the maternal ‘Viroflay’ map included 29,282 SNPs with 401.28cM total genetic distance. Both paternal and maternal maps have the expected number of six linkage groups (LGs). A non-recombining region with 5,678 SNPs (39 bin markers) co-segregates with sex type which located at 45.2 cM of LG1 in the ‘Cornell-NO.9’ map while indicates the sex determination region (SDR). Integration of two maps into a consensus map guided us to anchor an additional 1,242 contigs to six pseudomolecules from the published reference genome, which improved an additional 233Mb (23.4%) assembly based on spinach estimated genome size. Particularly, the X counterpart of SDR in our assembly is estimated around 18.4Mb which locates at the largest chromosome, as consensus with sex-biased FISH signals from previous cytogenetics studies. The region is featured by reduced gene density, higher percentage of repetitive sequences, and no recombination. Our linkage maps provide the resource for improving spinach genome *De-novo* assembly and identification of sex-determining genes in spinach.

Key words: Genome assembly, high-density genetic map, spinach, sex determination

Introduction

Linkage maps are one of the most effective and essential tools to improve genome assembly based on the degree of linkage of markers from different genomic positions (Consortium, 2012). In addition, linkage mapping deciphers the relationship between the molecular causes and phenotypic variation, which can be used to identify a genomic region underlying the traits (Grattapaglia & Sederoff, 1994; Mackay, 2001). F_1 segregating populations, known as "double pseudo-testcross", can be used to construct separate genetic maps from the two parents (Grattapaglia & Sederoff, 1994), which has been widely employed in woody plants, dioecious plants, and plants with self-incompatibility, such as apple (*Malus domestica*), Chinese white pear (*Pyrus × bretschneideri*), kiwifruit (*Actinidia chinensis*), and pineapple (*Ananas. comosus*) (de Sousa et al., 2013; Ming et al., 2015; Wu et al., 2014; Zhang et al., 2015).

Genotyping has traditionally been expensive and labor-intensive via PCR and gel electrophoresis assays. Recent advances in DNA sequencing technologies have empowered detecting large numbers of DNA markers with a rapidly falling cost. Restriction site-associated DNA markers (RAD-seq) and specific-locus amplified fragment sequencing (SLAF-seq) can produce tens of thousands of markers by capturing restriction site-associated DNA sequences at a low cost, as known as genotyping by sequencing (GBS) (Elshire et al., 2011; Peterson et al., 2012; Sun et al., 2013). However, the uneven distribution of restriction sites in plant genomes introduces gaps in the maps. Fortunately, the decreased cost of whole-genome shotgun (WGS) sequencing enables genotyping by resequencing to become affordable, which largely improves the capability of detecting numerous polymorphisms. This has been widely used for linkage mapping, yielded high-density genetic maps with several hundreds of thousands of SNPs or even millions of markers in several crops and a variety of population panels (Bowers et al., 2016; Chapman et al., 2015; Hahn et al., 2014).

The cultivated spinach (*Spinacea olracea*, 2n = 12) is a leafy vegetable with enriched vitamins and nutrients, it is widely cultivated in Asia, North America, and Europe area with an approximate annual production of 26.7 million tonnes worldwide (FAOSTAT, http://www.fao.org/faostat/en/#data/QC). Conventional spinach breeding was enhanced by molecular-based breeding resources, includes molecular markers, genetic maps, and whole-genome sequences (Ashraf & Foolad, 2013; Li et al., 2018; Xu et al., 2017a). The published spinach reference genome by Illumina sequencing yielded a 985 Mb assembly which is close to the estimated genome size by flow cytometry (989Mb) (Arumuganathan & Earle, 1991; Xu et al., 2017a). The assembly was further anchored to six pseudomolecules and matches its haplotype chromosome numbers from cytogenetics studies (Ellis & Janick, 1960; Janick et al., 1959). However, the chromosomal assembly only represented 465Mb (46.5%) of the genome due to numerous small-fragmented contigs or scaffolds and insufficient marker density to anchor those scaffolds. Therefore, better genetic maps will facilitate the improvement of spinach genome assembly for genomic studies.

Spinach is an annual dioecious (or occasional monoecy) plants with a short life cycle and its genetic sex determination is of interest (Ellis & Janick, 1960; Janick & Stevenson, 1955b). In flowering plants, hermaphrodites account for 89% of species while the rests are dioecious (6%) or monoecious (5%). The emergence of dioecious plants is closely related to sex determination. Specifically, sex determination in some species is controlled by a pair of sex chromosomes, as identified from 37 species of 16 families from angiosperms (Ming et al., 2011b; Renner, 2014b). Empirical 'two-factor' model proposed that sex chromosomes evolved from autosomes (Charlesworth et al., 2005). In step one, one recessive mutation (loss-of-function) encoded for male sterility and the other gain-of-function mutation for carpel suppression occurred. Followed by suppression of recombination via chromosomal rearrangements which form into a small non-recombing region along with YY viability. Further, the ceased recombination spreads to rest of the chromosome along with expanded Y chromosome by retrotransposons insertions and gene degradation, leads to YY lethality. Those accumulative mutations finally induce the shrinkage and loss of the Y chromosome.

In spinach, cytogenetic analysis has identified a pair of homomorphic XY sex chromosomes by detection of sex-biased signals from hybridization of 45S rDNA or DNA sequence to the largest pair of metacentric chromosomes between male and female spinach (Deng et al., 2013a; Deng et al., 2012; Lan et al., 2006). Those slight heteromorphs in between XY chromosomes and YY viability suggested that the Y-linked region is evolutionarily nascent (Wadlington & Ming, 2018b). The first spinach genetic map derived from microsatellites and ALFP markers generated seven linkage groups and a maker SO4 is 1.9cM distant to sex-determining loci at the largest linkage group, corresponded to the cytogenetics evidence (Khattak et al., 2006). Further, DNA microdissection of the Y chromosome expedited the identification of two sex-linked markers includes marker T11A and V20A (Akamatsu & Suzuki, 1999) and these two markers were encompassed by additional 10 RLFP markers in a13.4cM genetic region defined by linkage mapping (Takahata et al., 2016). The more recent SLAF-seq linkage mapping produced six linkage groups by 4,080 markers with 1,125.97cM distance and an average distance of 0.31cM between adjacent loci (Qian et al., 2017). Combined with bulk segregant analysis (BSA), two intervals (66.98 cM–69.72 cM and 75.48 cM–92.96 cM) from the largest group were proposed as candidate region for sex-determining loci. In addition, construction of BAC-BAC libraries of spinach DNA dissected 500Kb Y-linked non-recombining regions of the Y chromosome along with several Y-specific DNA makers, which provided the evidence of recombination suppression of Y-linked region (Kudoh et al., 2018b). However, neither the genomic position of non-recombining SDR, nor its size or the numbers of sex-linked genes from X/Y chromosomes were clear.

To improve spinach genome assembly and address questions above, we constructed two linkage maps from “Viroflay” × “Cornell-NO.9” cross using genome resequencing. As spinach is dioecious, we used F_1 progeny in a pseudo-test cross linkage mapping approach. Separate maps were constructed for male and female parents by scoring and merging segregating SNPs into bin markers, then grouped into chromosomes and ordered within chromosomes based on their segregation pattern. We further identified sequences with sex co-segregating bins as SDR and anchored remaining contigs to the six pseudomolecules from the published genome, to yield an improved genome assembly with X counterpart of the SDR defined. Our integrated maps and improved genome assembly can be used for further genomic studies, molecular breeding, and discovery of sex determination genes in spinach.

Materials and methods

Plant materials

One F_1 mapping population of dioecious spinach was generated from a cross ‘Viroflay’ (PI 217425) × ‘Cornell-NO.9’ (PI 347812). Plants were maintained in 20-22°C environment in a greenhouse at the University of Illinois at Urbana - Champaign for three months. Identification of sex type for each plant was based on flower morphology with three replications at different flower developmental stages to prevent false positives. Fresh leaves of each F_1 individual were collected and treated with liquid nitrogen and stored at -80 °C. DNA extraction was followed by CTAB method with 1% Edward Buffer and each DNA was separated by 1% agarose gel electrophoresis and quantified by Nanodrop spectrophotometer. The final DNA samples were diluted to 50ng/μL for quality check as DNA templates for sequencing.

WGS library construction and sequencing

The genomic DNA from two parents ('Viroflay' and 'Cornell-NO.9') and F_1 individuals were prepared for quality control before library construction. Briefly, all DNA samples were tested by gel electrophoresis and Qubit, and passed samples were prepared for sequencing by construction of paired-end libraries with an average insert size of 150bp using Genomic DNA Sample Prep kit and the Nextera Mate Pair Sample Preparation Kit (Illumina, San Diego, CA, USA). Following the instruction manual of Illumina NovaSeq. Each F_1 individual was sequenced at the depth about 6× and each parent was sequenced at 10× depth based on spinach genome size.

SNP calling and genotyping for population

Initially, raw DNA sequencing reads were trimmed by Trimmomatic to remove low-quality reads (Bolger et al., 2014) and the spinach reference genome (http://www.spinachbase.org/) was masked by pipelines which integrated detection of repeats by Repeat Masker and RepeatModeler (https://github.com/tangerzhang/popCNV) to reduce the possibility of erroneous alignment of SNP calling by repeat elements. The clean reads were further aligned to the masked genome by using bwa-mem function (Li & Durbin, 2010) and processed with GATK variants calling pipelines (https://software.broad institute.org/gatk/best-practices/workflow) (Walker et al., 2018). Basically, duplicated mapping reads were filtered out and only unique mapped reads were kept, the Indel realignment was performed to correct false positive callings around Indels. All SNPs generated above were further filtered by several criteria: Initial filtering was conducted by VCFtools (Danecek et al., 2011) in terms of the missing data percentage, numbers of alleles and minor allele frequency (MAF) (parameter: --max-missing 0.80 --maf 0.05 --min-alleles 2 --maxalleles 2). Secondly, the loci with > 2000 total mapping depth were discarded since those high-depth alignments were mainly attributed to erroneous repetitive-sequences mapping and calling. Since more than millions of SNPs were retained, SNPs with total quality score < 1000 from the last step were removed based on previous studies and distribution of a total quality score (Bowers et al., 2012b; Bowers et al., 2016). Finally, genotype code with 'nn × np' (markers with first parents as homozygous and second as heterozygous) for 'paternal map' and 'lm × ll' (markers with first parents as heterozygous and second as homozygous) for 'maternal map' were selected for map construction.

Genotyping proofreading and linkage map construction

Before map construction, genotype for each SNP was proofread based on segregation pattern of its adjacent SNPs. Namely, mistakenly called genotypes for SNPs from the same contig could be corrected based on its flanking SNPs in a few kilobase regions by assuming a low-chance of recombination in a small genomic region (Bowers et al., 2016). Based on 'majority rule', the genotype for SNPs could be determined for that given area by counting ratios of different genotypes for each individual. From SNPs selected for two maps, small contigs with less than 5 SNPs and genomic regions with inconsistent segregation patterns within 10 SNPs were discarded. The area of incorrect genotyping associated with high percentage of missing data were manually filled based on neighboring SNPs. Further, selected SNPs in each 100Kb genomic region on the same contig were merged into one bin by BinMarkers scripts (https://github.com/lileiting/BinMarkers-v2.git). By sorting and grouping bins from two maps, bins with missing data were filled. Representative genotypes for each contig was selected by removing duplicated bins from the same contig, as methods described from previous studies (Bowers et al., 2016). Finally, linkage mapping was constructed based on selected 'clean bins' from two maps above with Lep-Map3 mapping tool (Rastas, 2017). The passed bins were separated into different linkage groups using Separatechromosome2, by parameter LOD = 8.0 and theta = 0.3. The markers from major linkage groups were assigned recursively for 3 times by OrderMarkers2 function. The final maps were presented by Excel and linkageMapView package from R.

Genetic identification of SDR on the Y chromosome

Identification of sex co-segregating markers can identify SDR, which is particularly plausible if recombination is creased within a large region containing sex-determining genes. In this study, a careful screening of sex-linked markers was conducted by steps. Initially, sex phenotype for each F_1 individual was confirmed by genotypes of the two published sex co-segregating DNA markers including T11A and V20A (Akamatsu & Suzuki, 1999). The sex-linked bin markers were identified based on consistency between genotype and sex phenotype (heterozygous in all 41 males and homozygous in the 39 females). Those sex-linked bins were further manually confirmed by retrieving component SNPs before merging into bins. Finally, all the bins with expected non-recombing linkage pattern which possess a single locus on male genetic map were proposed as SDR on the Y chromosome.

Consensus map and genome assembly

Construction of linkage maps with an F_1 population takes the advantages of mapping two sets of recombination from male and female parents in a separate manner, which provides a reliable source for genome assembly. The missing hit from published reference genome by blast of few Y-specific markers indicated that reference genome contains no Y chromosome (Kudoh et al., 2018a). Thus, integration of two maps could enable us to assembly XX genotype, which is more accessible compared with splitting two haplotypes in XY genotype. To anchor rest hundred-thousands of contigs from published spinach genome (Xu et al., 2017a), consensus map and map-based genome assembly were constructed by ALLMAPS (Tang et al., 2015). Initially, genetic positions from 'Viroflay' map and 'Cornell-NO.9' map was assigned in the same weight scale for anchoring and ordering contigs. Further, synteny between each genetic map and genomic assembly was visualized, Pearson correlation coefficient between genomic assembly and each genetic map was calculated (cut-off: $\rho = 0.5$). Based on consensus map-based genome assembly, the chromosome numbers and linkage group names were reassigned based on the size of pseudomolecules.

Genomic identification and analysis of X counterpart of SDR

Sex linkage is closely related to the rise of sex chromosomes accompanying with diverged genomic landscape compared with autosomes. Initially, the genomic interval of X counterpart of SDR was defined by screening contigs containing sex co-segregating markers from male map and retrieved from map-based assembly. In addition, the coordinates of each protein-coding gene from published genome was mapped to new assembly by Blast (cut-off: E-value:0, identitiy:100, coverage: 80%) (Xu et al., 2017a). The repeat elements in new genome was identified by protocol described in previous section. A summary includes numbers of contigs, size of assembly, numbers of genes, and total length percentage of repeat elements were compared among chromosomal assembly, X counterpart of SDR, and autosomes. To further elucidate how sex chromosome evolution shape the degree of landscape variations among X counterpart of SDR, PAR (pseudo-autosomal region), and autosomes (Au), we make the boxplot and perform statistical test among those regions in terms of gene density and repeat elements popularity. Specifically, we defined the PAR as the two flanking regions of sex chromosome excluding SDR with an approximate 153.6Mb estimated as definition from previous studies (Otto et al., 2011). The numbers of genes in each million-base pair were counted based gene annotation and total length percentage of repeats in each Mb was calculated for analysis. Finally, we compared the variations by boxplot function in R and conduct Welch two-sample T-test ($\alpha = 0.05$) to detect significant levels among different genomic regions.

Identification of segregation distortion bins from consensus map

Resequencing-based genetic maps provided sources for studying genome-wide distribution of segregation distortion as substantially improved SNP coverage compared with GBS-based linkage maps. Chi-square test was applied to each bin marker from two maps at $\alpha = 0.05$ significance level and chi-square value at df = 1 scale was transformed to -Log10 manner. Those distorted bins were plotted based on coordinates from consensus map-based genome assembly. The numbers of homozygous and heterozygous genotypes for each locus were counted to classify the types of bias for each genotype.

Results

Population construction and DNA genotyping

The cross generated an F_1 population with 39 female and 41 male plants, which matched the expected 1:1 sex segregation ratio at the $\alpha = 0.05$ significance level. Resequencing of 80 F_1 individuals generated a total of 3,151,330,530 high-quality reads (964.26 Gb) with an average of 6.12× depth of DNA reads after data trimming (Table 3.1). Two parents were sequenced at about 12× with 165,904,471 (49.77Gb) clean reads. Reads alignment and SNP calling based on the masked reference genome generated 3,229,056 SNPs and those SNPs were classified based on marker genotype code from two parents: 749,807 markers were classified as 'nn × np', 423,352 as 'lm × ll', 787,691 as 'hk × hk' and 1,268,206 comprised of the remaining SNPs identified (Table 3.2). Initially, we removed loci with more than 10% missing data among F_1 progenies and loci with either one or two missing data from two parents, which kept 2,083,244 out of 3,229,056 of SNPs (Fig 3.1a). Further, canonical removal of SNPs based on low-quality alignment scores and high-depth removed 35% of the SNPs, leaving 362,297 SNPs for the paternal map and 200,051 SNPs for the maternal map for analysis (Fig 3.1b and 3.1c).

Genotyping proofreading and linkage map construction

Manual proofreading and extraction of unreliable SNPs generated 40,324 SNPs for 'Viroflay' (female) map and 256,636 SNPs for 'Cornell-NO.9' (male) map respectively (Fig. 3.1a). The reduction of SNPs from 'Cornell-NO.9' map attributed to the incorrect genotyping pattern from several genomic regions. Huge shrinkage of 'Viroflay' map SNPs mainly attributed to the high-percentage non-segregating genotype pattern of progenies (all progenies are either heterozygous or homozygous genotypes) and genotyping errors. This could be explained by the shared ancestors between accessions of female parents and reference genome or incorrect genotype of two parents for those loci. After cleaning SNPs for the whole population, a total of 940 bins identified from 'Viroflay' map and 2490 bins from 'Cornell-NO.9' map were used for map construction. At the LOD = 8.0 level, 738 bins (29,282 SNPs) from 'Viroflay' map were assigned into six linkage group, comprised of 401.28cM genetic distance (Table 3.3a); 2,378 bins (212,414 SNPs) from 'Cornell-NO.9' map were grouped to six linkage groups along with a total of 478.83cM genetic distance (Table 3.3b). The average distance for each locus of 'Viroflay' is 2.18cM, mostly attributed to the insufficient distribution of segregating markers from some genomic regions. The ‘Cornell-NO.9’ map presented a higher resolution of 1.61cM/locus on average, which is close to the maximum resolution of 80 individuals in an F_1 population (1.25cM). Among six linkage groups, LG1 from two maps share the highest density and accumulative distance for each map, with very few gaps greater than 5cM. Two graphical maps including genotype for each bin, crossing-over break point, and numbers of SNPs in each bin were summarized in a separate worksheet.

Consensus map-based genome assembly

Integration of re-sequencing-based maternal and paternal maps facilitated the improvement of current reference genome assembly quality. A total of 3,029 bins represented by 1,736 unique contigs merged from ‘Viroflay’ and ‘Cornell-NO.9’ maps were mapped to six linkage groups, along with a final 1,661 scaffolds anchored to six pseudomolecules in corresponded to a transformed coordinates of those contigs in the new genome (Fig 3.3). This consensus map-based assembly improved 23.4% of the estimated spinach genome to chromosomal level and our final chromosomal-assembly (698,320,906 bp) represented a total 70.1% of spinach genome (Table 3.4). Comparison of consensus map-based assembly between each map revealed the highly correlated genomic position and genetic position of scaffolds (Fig 3.4), indicated the reliability of ‘Viroflay’ map (ρ=0.9467 to ρ=0.988) and ‘Cornell-NO.9’ map (ρ=0.881 to ρ=0.974). Particularly, six linkage blocks which featured by a small genetic distance associated with a large genomic region from six pseudomolecules were identified, as consequences of reduced recombination near centromeric region or other heterochromatin regions. Positions of those blocks are consistent with the types of chromosomes classified by centromere positions from previous cytogenetics studies (Lan et al., 2006). As two metacentric chromosomes represented by LG1, four submetacentric chromosomes represented by LG3 and LG5, two acrocentric chromosomes by LG6, and four sub-telocentric chromosomes by LG2 and LG4.

Identification of SDR and regions with segregation distortion

Sex co-segregating makers provide direct evidence to map SDR and sex-linked genes. In our genetic maps, scanning of bin markers and confirmation of genotypes by respective component SNPs of ‘Cornell-NO.9’ (male) map identified 46 bins with sex co-segregating genotypes at the 45.18cM from LG1 (Table 3.6), leading us to define this corresponded non-recombining genomic region as the SDR from spinach Y chromosome (Fig 3.5). Identification of sex co-segregating SNPs from paternal map ('Cornell-NO.9' map) justified the XY type of sex determination in spinach as heterozygous loci for each male and homozygous locus for each female in mapping population. Besides, bin markers with segregation distortion were scanned based on chi-square test in terms of numbers of homozygous and heterozygous bins across loci from 80 F_1 individuals. A significance level of $\alpha = 0.05$ identified 397 out of 2,898 bins from the consensus map with segregation distortion. Among 397 distorted bins, 375 are homozygous biased and 22 are heterozygous biased. Positional mapping of these bins on map-based assembly identified five segregation distorted blocks from LG1, LG2, LG4, and LG5 (Fig 3.7). All five blocks were identified from telomeric regions of four pseudomolecules and the most severe distortion was identified from Chr2. Interestingly, most of the segregation distortions are homozygous biased, and those distorted regions might be related to the selection of favorable traits and partially fixed allelic diversity in spinach breeding. These distorted bins have increased the mapping coverage and accuracy as those distorted telomeric regions would have not been mapped.

Genomic identification and analysis of X counterpart of SDR

Screening of 46 sex co-segregating bins identified a genomic interval located on Chr1 from 110.01Mb to 129.42Mb consisted of contigs containing 39 sex co-segregating bins as the X counterpart of SDR. Particularly, this non-recombining region possesses the approximate same genomic position from FISH signals of a 45S rDNA where hybridized locates at the short arms of the largest pair of chromosomes (Deng et al., 2013b). A total of 216 genes were identified from this 18.4Mb sex-linked region in assembly based on gene annotation retrieved from previous reference genome (Table 3.6). A total of 16,713 protein-coding genes from five autosomes shared an average density of 31.86 genes per million base pair and 69.70% of repetitive sequences per million base pair. Nevertheless, the X counterpart of SDR was featured by significantly reduced genes density (12.10 genes/Mb) and increased repetitive sequences (84.10%) (Table 3.5). Genomic comparison among X counterpart of SDR, PAR, and autosomes revealed notable variations. The Welch two-sample T-test at $\alpha = 0.05$ level revealed the significantly diverged gene distribution between SDR and PAR ($P = 5.16*10^{-9}$), SDR and Au ($P = 9.63*10^{-15}$). These unevenly distributed genes corresponded to significantly increased repetitive sequences in SDR, supported by P = $1.51*10^{-10}$ and P = $6.78*10^{-8}$ from two sets of comparisons (Fig 3.6). Those diversified genomic landscapes match the expectation of sex chromosomes, as consequences of ceased recombination near sex-determining loci.

Discussion

Genotyping by genome re-sequencing substantially improve the resolution for detecting variants from the whole genome compared with the reduced genome coverage of the conventional GBS approaches (Mun et al., 2015; Xie et al., 2010; Zhou et al., 2015). Identification of a few millions of SNPs provided the chance for mapping regions with few or no restriction sites of the selected enzyme. Nevertheless, the accuracy of genotype could be largely related to population structure besides the sequencing depth for genotyping (Bowers et al., 2012a; Bowers et al., 2016). In our study, 12× depth resequencing generated 768,797 SNPs with one or two missing data for two parents, which could be largely attributed to the heterozygosity background of two parents. In addition, an approximate 75% reduction of SNPs for ‘lm × ll’ map attributed to erroneously typed parents genotype, supported by non-segregated patterns of F_1 progenies from those loci. Hence, accurate genotypes, especially for parents, is crucial, which can be improved by genotyping in higher depth.

Empirical screening of SNPs for contigs or scaffolds for each progeny based on the relation of adjacent SNPs substantially improve the genotyping accuracy for larger contigs to reduce missing data. However, small contigs are bearing lower numbers of SNPs and higher chance for inconsistent genotype for one progeny, which inhibits the genotype correction and increase genotype errors (Bowers et al., 2012a). In our analysis, 265,338 loci from contigs with < 5 SNP, genomic region with <10 inconsistent genotypes, and non-segregating regions were removed manually. Finally, 40,324 selected SNPs covered 1404 contigs for “Viroflay” map and 256,636 selected SNPs covered 2379 contigs for “Cornell-No.9” map was selected for map construction. Manual filtering of those SNPs is necessary for quality control as those incorrect or unreliable SNPs could inflate the final map distance by errors of SNP bins or introducing artifacts of recombination.

Resolution of genetic maps is largely dependent on both population size and accurate genotyping. The effective detection of recombination is limited by size of population given numerous DNA markers. In our F_1 population map, theoretical mapping resolution was estimated to be around 1.25cM for both Viroflay' and 'Cornell-NO.9' map with 80 individuals and our map supported the estimation based on average map distance of 2.18cM and 1.61cM for each map. Although 212,414 SNPs were mapped to "Cornell-NO.9" map, genomic region with co-segregating markers are commonly existed, especially the region bearing linkage disequilibrium Hence, a larger F_2 population could ideally improve the resolution by breaking linkage blocks for increased rate of recombination. Moreover, for two sets of bins generated here, 205 out of 2490 bins from 'nn × np' genotypes and 176 out of 913 bins from 'lm × ll' genotypes were failed to be mapped to major 6 linkage groups at the LOD = 8.0 level, mostly attributed to insufficient linkage caused by incorrect genotypes for SNP bins. Besides, the huge reduction of SNP bin numbers of 'Viroflay' map introduced several genetic gaps which are greater than 5cM to 'Viroflay' map while reduced around 60% of unique contigs compared with 'Cornell-NO.9' map. Hence, both enhanced genotyping and larger population size will be key factors for improving the mapping resolution.

Combination of multiple maps ensures the quality of map-based genome assembly since anchoring highly-correlated orders among maps largely prevents the false positive from a single map (Li et al., 2017; Tang et al., 2015). In our study, 'Cornell-NO.9' and 'Viroflay' map effectively anchors 96.68% (1,661 out of 1,736) contigs to six chromosomes, validated by highly correlated position from genetic maps and genome assembly. In all, our two robust genetic maps significantly improve the spinach genome assembly, however, the existed gaps among contigs remained unresolvable due to limited N50 size of contigs from Illumina sequencing due to repetitive sequences. Long reads sequencing in future studies will be effective to address the problem and largely improve genome quality.

Recombination suppression plays a substantial role to prevent sex reversal from dioecy to hermaphrodism, and also provides the evidence to define sex chromosomes and non-recombining SDR (Charlesworth & Charlesworth, 1978; Liu et al., 2004; Ma et al., 2004). As one signature of dioecious plants with sex chromosomes, sex co-segregating DNA markers from recombination suppressed-region were identified from papaya (*Carica papaya*), asparagus (*Asparagus officinalis*) and white campion (*Sativa latifolia*) (Delph et al., 2010; Harkess et al., 2017b; Ming et al., 2011b). Genomic analysis of sex-linked genomic region provided the basis of SDR landscape in terms of portions of genes to repetitive sequences (Harkess et al., 2017b). Herein, we identified 5,678 sex co-segregating SNPs (39bins) which is consistent with the genotype pattern of T11A and V20A DNA markers from previous studies. The diversified landscape patterns of genes and repetitive sequences across SDR, PAR, and Au provided an additional justification regarding the sex-linked region we identified (Fig 3.6). These unique features from SDR could be potentially attributed to repetitive sequences insertions and gene degradation during sex chromosomes evolution, as discovered from several previous studies (Hobza et al., 2015; Kejnovsky et al., 2009; Wang et al., 2012). Hence, spinach sex chromosomes are at the second stage of sex chromosome evolution with suppression of recombination at the SDR and viable YY genotype (Wadlington & Ming, 2018b).

To further validate the SDR we identified, 12 sex-linked genes derived from *de-novo* transcriptome assembly from previous studies were retrieved and blast against the published reference genome (Okazaki et al., 2019; Xu et al., 2017a). We found that 7 out of 12 genes from transcriptome assembly were mapped to 18.4Mb genomic region we identified (Table 3.7), which provided additional evidence of sex-linked genes we identified. The size of SDR of the Y chromosome defined by genetic mapping is larger than the divergent XY sequences defined by DNA sequence identity as shown in papaya SDR (Wang et al. 2012). The current SDR in spinach might be exaggerated due to a small population. Fine mapping based on a larger F_2 population will be more accurate for delimiting non-recombining SDR boundaries. The genetically defined SDR borders solidified the existence of sex chromosomes in spinach. These 5,678 co-segregating SNPs and 18.4 Mb sequence of X counterpart of SDR provide valuable sources for exploring sex chromosome evolution and studying sex determination candidate genes in dioecious spinach.

Conclusion

Resequencing-based ‘Viroflay’ and ‘Cornell-NO.9’ spinach genetic maps significantly improved marker density and coverage with hundreds of thousands of SNPs from the whole genome while facilitated consensus map-based genome assembly with 23.4% improvement and high correlation between two maps. As an enhanced genomic resource, the map could positively impact the QTL mapping, cloning of essential agronomic trait related genes, and genomic selection-based breeding in spinach. In addition, genomic identification of f5,678 SNP sex co-segregating SNPs from the 18.4Mb region provided valuable estimation in terms of the physical position of sex determination region with strong evidence of reduced gene density, increased repetitive sequence insertions and suppressed recombination. Those results advanced sex chromosome study by serving as genomic resources of detecting sex-specific genes, building gene networks, and analyzing gene function in spinach sex chromosomes.

Data availability statements

The genomic sequences of spinach reference genome and gene annotation can be found in the European Nucleotide spinachBase (http://www.spinachbase.org/).

Tables and figures

Table 3.1 Summary of resequencing data

Samples	clean reads numbers	Average depth	GC content (%)	Q20 (%)	Q30 (%)
Cornell-NO.9	76,112,154	11.41	40.55	97.02	93.11
Viroflay	89,792,317	13.47	40.95	96.97	93.03
F_1 progenies	3,151,330,530	6.12	39.15	96.27	91.73

Table 3.2 Segregation pattern of SNPs markers

Categories	Numbers of SNP
Total combined vcf SNP	3,229,056
Parents genotypes	
‘lm x ll’	423,352
‘nn x np’	749,807
‘hk x hk’	787,691
others	1,268,206

Table 3.3a Summary of “Viroflay” linkage map

LG	SNP numbers	Bin numbers	Total Length (cM)	Average distance (cM)	SNP density (per Kb)
1	7725	196	92.29	1.77	0.23
2	7499	127	62.35	1.73	0.42
3	2070	85	58.98	2.46	0.26
4	5745	152	80.512	2.12	0.35
5	3842	122	50.87	2.67	0.25
6	2401	56	56.27	2.34	0.64
Combined	29,282	738	401.28	2.18	0.36

Table 3.3b Summary of “Cornell-NO.9” linkage map

LG	SNP numbers	Bin numbers	Total Length (cM)	Average distance (cM)	SNP density (per Kb)
1	49,905	518	92.89	1.57	0.57
2	33,108	332	77.93	1.73	0.54
3	43,853	455	77.33	1.58	0.80
4	24,901	328	72.75	1.73	0.66
5	34,071	444	96.57	1.51	0.77
6	26,576	301	59.36	1.52	0.56
Combined	212,414	2,378	476.83	1.60	0.64

Table 3.4 Summary of genome assembly based on consensus map

	Anchored	Oriented	Unplaced
Bins (unique)	3,020	739	100
Bins per Mb	4.3	5.7	0.3
N50 Scaffolds	467	126	52
Scaffolds	1,661	131	76,068
Scaffolds with 1 marker	1,252	0	53
Scaffolds with 2 markers	141	25	22
Scaffolds with 3 markers	63	13	1
Scaffolds with >=4 markers	205	93	0
Total bases	698,320,906	128,858,779	298,147,897
Percentage	70.10%	12.90%	29.90%

Table 3.5 Statistics summary of contigs containing sex-linked regions

	Chromosomal assembly	Sex-linked region	Autosomes
Numbers of contigs	1,661	38	1,304
Total base pair (bp)	698,320,906	18,408,487	526,248,241
Average contig length (bp)	420,422	471,915	413,464
Total Repeats density (%)	69.70%	84.10%	69.51%
Protein coding genes (Per Mb)	21,015 (30.09)	216 (11.14)	16,713 (31.76)

Table 3.6 Summary of sex-linked genomic region

Reference region	Num of genes	Assembly region		bin size	SNPs
SpoScf_01759	2	111014346	111138316	88548	71
SpoScf_02108	3	111138417	111230141	2212	32
SpoScf_02117	2	111230242	111321366	62191	36
SpoScf_01981	0	111321467	111423794	66823	32
SpoScf_02017	2	111423895	111522435	21703	9
SpoScf_03182	0	111522536	111554701	2865	9
Super_scaffold_230	6	111554802	112063131	498614	93
SpoScf_04449	0	112063232	112075079	1652	6
SpoScf_00605	10	112075180	112429061	268781	139
SpoScf_01771	3	112429162	112552036	67706	25
Chr3	6	112552137	113552136	566078	278
SpoScf_03739	0	113552237	113571144	16200	14
SpoScf_01907	2	113571245	113679252	9660	22
SpoScf_02507	0	113679353	113743053	21342	19
Chr4	6	113743154	114743153	400234	483
Chr4	17	114743254	115743253	305979	115
SpoScf_01132	12	115743354	115954643	172676	177
SpoScf_02300	1	115954744	116031584	29917	18
Super_scaffold_232	31	116031685	118689962	2583030	771
Chr4	15	118690063	119690062	93650	129
SpoScf_01622	1	119690163	119830844	98464	73
SpoScf_03082	1	119830945	119866562	10892	21
Super_scaffold_60	20	119866663	122394076	2500299	959
Chr4	16	122394177	123394176	687849	201
Chr3	7	123394277	123855564	320962	300
SpoScf_00867	0	123855665	124127063	199842	73
SpoScf_01507	1	124127164	124279924	90498	95
SpoScf_01915	2	124280025	124386949	4234	35
Chr4	10	124387050	125387049	990895	197
Chr3	12	125387150	126387149	960637	192
Super_scaffold_208	4	126387250	126989067	80907	139
SpoScf_02328	2	126989168	127063152	24312	17
SpoScf_01429	1	127063253	127226441	97805	31
SpoScf_01919	2	127226542	127333019	63003	16
Super_scaffold_66	9	127333120	129084582	81011	767
SpoScf_01529	8	129084683	129235934	85766	42
SpoScf_01910	2	129236035	129343567	3358	7
SpoScf_03757	0	129343668	129362289	3584	11
SpoScf_02571	0	129362390	129422833	36790	24
Total	**216**	**total size**	**18408487**	**11620969**	**5678**

Table 3.7 Summary of sex-linked unigenes from previous studies

Transcriptome ID	NCBI ID	Position	Start	End	Occurrence in X counterpart of SDR
comp18846_c0_seq1	XP_021843860.1	chr4	55239847	55254080	No
comp32199_c0_seq1	XP_021850419.1	Super_scaffold_232	702718	718467	Yes
comp32303_c0_seq1	XP_021854084.1	SpoScf_01507	27596	44206	Yes
comp32347_c0_seq1	XP_021854033.1	chr6	30666497	30670688	No
comp41371_c0_seq2	XP_021846643.1	chr6	39439485	39447115	No
comp45039_c0_seq1	XP_021847201.1	Super_scaffold_60	725094	729796	Yes
comp50243_c0_seq2	XP_021846152.1	Super_scaffold_66	1002692	1007550	Yes
comp50308_c0_seq1	XP_021863440.1	chr4	56362426	56390169	No
comp41527_c0_seq1	XP_021857279.1	chr4	59061395	59064904	Yes
comp49566_c0_seq4	XP_021850779.1	chr3	64556952	64565535	No
comp43645_c0_seq1	XP_021848842.1	Super_scaffold_232	1183999	1186968	Yes
comp34398_c0_seq1	XP_021836980.1	SpoScf_02117	66941	69416	Yes

Figures

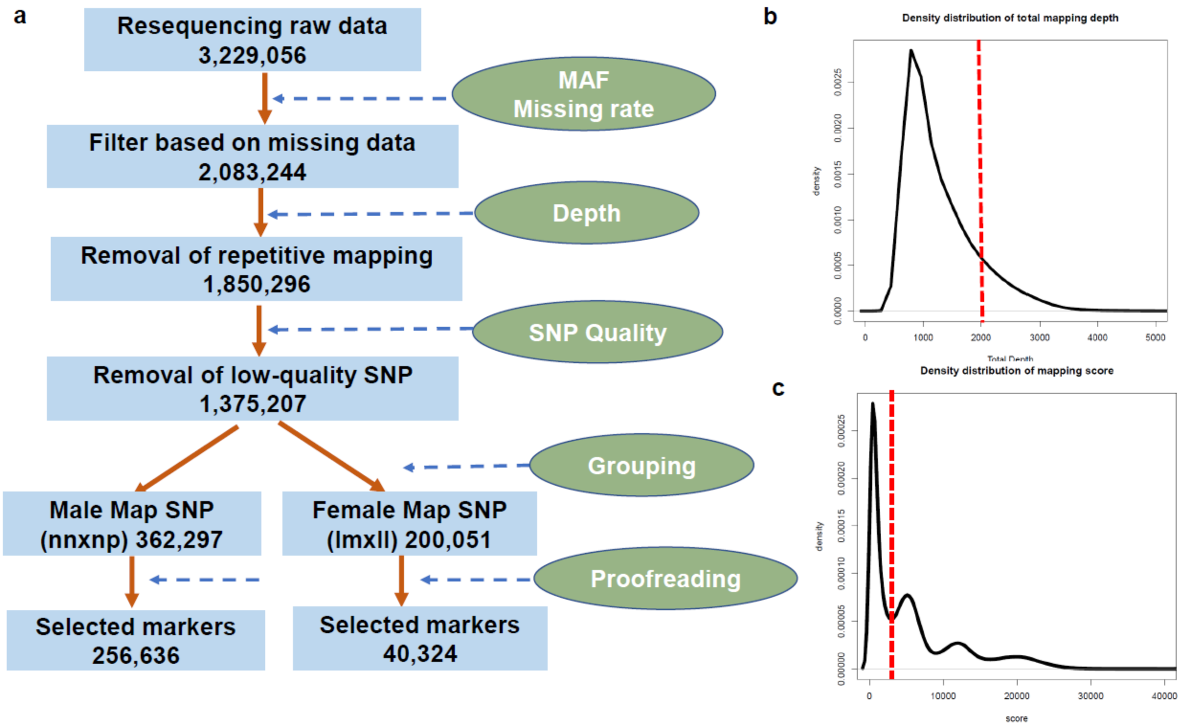

Fig 3.1 Workflow of SNP filtering for linkage mapping

a. Workflow of SNP filtering based on portions of missing data, sequencing depth and mapping quality.

b. Density distribution of total mapping depth for each SNP, the red line marked the threshold (Total depth = 2000) for SNP filtering.

c. Density distribution of mapping quality score for each SNP, the red line indicated the cut-off (combined quality score = 1000) for SNP filtering.

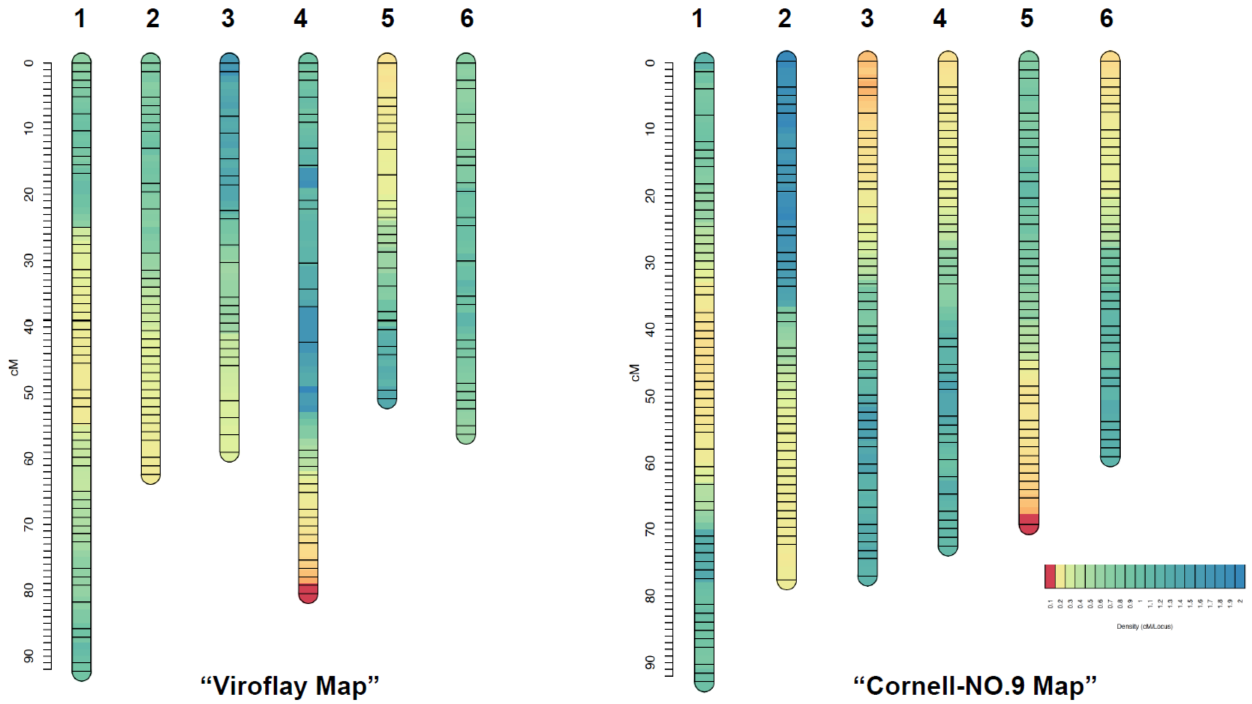

Fig 3.2 Genetic map of ‘Viroflay × Cornell-No.9’

Distribution of linkage groups (LGs) of 2,285 bins for Cornell-No.9 map and 737 bins for Viroflay map was presented. The scale bar indicated the density (cM/Locus), horizontal numbers indicated LG number and vertical scale indicated the linkage distance in centi-Morgan (cM).

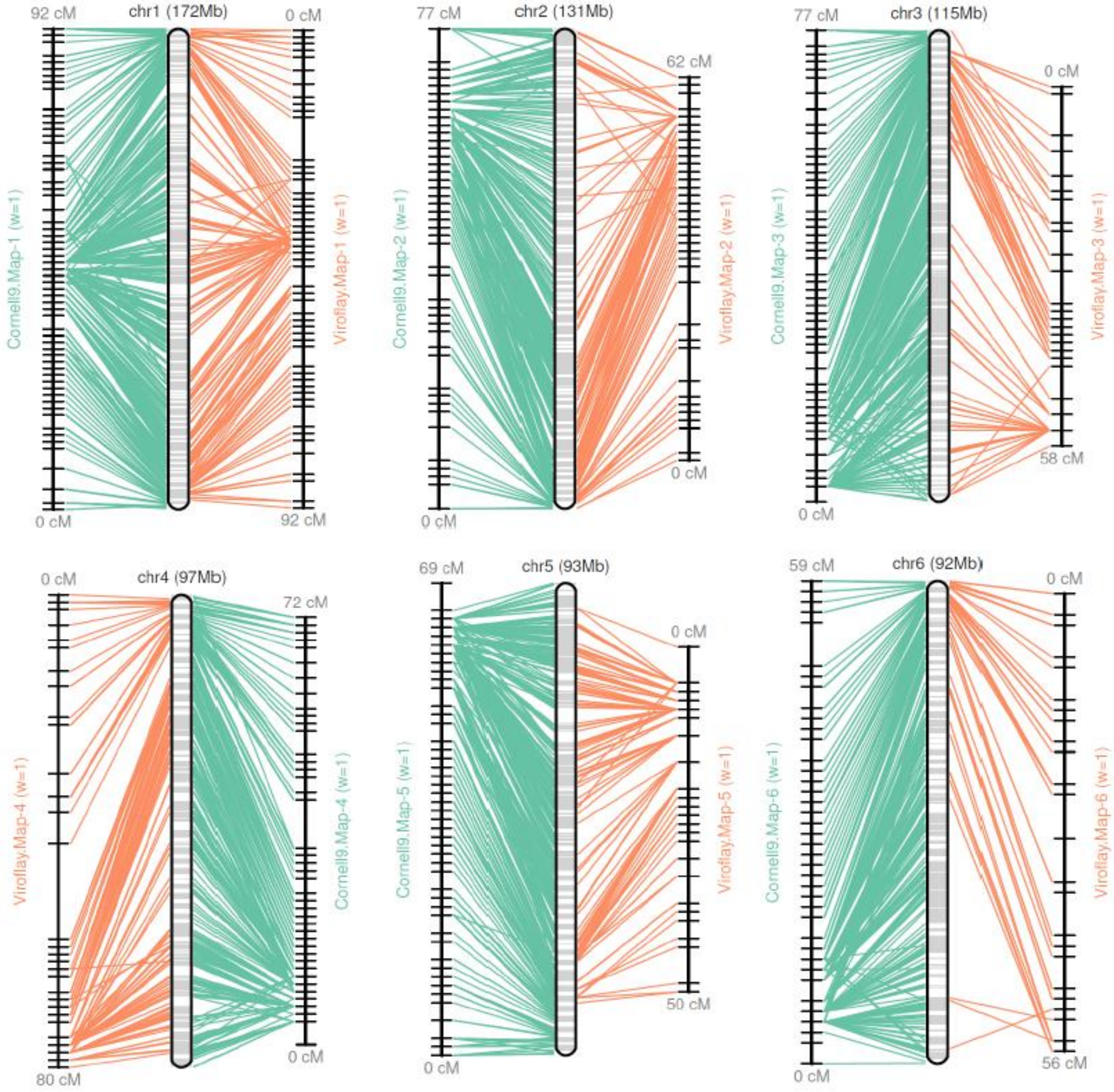

Fig 3.3 Consensus mapping-based assembly of spinach genome

Correspondence of genomic position of contigs and bins were connected in orange lines for "Viroflay" map and green lines for "Cornell-NO.9" map. The "grey shaded" bars represented integrated pseudomolecules from two maps.

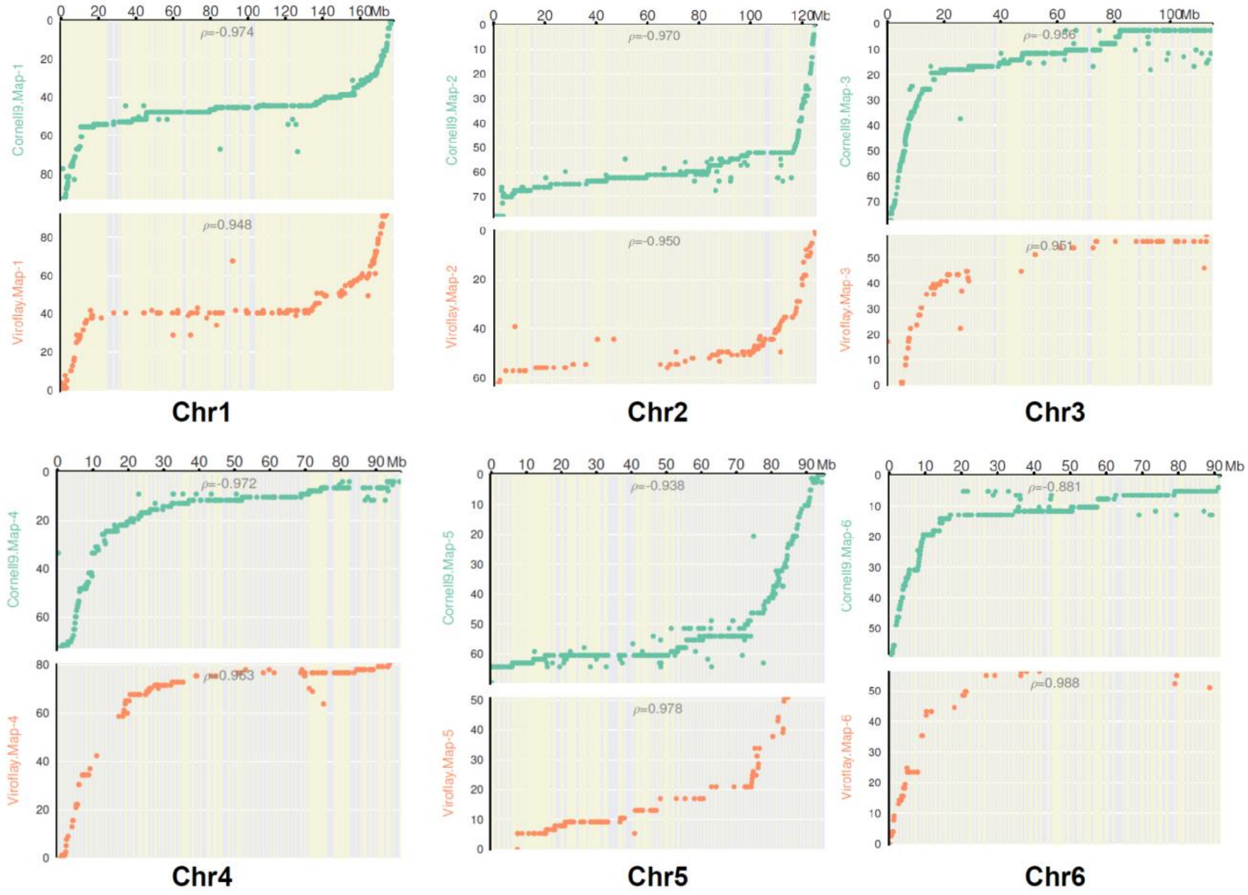

Fig 3.4 Correlation between pseudomolecules physical position and two maps

Correlation of spinach pseudomolecules and “Cornell-No.9” map was plotted in orange dots and green dots for “Viroflay” map. The horizontal direction indicated physical size of pseudomolecules and vertical direction indicated genetic distance for each map.

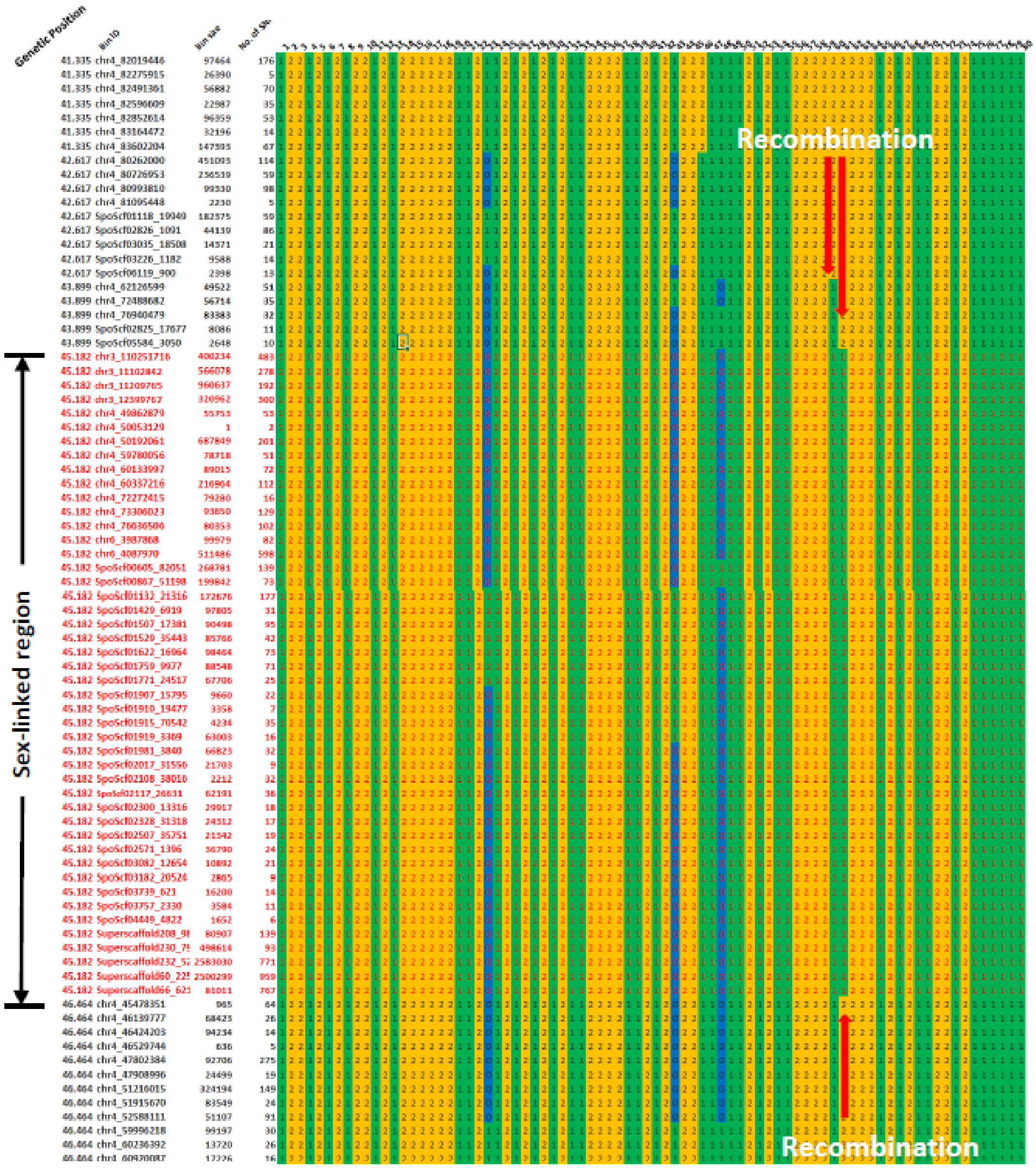

Fig 3.5 Graphical mapping of sex-linked region from 'Cornell-NO.9' Map

Genotypes of 79 bins were ordered based on genetic distance, genotype for each cell was encoded as "1" for homozygous from female parents (Orange), "2" for heterozygous from male parents (Green), "0" for missing data (blue). The SDR was marked in red font.

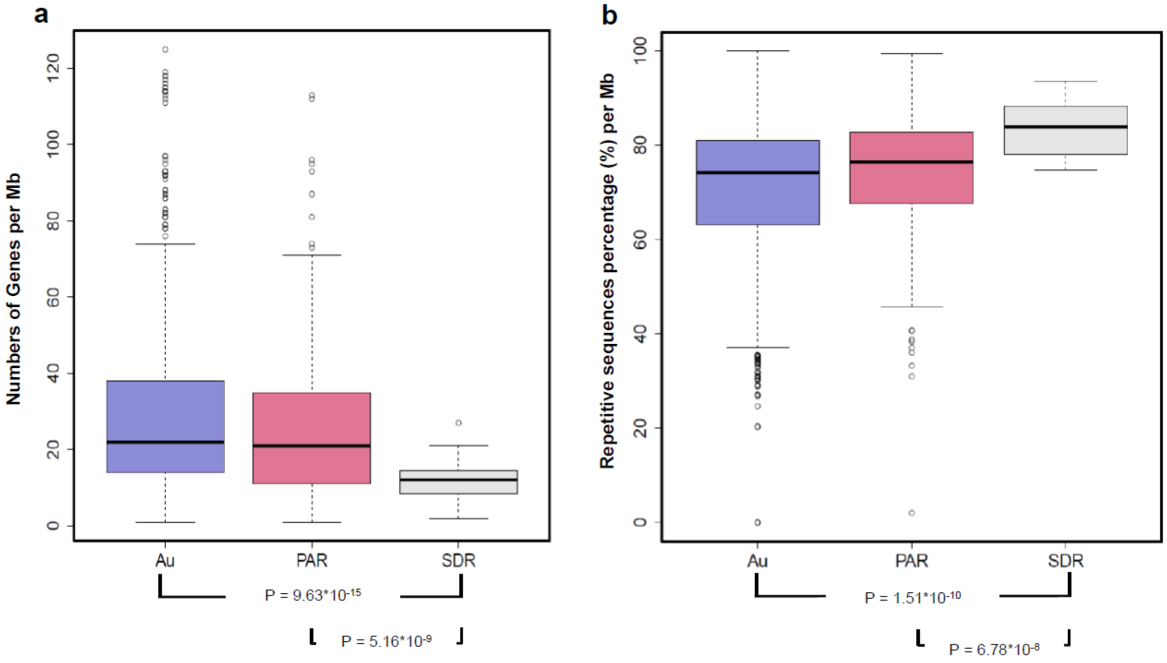

Fig 3.6 Genomic feature summary of sex-linkage effects

a. Comparison of gene density among autosomes (Au), Pseudo-autosomal region (PAR) and sex determination region (SDR).
b. Comparison of repetitive sequences distribution among autosomes (Au), Pseudo-autosomal region (PAR) and sex determination region (SDR).

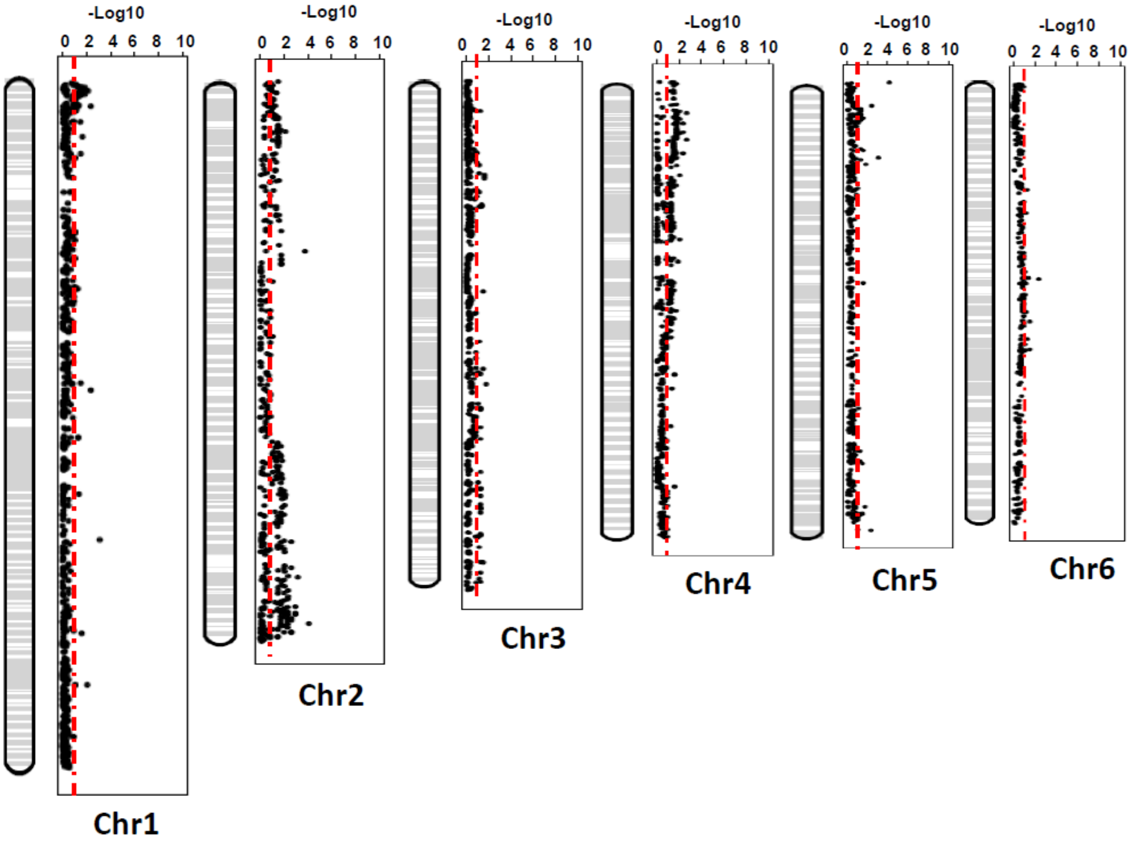

Fig 3.7 Distribution of segregation distortion Bins from spinach genome

Distribution of distorted segregating bins across six consensus map-based pseudomolecules. The guideline in red indicates the significance cut-off of the chi-square test at α = 0.05 based on -Log10 transformation from X^2 value (cut-off = 3.841).

Chapter IV. Genomic analysis of structural variations and evolutionary forces of sex chromosomes in spinach

Abstract

Suppressed recombination near sex-determining loci of dioecious spinach was identified by Y-specific genes, sex co-segregating DNA markers, and high-density genetic maps. However, the global genomic structure of the Y-specific region (MSY) and X counterpart remains unclear, which hinders the identification and functional studies of sex-determining genes. In this study, we anchored a chromosomal level genome of YY spinach with Hi-C, genetic maps, and reference guided correction by an equivalent XX genome assembly. A high-quality YY genome anchored almost 99% of the genome to six pseudomolecules. Comparative studies revealed a surprisingly large 39.26 Mb MSY and an equivalent 38.92 Mb X counterpart feature by three inversions and three collinear regions. Both the large MSY and X counterpart were the consequence of a recent *Spinacia* genus-specific and genome-wide LTR-RTs burst dated 0.15-0.2 MYA. Estimation of gene divergence revealed a 6 Mb interval with continuously declined pair gene identities which matches the border of one large inversion and leads to the diversification of three evolutionary strata dated from 0.1-0.3 MYA. The time course estimations indicate a pair of young sex chromosomes, which is supported by YY viability and limited structural rearrangements between XY chromosomes. Among these rearrangements, the largest genomic inversion is thought to be the main cause of the MSY progression, and this small 6 Mb interval could be an ideal region for further investigating sex-determining genes.

Key words: Genome assembly, MSY and X counterpart, LTR-RTs burst, gene divergence

Introduction

Sex chromosomes are widely found among eukaryotic lineages (animals, plants, and fungi), which evolved independently and multiple times (Ming et al., 2011b; Ming et al., 2007; Renner & Ricklefs, 1995). Despite diverse origins of sex chromosomes, a common selective pressure includes the sexual selection and sex maintenance are shared by all sex chromosomes and these evolutionary forces lead to a different genomic landscape between two homologs which differs as autosomes. Established theoretical and experimental evidence suggests that sex chromosomes evolved from autosomes initiated by two mutations relates to male and female sterility. The subsequent accumulative deleterious mutations, gene degradation, and structural variations evoked by ceased recombination around the sex-determining loci further expedite the divergence between two homologs and finally formed into a male-specific region on the Y chromosome (MSY).

Characterizing genomic sequences of MSY and its X counterpart is fundamental to understand the evolutionary forces accompanying sex chromosome evolution (Charlesworth & David, 2004). To date, very few complete sequences of MSY have been sequenced, due to wide-spread repetitive elements and difficulties of splitting female-specific sequences from MSY assembly. In papaya, two inversions were identified between the HSY (8.1 Mb) and X counterpart (3.5 Mb) of sex chromosomes derived from BAC-BAC sequencing. The inversions caused recombination suppressions and led to the subsequent spread of structural variations, supported by expanded HSY by retro-transposable elements. The evolutionary strata matched the position of two inversions and provided a strong justification of progression and expansion of the HSY (Wang et al., 2012). Another recent sequenced dioecious shrub *Salix purpurea* featured a large-scale palindromic structure of sex-specific region on its W chromosome and expanded gene numbers translocated from autosomes. The presence of palindrome indicated an adaptation of evolution shared by sex chromosomes from both plants and animals (Zhou et al., 2020).

The lack of genomic sequences further hindered the studies of sex differentiation-related genes, particularly the identification of sex-determining genes (Charlesworth, 2013b; Vyskot & Hobza, 2004). In general, genetic mapping of sex-determining genes is less feasible as no recombination around the loci. To date, sex determinants or candidates for sex differentiation were identified from very few dioecious species. For instance, a young Y chromosome in garden asparagus was identified with only 750Kb MSY genomic non-recombining region defined. The two Y-encoded factors with this region, named *SOFF* and *aspTDF*, act independently to suppress the carpel development and promote stamen development, respectively (Harkess et al., 2017c). The genus-wide genome sequencing of the *Phoenix* genus and screening of sex-specific sequences by Kmer analysis identified a 2Mb Y-specific region shared by 14 species under the genus (Torres et al., 2018). In kiwifruit, transcriptome of floral-biased expression pattern and DNA sequencing revealed two Y-encoded sex-determining loci *FrBy* and *SyGl*, which is consistent with "two-gene" model (Akagi et al., 2018; Akagi et al., 2019). In all, genomic sequences provided crucial references to identify key genes involved in sex determination.

The dioecy of garden spinach is maintained by a pair of homomorphic XY sex chromosomes as evidence from cytogenetics studies and crosses (Deng et al., 2013b; Deng et al., 2013c; Ellis & Janick, 1960; Lan et al., 2006). Identification of andromonoecious spinach (a plant with both hermaphrodite flowers and make flowers) and viable YY spinach further indicated its nascent sex chromosomes (Wadlington & Ming, 2018b). From molecular level, sex co-segregating DNA markers, *De-novo* assembled sex-linked genes ,and identification of a sex-linked scaffold revealed effects of suppressed recombination near sex-determining loci (Kudoh et al., 2018a; Okazaki et al., 2019)(unpublished data). A large non-recombining MSY was mapped to the largest linkage group of spinach with 4000 sex co-segregating SNPs along with approximate 18.41 Mb of X counterpart estimated (Unpublished data). Nevertheless, it was still unclear of the genomic size of MSY on the Y chromosome and its relation to X due to chimeric assemblies introduced from homologous sequences from the XY genotype. However, the YY viability provided a valuable chance to assemble MSY region. Further comparative analysis between the MSY and X counterpart could shed light on the molecular basis of spinach sex chromosome evolution.

Here we used the genetic maps and Hi-C data to order contigs of spinach YY sequences into a chromosomal level genome. We defined the genomic position of MSY and its X counterpart by scanning sex co-segregating bin markers on the Y chromosome and using synteny to anchor its counterpart on the X chromosome. We characterized the genomic landscape of gene divergence, chromosomal rearrangement, and time-courses of repeat elements from X/Y chromosomes, which helps us infer forces during sex chromosome evolution. We further surveyed the evidence of MSY expansion based on comparison of orthologous genes and repetitive sequences among *Spinacia* genus and other genera under the same family. The discovery of SVs and gene divergences on the Y chromosome details the molecular basis of sex chromosome evolution and provides robust evidence of SDR expansion after ceased recombination during the second evolutionary stage of spinach sex chromosomes.

Methods and materials

Genome assembly of YY genome by genetic maps and Hi-C

The genetic maps of “Viroflay” × “Cornell-NO.9” were derived from methods and population of the last chapter using PacBio-sequenced YY scaffold-level genome as a reference for SNP calling. The purpose for genetic mapping is to assist grouping of fragmented contigs due to weak crosslinks by limited N50 size from primary genome assembly. Initially, the two genetic maps for two parents were generated and the distribution of contigs in six linkage groups was retrieved. A reference genome of female spinach (XX genotype) (unpublished data) was used to orient the orders of contigs by RaGOO (https://github.com/malonge/RaGOO) (Alonge et al., 2019). These two sources of information were integrated as guidance information for assembly. Finally, the sex-linked contigs from “Cornell-NO.9” map was confirmed from Hi-C assembly by counting the occurrence of sex co-segregating contigs from six pseudomolecules.

Genomic analysis between MSY and X counterpart

Mummer4.0 (Marcais et al., 2018) (https://github.com/mummer4/mummer) pipeline was performed to compare the XX and YY genome assemblies with respective full genomic sequences (parameter settings: minimum length for match = 2000bp, maximal matches that are unique in both sequences). The genome-wide identification of syntenic genes was further conducted by MCscan pipelines (https://github.com/tanghaibao/jcvi/wiki/MCscan-(Python-version)) (Parameter settings: cs-score >= 0.65). The corresponded genomic region and syntenic genes within the MSY and X counterpart were retrieved for further SDR micro-synteny analysis. We also characterized the genomic landscape of MSY and X counterpart in terms of the distribution of genes, repetitive sequences, and chromosomal rearrangements. Specifically, gene numbers and percentage of repeat elements each 200 Kb region was calculated then plotted in a 500Kb size sliding windows by heatmap.

Estimation of pair-genes divergence between X and Y chromosome

The divergence time of X and Y chromosome could be dated by calculating the mutation rates among Y-encoded genes because these mutations provided a molecular clock to estimate the ages of sex chromosomes (Charlesworth, 2013a). We made a careful selection of sex-linked transcripts among syntenic genes based on sequences identities and aligned lengths of transcripts sequences derived from each genome. The transcripts with low identities (<85%) and transcripts with short alignment length (hits < 60% of query length) were not included for divergence analysis. Particularly, we used an easy_KaKs pipeline (https://github.com/tangerzhang/FAFU-cgb/blob/master/easy_KaKs) to calculate the rate of synonymous substitution (Ks) of transcripts from each sex pair genes and plot the distribution in boxplot each 5 Mb genomic region (settings: approximate methods: YN) (Wang et al., 2010; Yang & Nielsen, 2000). The potential evolutionary strata were divided based on the distribution of Ks value in each boxplot and the rate of nucleotide substitution was set to 6.5e-9 as substitutions per site per year to infer ages of the Y chromosome. Besides, *de-novo* gene annotation yielded different numbers of genes between X and Y chromosomes due to the different continuity and quality of two assemblies. Thus, the biases introduced by different annotations were alleviated by comparison of MSY gene model and XX genome assembly sequences. Those hits from the X chromosome with 100% of alignment length (coverage) were retrieved as conserved X/Y gene pairs. The numbers of SNPs, INDELs, and identities of each conserved gene pair was plotted in a sliding window across the MSY genomic region.

Evidence of MSY expansion by comparative analysis with other species

As the only dioecious genus under the Amaranthaceae family, the evidence of sex chromosome evolution of the *Spinacia* genus might be enhanced through comparing the MSY to a corresponded region of other closely hermaphrodite species. In our analysis, we retrieved the gene annotations of quinoa and sugar beet from the Phytozome database (https://phytozome.jgi.doe.gov/pz/portal.html) to study the syntenic pair genes distribution between genes from spinach MSY and corresponded part of other species by McScan pipeline.

The region with > 10 collinear genes was used for further micro-synteny analysis includes the size variations of regions and the numbers of genes.

Comparative analysis of repeats elements evolution among five genera

The sequenced genomes of six species from respective genera of Amaranthaceae family provide opportunities to study evolution of repetitive elements among genera. To understand the distribution and time scale of each type of repetitive sequence, we made a comparison among them in terms of the fraction of repetitive elements with different Kimura distance levels. Initially, we performed the genome-wide annotation of repeat elements by a combination of RepeatModler and RepeatMasker which integrated in a set of personalized scripts (Smit & Hubley, 2010; Tarailo-Graovac & Chen, 2009). Briefly, detection of repeats by *De-novo* detection was served as a database to further identify rest repeats with a > 80% identities to repeats existed in database. Those repeats were classified by TEclass and tandem repeat finder packages: DNA elements, LTR elements, SINEs, LINEs, and unknown elements. We further calculated and visualized the kimura substitution rate level distribution (X-axis) among repeats class and their corresponded percentage of genome size (Y-axis) by createRepeatLandscape.pl scripts (a function from RepearMasker package) (https://github.com/rmhubley/RepeatMasker/blob/master/util/createRepeatLandscape.pl). Finally, we presented the landscape of each genome based taxonomical relationship of six genera (https://www.ncbi.nlm.nih.gov/taxonomy).

Estimation of LTR-RTs insertion in spinach genome

As one of the substantial components of repetitive sequences in plant genomes, distribution and time-course analyses of LTR-RTs could be informative for genome size evolution. Although the LTR-RTs evolved independently among species with substantial variations, the time-course of LTR-RTs in the same species can be estimated by molecular clock based on the divergence of flanking LTRs of the same retrotransposon. Thus, we performed the precise annotation of LTR-RTs by LTR_retriever pipeline (https://github.com/oushujun/LTR_retriever/blob/master/LTR_retriever) to study the divergence of LTR-RTs. Each intact retrotransposon which includes two long terminal repeats (LTRs), two target site duplications (TSDs), and coding regions encoded by protease, reverse transcriptase, integrase, and ribonuclease H domains was chosen for analysis. The insertion time (T) of those LTR retrotransposon was calculated using the formula *T=K/2r*, where K is the distance and r is the rate of nucleotide substitution, which was set to 1.3e-8 as substitutions per site per year, as reported from rice (Ma & Bennetzen, 2004). Moreover, we made an LTR-RTs insertion time comparison among MSY, the Y chromosome, and whole-genome level.

Results

Construction of genetic maps and genome assembly of YY genome

The initial YY genome assembly by CANU pipeline generated a total of 45,306 contigs with an N50 size of 26,493 bp. The SNP calling pipelines and genetic mapping protocol from last chapter enabled us to generate 302,879,667 SNPs across F_1 mapping population. A total of 36,9524 SNPs which represented by 19,815 bins were anchored to six linkage groups (LGs) in male map and comprised of 428.99 cM genetic distance with an average 1.53cM genetic distance between each two loci (Table 4.1). The female map is composed of 41,876 SNPs (5,362 bins) across the six LGs with an accumulative 412.78 cM and an average of 2.45 cM distance. As consequences of Hi-C and reference guided correction, 98.92% (950,323,615bp) of YY genome was anchored to six pseudomolecules (Figure 4.1a) (Table 4.2) and highly correlated with the order from 'male' genetic map (Figure 4.1b). An estimated of more than 70% of sex-linked region represented by sex co-segregating bins were mapped to a genomic interval from the largest chromosome, as consensus with previous cytogenetics studies (Table 4.3).

Genomic landscape of MSY and X counterpart

Comparison of two genomes by Mummer 4.0 with complete genomic sequences identified a total of 4,967 sets of blocks with an average length of 6,765bp from alignment. The short length of genomic blocks between two genomes can be explained by the different accessions of two genomes, a modest N50 size of contigs, and highly repetitive nature of spinach genome. Among those blocks, the non-recombining regions derived from linkage map were mapped to a genomic interval with an approximate 39.26 Mb as the MSY and a smaller 38.92 Mb as its X counterpart (Fig 4.2b). We found that the non-recombing MSY is featured by some noticeable genomic rearrangements, particularly, three inversions (140-142 Mb, 150-160Mb and 165-168 Mb on the Y chromosome) were identified and some region without correspondence between Y and X sequences were proposed as Y-specific region (141-145 Mb on the Y chromosome). The genome-wide identification revealed a total of 30,733gene pairs from six groups, which anchored a 78.97% of the genes from annotation (Fig 4.2a) and 391 pairs of syntenic genes were found from MSY region we defined (Cut-off: > 80% identities, > 60% coverage). Besides, a comprehensive genomic landscape was presented by combining micro-synteny of the syntenic genes, distribution of gene density, and density of repeats elements insertions (Fig 4.3). We found a higher average of repetitive sequences within the region we defined (80.62% per Mb) compared with rest of genome (70.43% per Mb), as corresponded to a lower average gene density. Specifically, few sex-specific regions coincide with a pronounced repetitive sequences distribution (> 90%). These insertions and inversions could be closely related to the expansion of MSY, as the consequences of ceased recombination.

Divergence of gene pairs matches patterns of genomic rearrangement

Among 1,201 annotated genes from MSY region, only 391 genes were classified as pair-genes based on gene model from both X/Y chromosomes due to discrepancy of annotation and assembly. Another 372 pair-genes were identified by MSY gene model against sequence of X counterpart of MSY; 233 pair-genes were defined by MSY gene model against X chromosome; the rest 228 genes are categorized as Y-specific genes. Classifications of pair-genes enabled us to estimate divergence of X/Y genes and their relations to genomic rearrangement. Overall, the pattern of pair-gene divergence defined three evolutionary strata in terms of Ks value distribution, including the oldest stratum which dated around 0.193 *MYA* (147-160 Mb of the Y chromosome)*,* a stratum with slight recent divergence around 0.172 *MYA* (160-170 Mb of the Y chromosome), and the most recent stratum dated around 0.137 *MYA* (130-147 Mb of the Y chromosome) (Fig 4.4). Particularly, a small gene cluster which reside the one boarder of inversion (148 Mb at the Y chromosome) revealed a significant pair-gene divergence compared with the rest genomic region of MSY. These distributions were further supported by sequence identities and numbers of variations (SNPs and INDELs) of conserved pair-genes (Fig 4.5a and Fig 4.5b). These reduced gene identities are mostly contributed by SNPs from both intron and exon region. The peak of reduced identities also overlapped the appearance of one board of genomic inversion.

Evolution of repeat elements among six genera

The comparison of repeat elements among six genera under family Amaranthaceae revealed different proliferation of repeat elements (Fig 4.4). Initially, the peak of repeat elements insertions is more recently from Beta, Kochia, Suaeda, Chenopodiumand, and Spinacia genus, compared with Amaranthus genus, as supported by a lower score of Kimura substitution level. Moreover, we found that a most recent burst of repetitive elements (Kimura substitution level < 5) exclusively existed in spinach genome and contributed significantly by LTR-RTs from *Copia* and *Gypsy* family. Interestingly, the more recent speciation for species under each genus associated with a higher fraction of repeat elements and a more recent burst of those elements. For instance, the genome of the most ancient *Amaranthus hypochondriacus* is the least repetitive among six genera but the more recently evolved spinach is the most repetitive by recent burst of repeat elements. Hence, we proposed that the speciation under family Amaranthaceae is closely related to repeat elements evolution. As the only dioecious species in this study, the more recent burst of LTR-RTs might be related to sex chromosome evolution.

Time course of LTR-RTs insertions in spinach genome

The LTR-retriever pipeline identified a total of 2,241 intact LTR-RTs with two flanking conserved LTRs from six pseudomolecules. Comparison of these LTR-RTs among genome-wide, the Y chromosome, and MSY indicated a large-scale burst shared by three regions which can be dated around 0.15 million of years (MYA) and this proliferation played a substantial role of genome size expansion. Particularly, a unique smaller peak of the LTR-RTs burst around 0.75 MYA (Fig 4.8a) was detected and these slight ancient LTR-RTs distributed on two sides of MSY (Fig 4.8b). Those older LTR-RTs within MSY distinguishes the subsequent genome-wide burst of repeat elements, which indicates potential roles during sex chromosome evolution.

Discussion

Genome sequencing of sex chromosomes is largely hampered by the complicated genomic structure of non-recombining SDR, includes wide-spread repetitive elements, a mixture of sex-specific reads, and limited information for orientations and orders. In our studies, the feasibility of assembly is substantially empowered by the YY genotype which eliminates chimeric assembly in the XY genotype. We further integrated a reference XX genome, the Hi-C data, and genetic maps as robust guidance to anchor contigs into higher chromosomal level. Although the continuity of sequences is hindered due to fragmented contigs, the genome has substantially improved the chromosomal level assembly (98.9%) compared with previous genome assembly (46.5%). The MSY and X counterpart we identified from two sets of assemblies could provide informative sources to make genomic and evolutionary studies.

Transposable elements, particularly the LTR-RTs, substantially contributed to the recent genome size variations among closely related genera, families, and even species under the same genus, such as genera from grass family and species under *Oryza* genus (Neumann et al., 2006; Piegu et al., 2006; Zuccolo et al., 2007). As the only dioecious genus in our study, the genus-specific repeats abundance in *Spinacia* genus revealed a putative correlation between dioecy and the pronounced repeats activities. However, it is inconclusive to state the role of repeat elements in the progression of MSY due to their ubiquitous distribution in whole genome. A significant higher fraction of repetitive elements was identified from both MSY and X counterpart which can be dated in the same time course as majority of LTR-RTs from whole-genome and sex chromosomes (Fig 4.8a). Thus, this shared recent LTR-RTs burst contributed significantly to the large size of both MSY and X but less likely to be the forces of the MSY expansion after ceased recombination considering the different time course of two independent events. Besides, a small-scale LTR-RTs insertion dated around 0.75 MYA were identified from collinear regions of the MSY with unknown reasons (Fig 4.8b). These LTR-RTs are less relevant to the subsequent MSY expansion, but these insertions might set up a pre-condition of ceased recombination near sex-determining loci.

Chromosomal rearrangement near the sex-determining loci is another factor closely related to the progression of MSY, which includes inversions, insertions, and translocations (Charlesworth & David, 2004; Vyskot & Hobza, 2004). The scale of rearrangements correlates with the progression of sex determination region and stages of sex chromosomes, as identified from asparagus, papaya, white campion, and willows (Wang et al., 2012) (Bergero et al., 2007b; Harkess et al., 2017c; Zhou et al., 2020). Comparison of MSY and its X counterpart in our studies identified three inversions and two Y-specific insertions (Fig 4.3), either features by gene divergence of gene pairs or extremely low gene density. The patterns of continuously declined gene identities among gene pairs from 146-153 Mb of Y chromosome could further lead to a higher chance of mutations, selections, and subsequent expansion during evolution. Therefore, this small region could be a candidate region to narrow down sex-determining loci combining with gene annotation and transcriptome data. The two Y-specific regions near the large genomic inversion (140-144 Mb and 150-152 Mb) are most repetitive across the MSY (Fig 4.3a). Although the estimated time of those insertions overlapped with the repeat burst from genome-wide, the LTR-RTs within those regions might still contribute to the proliferation of sex-specific region after the ceased recombination.

Identification of expanded MSY or X correspondence compared with respective orthologous region of outgroup species also provided valuable information of MSY expansion, as identified from papaya, chicken, human with their respective outgroup species (Bellott et al., 2010; D'Esposito et al., 2003; Gschwend et al., 2012; Ross et al., 2005). For instance, a 41% greater genomic size was estimated from papaya sex-linked orthologous-X sequences compared with its relative *V. monica* which possesses a larger genome. The phenomenal difference provides a strong clue of a recent expanded X chromosome in papaya by retro-transposable elements insertions. In our studies, we analyzed the same pattern derived from comparison of spinach, quinoa, and sugar beet. Given that a significant size expansion of few regions where resides at the boarder of spinach MSY were identified (Fig 4.8), those region with increased size will be less informative because of the relative smaller genome size of quinoa and sugar beet compared with spinach.

In all, we proposed that the evolution of sex chromosomes in spinach was mainly driven by a large genomic inversion. Both the equivalent size of non-recombining regions and similar gene numbers from X and Y chromosomes indicated its early-stage evolutionary status with limited numbers of gene loss, a similar genomic landscape between homologs, and incipient pair-gene divergences. Nevertheless, an approximate 20% size of non-recombining MSY compared with the complete Y chromosome is rare since an expected small non-recombing region of sex chromosomes in this stage. Even though the recent genome-wide scale LTR-RTs burst partially explained the major reason for expanded large size, the reasons for this large suppressed recombination region with hundreds of gene pairs are not fully resolved. A larger mapping population and integrated transcriptome studies will be needed to fine map the boarder and refine candidate genes involved in sex differentiation.

Conclusion

Sequencing of YY spinach yielded valuable genomic sequences of the MSY region for sex-chromosomal comparative analysis. Although a large non-recombining region shared by both XY chromosomes triggered by a recent genus-specific and genome-wide LTR-RTs burst, the large-scale Y-specific insertions and degradation of Y-encoded genes were not found, matches the landscape of nascent sex chromosomes and features of YY viability. The divergence between each sex-linked pair-gene within MSY defined three evolutionary strata with an early stage of gene divergence. Particularly, genomic region surrounding one large inversion with the lowest pair-gene identities was identified as the oldest stratum which can be dated around 0.27 MYA with some potential gene degradation due to Y-genic insertions or small-scale inter-genic Y-specific insertions. We proposed the genomic inversion as a major evolutionary force related to progression of MSY. A small 6 Mb genomic region with the highest pair-gene divergence will be an ideal region for further identifying sex-determining genes.

Tables and figures

Table 4.1 Genetic map of "Cornell-NO.9" derived from YY genome

LG	SNP numbers	Bin numbers	Total Length	Average distance (cM)
1	49,905	518	92.89	1.57
2	33,108	332	77.93	1.73
3	43,853	455	77.33	1.58
4	24,901	328	72.75	1.73
5	34,071	444	96.57	1.51
6	26,576	301	59.36	1.52
Total	212,414	2,378	476.83	1.60

Table 4.2 Summary of genomic assembly of YY genotype of spinach

Group	**Size (bp)**
group1	217,806,951
group2	161,159,065
group3	126,850,049
group4	162,104,220
group5	126,542,630
group6	155,860,700
Total	950,323,615
Anchor rate	98.91%

Table 4.3 Occurrence of Sex-linked bins in YY genotype assembly

Group	Contig No.	Total length
group1	990	23,341,150
group2	29	794,568
group3	11	229,417
group4	26	584,113
group5	14	336,035
group6	15	359,744
Sex-linked	686	17,696,156

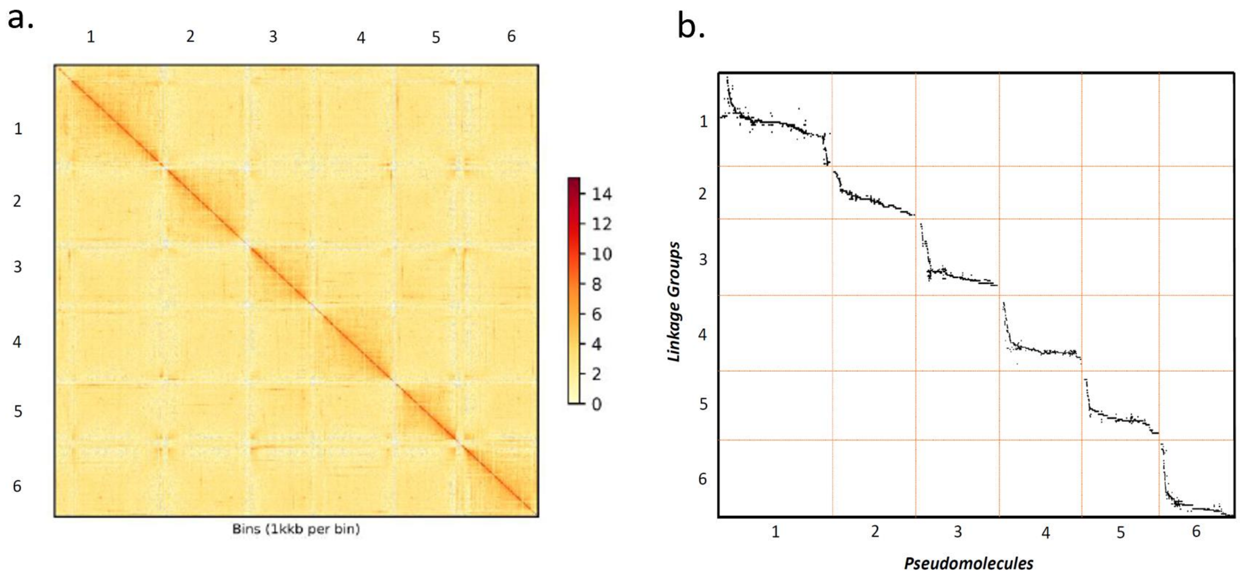

Fig 4.1 The Hi-C based genome assembly of YY genome

a. The assembly derived from 45,306 contigs from a 63X Pac-bio sequencing data and a 98.91% of the contigs were anchored to six pseudomolecules by integrating a XX reference genome, Hi-C data, and genetic maps. The continuum of genome assembly was presented in a heat map with a resolution of 1000Kb genomic region.
b. Matrix plot of “Cornell-NO.9” genetic map and genome assembly of six pseudomolecules of YY spinach genome. The horizontal direction represents genomic position and vertical direction represents the linkage distance.

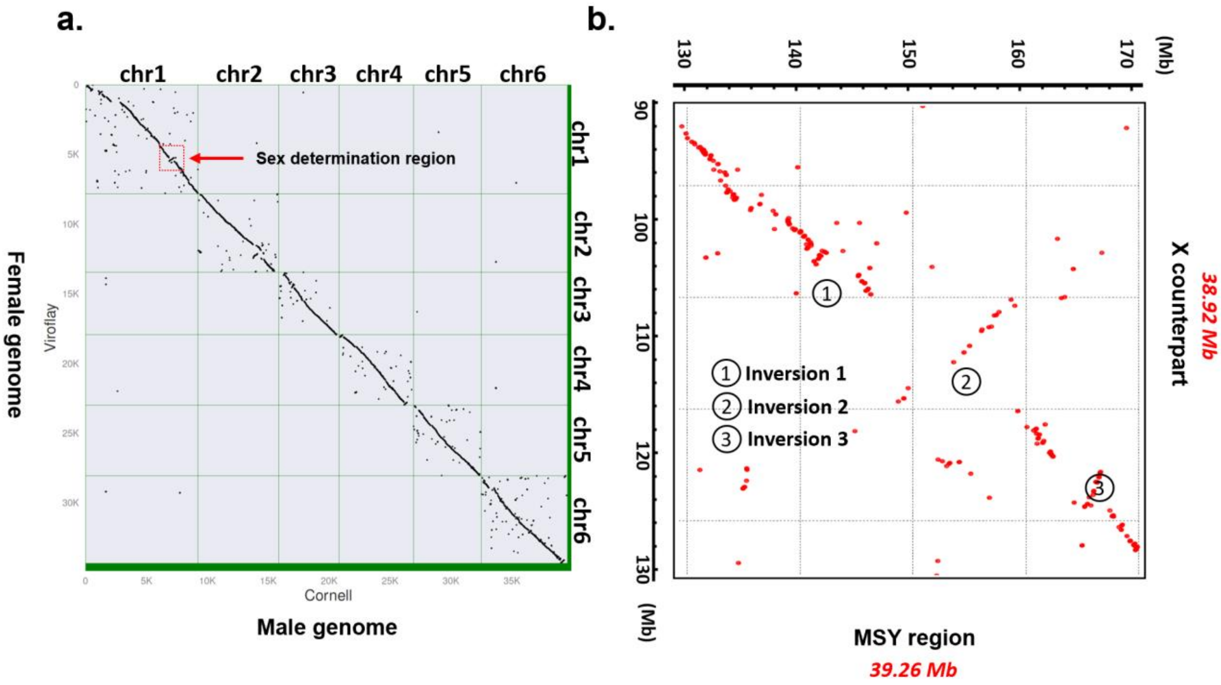

Fig 4.2 Genomic synteny between Female and male Genome

a. The 30,733 (>80%) of syntenic genes identified in between female and male spinach genome by MCscan pipelines (https://github.com/tanghaibao/jcvi/wiki/MCscan-(Python-version)).

b. Full-length genomic sequence alignment between MSY and X counterpart derived from Mummer4 pipelines (https://mummer4.github.io/), the numbers from “1-3” represents three genomic inversions.

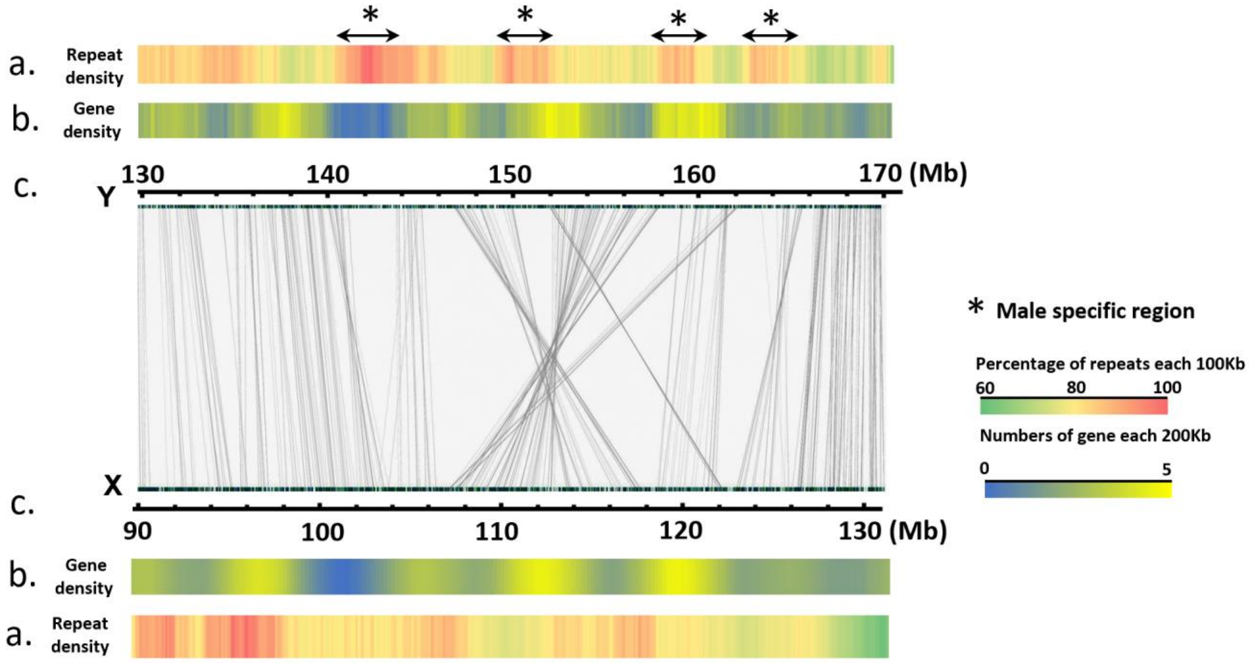

Fig 4.3 Genomic landscape of MSY and X counterpart

a. The repeat elements density was represented by heat map in a "green-red" scale based on the percentage of masked elements derived from RepeatModeler and RepeatMakser
b. The gene density was represented by heat map in a "blue-yellow" scale based on gene annotation from "Virofaly" and "Cornell-NO.9" genome assemblies.
c. The distribution of syntenic genes from MSY and X counterpart retrieved from MCscan pipelines, a total of 391 gene pairs were included.

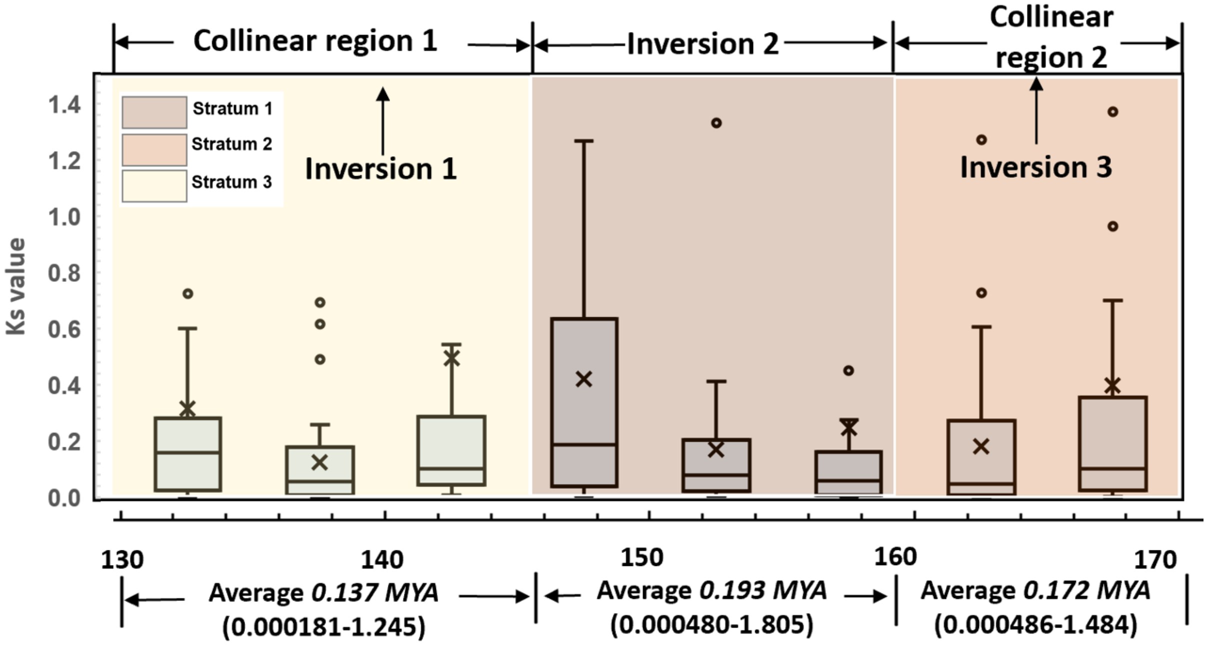

Fig 4.4 Estimation of gene divergence between MSY and X counterpart

The pair-wise comparison of pair genes defined by McscanX pipeline was represented in boxplots. Ks value for each gene pairs were represented in an independent box plot (5.0 million base pair for each group)

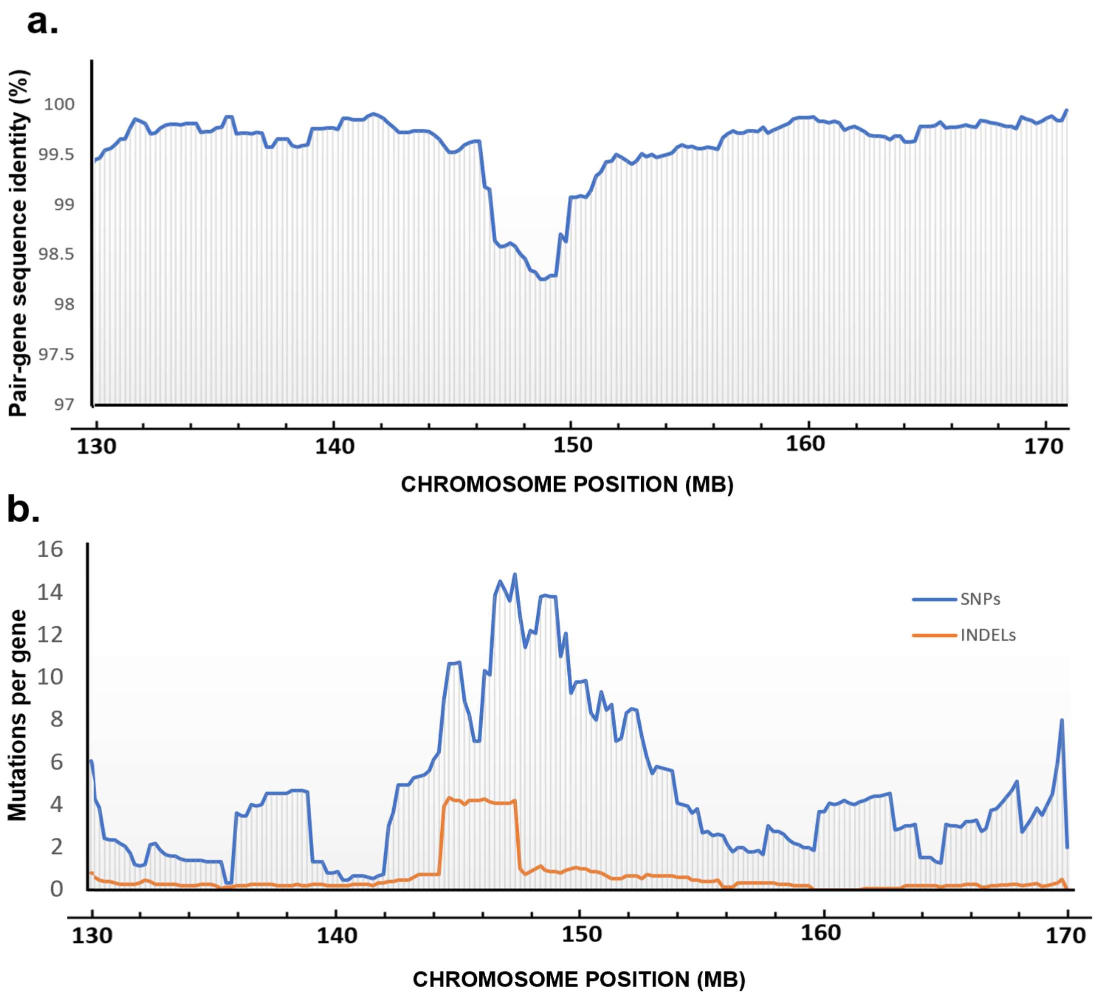

Fig 4.5 Distribution of identities and mutations of conserved pair-genes from MSY

a. The pair-wise comparison of genomic sequences between 239 conserved genes between MSY and X counterpart, The X axis stands for genomic position on Y chromosome and Y axis stands for identities of gene pairs.

b. The statsitics of numbers of SNPs and numbers of small INDELs within 239 genes pairs between MSY and X counterpart. The X axis stands for genomic position on Y chromosome and Y axis stands for numbers of SNPs and INDELs.

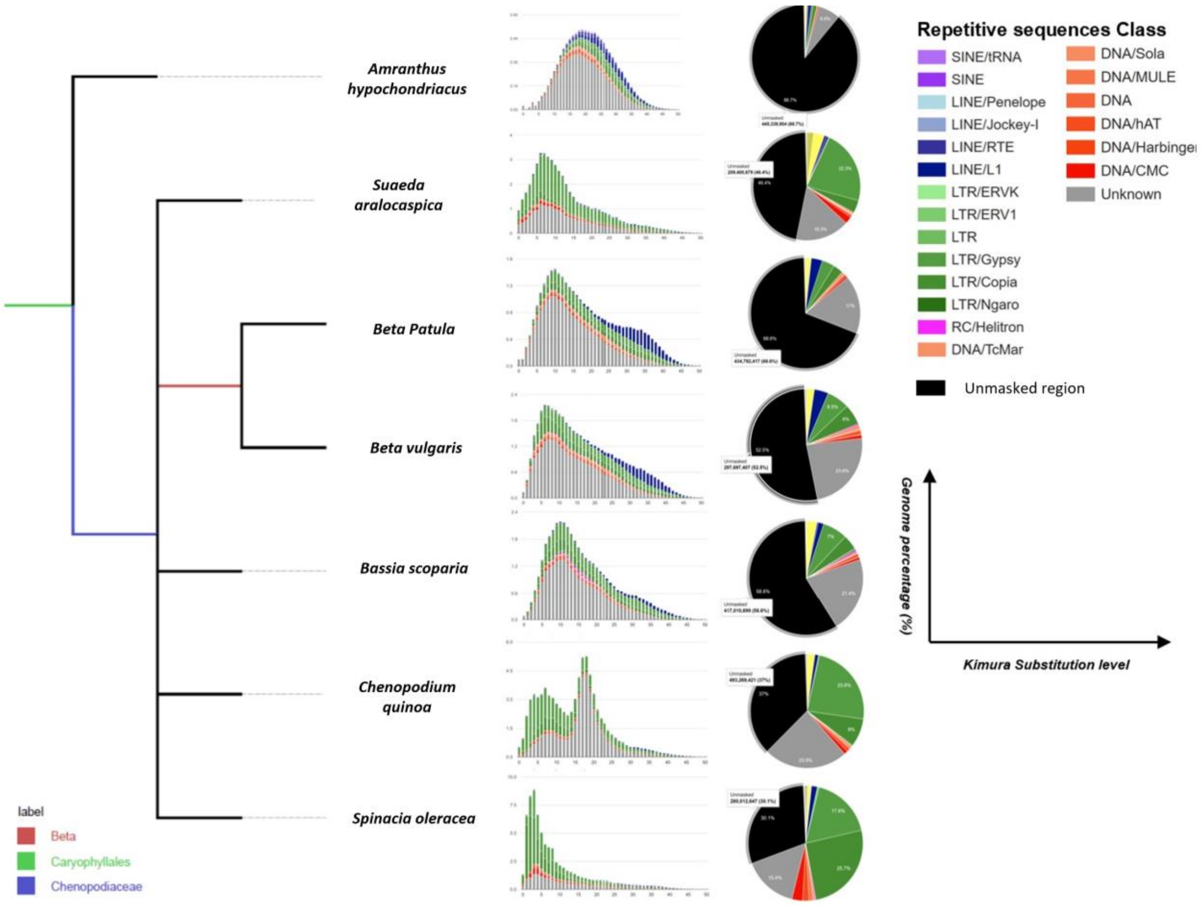

Fig 4.6 Evolution of repeat elements among six genera

The left taxonomy tree stands for the relative relations among seven species from six genera under the family Amaranthaceae. The percentage of each class of repeats elements was represented in a respective pie chart. The repeats in different Kimura distance level were stands in X axis, in accordance to respective contribution to genome size on Y axis

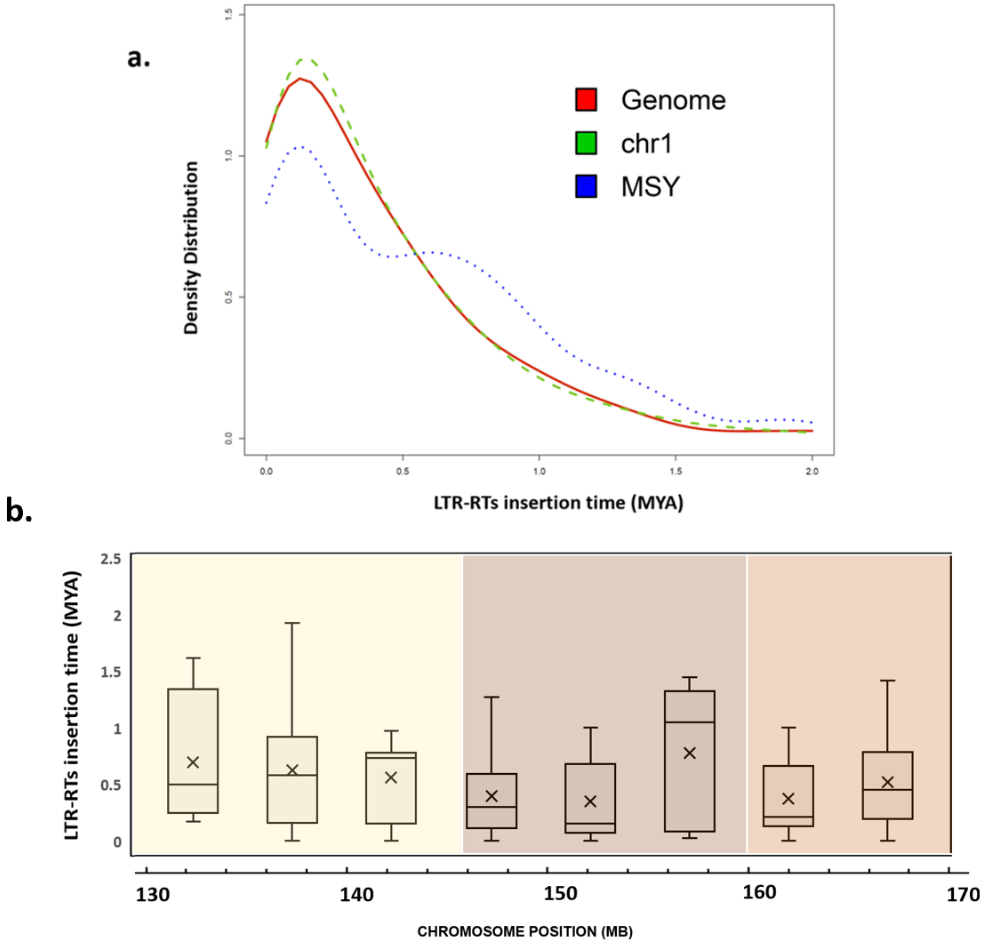

Fig 4.7 Time course of intact LTR-RTs in spinach genome

a. Distribution of intact LTR-RTs estimated insertion time by genomic region was plot by density plot. The red line represents the genome-wide, green line represents sex chromosome and blue lines represent MSY. The time estimation was performed by using formula T = *K/2r*, where K is the distance and r is the rate of nucleotide substitution, which was set to 1.3e-8 substitutions per site per year

b. Distribution of intact LTR-RTs insertion time of MSY region was plotted by boxplot in each 5Mb genomic region. The estimation of time was derived as Fig 4.8a described.

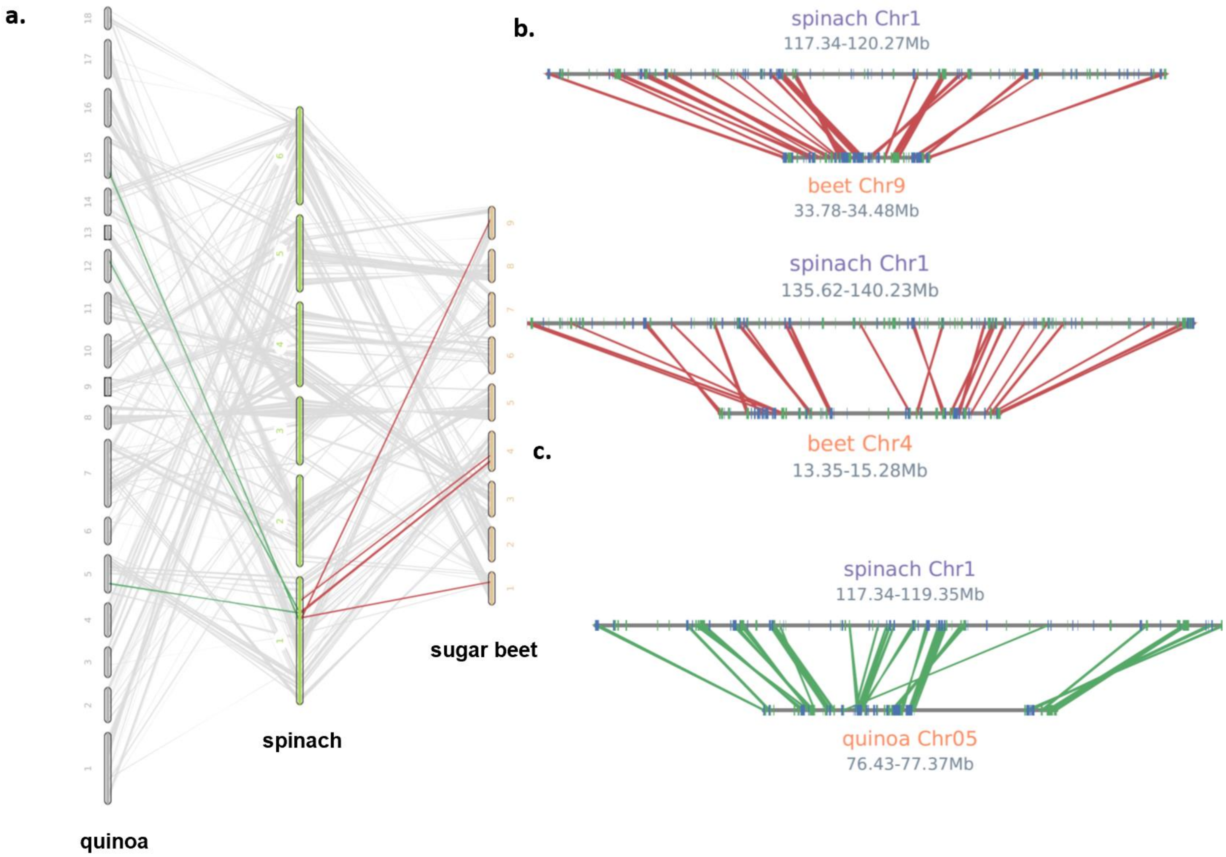

Fig 4.8 Intra-specific comparison between MSY and orthologous region

a. Genome-wide macro-synteny between spinach and quinoa, spinach and sugar beet. Grey lines connect orthologous genes. Green lines connect the collinear genes from MSY and corresponded region from quinoa; Red lines connect the collinear genes from MSY and sugar beet counterpart.

b. Micro-synteny between MSY from spinach Y chromosome and orthologous region from chr4 and chr9 of the sugar beet genome.

c. Micro-synteny between MSY from spinach Y chromosome and orthologous region from chr5 of the quinoa genome.

References

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

Akagi, T., Henry, I. M., Ohtani, H., Morimoto, T., Beppu, K., Kataoka, I., & Tao, R. (2018). A Y-Encoded Suppressor of Feminization Arose via Lineage-Specific Duplication of a Cytokinin Response Regulator in Kiwifruit. *Plant Cell, 30*(4), 780-795. doi: 10.1105/tpc.17.00787

Akagi, T., Henry, I. M., Tao, R., & Comai, L. (2014). A Y-chromosome–encoded small RNA acts as a sex determinant in persimmons. *Science, 346*(6209), 646-650.

Akagi, T., Pilkington, S. M., Varkonyi-Gasic, E., Henry, I. M., Sugano, S. S., Sonoda, M., . . . Tao, R. (2019). Two Y-chromosome-encoded genes determine sex in kiwifruit. *Nat Plants, 5*(8), 801-809. doi: 10.1038/s41477-019-0489-6

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

Akamatsu, T., Suzuki, T., & Uchimiya, H. (1998). Determination of male or female of spinach by using DNA marker. *Sakata no tane KK, Japan*.

Akiyama, Y., Conner, J. A., Goel, S., Morishige, D. T., Mullet, J. E., Hanna, W. W., & Ozias-Akins, P. (2004). High-resolution physical mapping in Pennisetum squamulatum reveals extensive chromosomal heteromorphism of the genomic region associated with apomixis. *Plant Physiology, 134*(4), 1733-1741.

Alonge, M., Soyk, S., Ramakrishnan, S., Wang, X., Goodwin, S., Sedlazeck, F. J., . . . Schatz, M. C. (2019). RaGOO: fast and accurate reference-guided scaffolding of draft genomes. *Genome Biol, 20*(1), 224. doi: 10.1186/s13059-019-1829-6

Arumuganathan, K., & Earle, E. (1991). Nuclear DNA content of some important plant species. *Plant molecular biology reporter, 9*(3), 208-218.

Ashraf, M., & Foolad, M. R. (2013). Crop breeding for salt tolerance in the era of molecular markers and marker-assisted selection. *Plant Breeding, 132*(1), 10-20.

Bellott, D. W., Skaletsky, H., Pyntikova, T., Mardis, E. R., Graves, T., Kremitzki, C., . . . Wilson, R. K. (2010). Convergent evolution of chicken Z and human X chromosomes by expansion and gene acquisition. *Nature, 466*(7306), 612-616.

BEMIS, W. P., & WILSON, G. B. (1953). A NEW HYPOTHESIS EXPLAINING THE GENETICS OF SEX DETERMINATION: In Spinacia oleracea L. *Journal of Heredity, 44*(3), 91-95. doi: 10.1093/oxfordjournals.jhered.a106370

Bergero, R., & Charlesworth, D. (2009). The evolution of restricted recombination in sex chromosomes. *Trends Ecol Evol, 24*(2), 94-102. doi: 10.1016/j.tree.2008.09.010

Bergero, R., Forrest, A., Kamau, E., & Charlesworth, D. (2007a). Evolutionary strata on the X chromosomes of the dioecious plant Silene latifolia: evidence from new sex-linked genes. *Genetics*.

Bergero, R., Forrest, A., Kamau, E., & Charlesworth, D. (2007b). Evolutionary strata on the X chromosomes of the dioecious plant Silene latifolia: evidence from new sex-linked genes. *Genetics, 175*(4), 1945-1954. doi: 10.1534/genetics.106.070110

Bolger, A. M., Lohse, M., & Usadel, B. (2014). Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics, 30*(15), 2114-2120.

Boualem, A., Fergany, M., Fernandez, R., Troadec, C., Martin, A., Morin, H., . . . Pitrat, M. (2008). A conserved mutation in an ethylene biosynthesis enzyme leads to andromonoecy in melons. *Science, 321*(5890), 836-838.

Boualem, A., Troadec, C., Kovalski, I., Sari, M. A., Perl-Treves, R., & Bendahmane, A. (2009). A conserved ethylene biosynthesis enzyme leads to andromonoecy in two cucumis species. *PLoS One, 4*(7), e6144. doi: 10.1371/journal.pone.0006144

Bowers, J. E., Bachlava, E., Brunick, R. L., Rieseberg, L. H., Knapp, S. J., & Burke, J. M. (2012a). Development of a 10,000 locus genetic map of the sunflower genome based on multiple crosses. *G3: Genes, Genomes, Genetics, 2*(7), 721-729.

Bowers, J. E., Nambeesan, S., Corbi, J., Barker, M. S., Rieseberg, L. H., Knapp, S. J., & Burke, J. M. (2012b). Development of an ultra-dense genetic map of the sunflower genome based on single-feature polymorphisms. *PloS one, 7*(12), e51360.

Bowers, J. E., Pearl, S. A., & Burke, J. M. (2016). Genetic mapping of millions of SNPs in safflower (Carthamus tinctorius L.) via whole-genome resequencing. *G3: Genes, Genomes, Genetics, 6*(7), 2203-2211.

Casselman, A. L., Vrebalov, J., Conner, J. A., Singhal, A., Giovannoni, J., Nasrallah, M. E., & Nasrallah, J. B. (2000). Determining the physical limits of the Brassica S locus by recombinational analysis. *The Plant Cell, 12*(1), 23-33.

Castillo, E. R., Marti, D. A., & Bidau, C. J. (2010). Sex and neo-sex chromosomes in Orthoptera: a review. *Journal of Orthoptera Research,* 213-231.

Chapman, J. A., Mascher, M., Buluç, A., Barry, K., Georganas, E., Session, A., . . . Oliker, L. (2015). A whole-genome shotgun approach for assembling and anchoring the hexaploid bread wheat genome. *Genome biology, 16*(1), 26.

Charlesworth, B., & Charlesworth, D. (1978). A model for the evolution of dioecy and gynodioecy. *The American Naturalist, 112*(988), 975-997.

Charlesworth, B., & Charlesworth, D. (2000). The degeneration of Y chromosomes. *Philos Trans R Soc Lond B Biol Sci, 355*(1403), 1563-1572. doi: 10.1098/rstb.2000.0717

Charlesworth, B., Sniegowski, P., & Stephan, W. (1994). The evolutionary dynamics of repetitive DNA in eukaryotes. *Nature, 371*(6494), 215-220.

Charlesworth, D. (2013a). Plant sex chromosome evolution. *Journal of experimental botany, 64*(2), 405-420.

Charlesworth, D. (2013b). Plant sex chromosome evolution. *Journal of experiemntal Botany*. doi: 10.1093/jxb/err31310.1093/jxb/ers322

Charlesworth, D. (2016). Plant Sex Chromosomes. *Annu Rev Plant Biol, 67,* 397-420. doi: 10.1146/annurev-arplant-043015-111911

Charlesworth, D., Charlesworth, B., & Marais, G. (2005). Steps in the evolution of heteromorphic sex chromosomes. *Heredity, 95*(2), 118.

Charlesworth, D., & David, S. G. (2004). The evolution of dioecy and plant sex chromosome systems *Sex determination in plants* (pp. 25-50): Garland Science.

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. *Nature Genetics, 39*(12), 1517-1521. doi: 10.1038/ng.2007.20

Consortium, I. B. G. S. (2012). A physical, genetic and functional sequence assembly of the barley genome. *Nature, 491*(7426), 711.

Cronk, Q. C. (2005). Plant eco-devo: the potential of poplar as a model organism. *New Phytol, 166*(1), 39-48. doi: 10.1111/j.1469-8137.2005.01369.x

D'Esposito, M., Graves, J., Kirby, P., Matarazzo, M., Ciccodicola, A., Rocchi, M., . . . Ventura, M. (2003). Complex Events in the Evolution of the Human Pseudoautosomal Region 2 (PAR2).

Dai, X., Hu, Q., Cai, Q., Feng, K., Ye, N., Tuskan, G. A., . . . Yin, T. (2014). The willow genome and divergent evolution from poplar after the common genome duplication. *Cell Res, 24*(10), 1274-1277. doi: 10.1038/cr.2014.83

Danecek, P., Auton, A., Abecasis, G., Albers, C. A., Banks, E., DePristo, M. A., . . . Sherry, S. T. (2011). The variant call format and VCFtools. *Bioinformatics, 27*(15), 2156-2158.

de Sousa, N., Carlier, J., Santo, T., & Leitão, J. (2013). An integrated genetic map of pineapple (Ananas comosus (L.) Merr.). *Scientia Horticulturae, 157*, 113-118.

Delph, L. F., Arntz, A. M., Scotti-Saintagne, C., & Scotti, I. (2010). The genomic architecture of sexual dimorphism in the dioecious plant Silene latifolia. *Evolution, 64*(10), 2873-2886.

Deng, C.-l., Qin, R.-y., Cao, Y., Gao, J., Li, S.-f., Gao, W.-j., & Lu, L.-d. (2013a). Microdissection and painting of the Y chromosome in spinach (Spinacia oleracea). *Journal of plant research, 126*(4), 549-556.

Deng, C., Qin, R., Gao, J., Cao, Y., Li, S., Gao, W., & Lu, L. (2012). Identification of sex chromosome of spinach by physical mapping of 45s rDNAs by FISH. *Caryologia, 65*(4), 322-327.

Deng, C., Qin, R., Gao, J., Cao, Y., Li, S., Gao, W., & Lu, L. (2013b). Identification of sex chromosome of spinach by physical mapping of 45s rDNAs by FISH. *Caryologia, 65*(4), 322-327. doi: 10.1080/00087114.2012.760879

Deng, C. L., Qin, R. Y., Cao, Y., Gao, J., Li, S. F., Gao, W. J., & Lu, L. D. (2013c). Microdissection and painting of the Y chromosome in spinach (Spinacia oleracea). *J Plant Res, 126*(4), 549-556. doi: 10.1007/s10265-013-0549-3

Dohm, J. C., Minoche, A. E., Holtgräwe, D., Capella-Gutiérrez, S., Zakrzewski, F., Tafer, H., . . . Reinhardt, R. (2014a). The genome of the recently domesticated crop plant sugar beet (Beta vulgaris). *Nature, 505*(7484), 546.

Dohm, J. C., Minoche, A. E., Holtgrawe, D., Capella-Gutierrez, S., Zakrzewski, F., Tafer, H., . . . Himmelbauer, H. (2014b). The genome of the recently domesticated crop plant sugar beet (Beta vulgaris). *Nature, 505*(7484), 546-549. doi: 10.1038/nature12817

Duangjai, S., Wallnöfer, B., Samuel, R., Munzinger, J., & Chase, M. W. (2006). Generic delimitation and relationships in Ebenaceae sensu lato: evidence from six plastid DNA regions. *American journal of Botany, 93*(12), 1808-1827.

Ellis, J., & Janick, J. (1960). THE CHROMOSOMES OF SPLNACIA OLERACEA. *American journal of Botany, 47*(3), 210-214.

Elshire, R. J., Glaubitz, J. C., Sun, Q., Poland, J. A., Kawamoto, K., Buckler, E. S., & Mitchell, S. E. (2011). A robust, simple genotyping-by-sequencing (GBS) approach for high diversity species. *PloS one, 6*(5), e19379.

Geraldes, A., Hefer, C. A., Capron, A., Kolosova, N., Martinez-Nunez, F., Soolanayakanahally, R. Y., . . . Cronk, Q. C. (2015). Recent Y chromosome divergence despite ancient origin of dioecy in poplars (Populus). *Mol Ecol, 24*(13), 3243-3256. doi: 10.1111/mec.13126

Grattapaglia, D., & Sederoff, R. (1994). Genetic linkage maps of Eucalyptus grandis and Eucalyptus urophylla using a pseudo-testcross: mapping strategy and RAPD markers. *Genetics, 137*(4), 1121-1137.

Gschwend, A. R., Yu, Q., Tong, E. J., Zeng, F., Han, J., VanBuren, R., . . . Paterson, A. H. (2012). Rapid divergence and expansion of the X chromosome in papaya. *Proceedings of the National Academy of Sciences, 109*(34), 13716-13721.

Hahn, M. W., Zhang, S. V., & Moyle, L. C. (2014). Sequencing, assembling, and correcting draft genomes using recombinant populations. *G3: Genes, Genomes, Genetics, 4*(4), 669-679.

Harkess, A., Zhou, J., Xu, C., Bowers, J. E., Hulst, R., Ayyampalayam, S., . . . Kakrana, A. (2017a). The asparagus genome sheds light on the origin and evolution of a young Y chromosome. *Nature communications, 8*(1), 1279.

Harkess, A., Zhou, J., Xu, C., Bowers, J. E., Van der Hulst, R., Ayyampalayam, S., . . . Kakrana, A. (2017b). The asparagus genome sheds light on the origin and evolution of a young Y chromosome. *Nature communications, 8*(1), 1279.

Harkess, A., Zhou, J., Xu, C., Bowers, J. E., Van der Hulst, R., Ayyampalayam, S., . . . Chen, G. (2017c). The asparagus genome sheds light on the origin and evolution of a young Y chromosome. *Nat Commun, 8*(1), 1279. doi: 10.1038/s41467-017-01064-8

Henry, I. M., Akagi, T., Tao, R., & Comai, L. (2018). One Hundred Ways to Invent the Sexes: Theoretical and Observed Paths to Dioecy in Plants. *Annual Review of Plant Biology, Vol 69, 69,* 553-575. doi: 10.1146/annurev-arplant-042817-040615

Hobza, R., Kubat, Z., Cegan, R., Jesionek, W., Vyskot, B., & Kejnovsky, E. (2015). Impact of repetitive DNA on sex chromosome evolution in plants. *Chromosome Research, 23*(3), 561-570.

Iizuka, M., & Janick, J. (1962). Cytogenetic analysis of sex determination in Spinacia oleracea. *Genetics, 47*(9), 1225.

Jaarola, M., Martin, R. H., & Ashley, T. (1998). Direct evidence for suppression of recombination within two pericentric inversions in humans: a new sperm-FISH technique. *The American Journal of Human Genetics, 63*(1), 218-224.

Jamilena, M., Mariotti, B., & Manzano, S. (2008). Plant sex chromosomes: molecular structure and function. *Cytogenetic and genome research, 120*(3-4), 255-264.

Janick, J. (1954). *A genetic study of the heterogametic nature of the staminate plant in spinach (Spinacia oleracea L.).* Paper presented at the Proc Am Soc Hort Sci.

Janick, J. (1955a). *Environmental influences on sex expression in monoecious lines of spinach.* Paper presented at the Proc. Amer. Soc. Hort. Sci.

Janick, J. (1955b). *Inheritance of sex in tetraploid spinach.* Paper presented at the Proc. Amer. Soc. Hort. Sci.

Janick, J. (1957). The effects of polyploidy on sex expression in spinach. *J. of Hered, 48,* 150-156.

Janick, J., Mahoney, D., & Pfahler, P. (1959). The trisomics of Spinacia oleracea. *Journal of Heredity, 50*(2), 47-50.

Janick, J., & Stevenson, E. (1955a). Genetics of the monoecious character in spinach. *Genetics, 40*(4), 429.

Janick, J., & Stevenson, E. C. (1955b). Genetics of the Monoecious Character in Spinach. *Genetics, 40*(4), 429-437.

Janousek, B., & Mrackova, M. (2010). Sex chromosomes and sex determination pathway dynamics in plant and animal models. *Biological journal of the Linnean Society, 100*(4), 737-752.

Jarvis, D. E., Ho, Y. S., Lightfoot, D. J., Schmockel, S. M., Li, B., Borm, T. J., . . . Tester, M. (2017). The genome of Chenopodium quinoa. *Nature, 542*(7641), 307-312. doi: 10.1038/nature21370

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, 102*(6), 533.

Khattak, J. Z., Torp, A. M., & Andersen, S. B. (2006). A genetic linkage map of Spinacia oleracea and localization of a sex determination locus. *Euphytica, 148*(3), 311-318.

Kim, K. E., Peluso, P., Babayan, P., Yeadon, P. J., Yu, C., Fisher, W. W., . . . Li, J. (2014). Long-read, whole-genome shotgun sequence data for five model organisms. *Scientific data, 1*, 140045.

Kondo, M., Hornung, U., Nanda, I., Imai, S., Sasaki, T., Shimizu, A., . . . Shimizu, N. (2006). Genomic organization of the sex-determining and adjacent regions of the sex chromosomes of medaka. *Genome research, 16*(7), 815-826.

Koren, S., Walenz, B. P., Berlin, K., Miller, J. R., Bergman, N. H., & Phillippy, A. M. (2017). Canu: scalable and accurate long-read assembly via adaptive k-mer weighting and repeat separation. *Genome research*, gr. 215087.215116.

Kudoh, T., Takahashi, M., Osabe, T., Toyoda, A., Hirakawa, H., Suzuki, Y., . . . Onodera, Y. (2018a). Molecular insights into the non-recombining nature of the spinach male-determining region. *Mol Genet Genomics, 293*(2), 557-568. doi: 10.1007/s00438-017-1405-2

Kudoh, T., Takahashi, M., Osabe, T., Toyoda, A., Hirakawa, H., Suzuki, Y., . . . Onodera, Y. (2018b). Molecular insights into the non-recombining nature of the spinach male-determining region. *Molecular Genetics and Genomics, 293*(2), 557-568.

Kurtz, S., Phillippy, A., Delcher, A. L., Smoot, M., Shumway, M., Antonescu, C., & Salzberg, S. L. (2004). Versatile and open software for comparing large genomes. *Genome biology, 5*(2), R12.

Löve, A. (1944). Cytogenetic studies on Rumex subgenus Acetosella. *Hereditas, 30*(1-2), 1-135.

Lan, T., Zhang, S., Liu, B., Li, X., Chen, R., & Song, W. (2006). Differentiating sex chromosomes of the dioecious Spinacia oleracea L. (spinach) by FISH of 45S rDNA. *Cytogenet Genome Res, 114*(2), 175-177. doi: 10.1159/000093335

Lewis, D. (1942). The evolution of sex in flowering plants. *Biological Reviews, 17*(1), 46-67.

Li, H., & Durbin, R. (2010). Fast and accurate long-read alignment with Burrows–Wheeler transform. *Bioinformatics, 26*(5), 589-595.

Li, L., Deng, C. H., Knabel, M., Chagne, D., Kumar, S., Sun, J., . . . Wu, J. (2017). Integrated high-density consensus genetic map of Pyrus and anchoring of the 'Bartlett' v1.0 (Pyrus communis) genome. *DNA Res, 24*(3), 289-301. doi: 10.1093/dnares/dsw063

Li, S.-F., Wang, B.-X., Guo, Y.-J., Deng, C.-L., & Gao, W.-J. (2018). Genome-wide characterization of microsatellites and genetic diversity assessment of spinach in the Chinese germplasm collection. *Breeding science, 68*(4), 455-464.

Li, Z., Wang, S., Tao, Q., Pan, J., Si, L., Gong, Z., & Cai, R. (2012). A putative positive feedback regulation mechanism in CsACS2 expression suggests a modified model for sex determination in cucumber (Cucumis sativus L.). *Journal of experimental botany, 63*(12), 4475-4484.

Liu, Z., Moore, P. H., Ma, H., Ackerman, C. M., Ragiba, M., Yu, Q., . . . Stiles, J. I. (2004). A primitive Y chromosome in papaya marks incipient sex chromosome evolution. *Nature, 427*(6972), 348.

Loptien, H. (1979). Identification of the sex chromosome pair in asparagus (Asparagus officinalis L.). *Z. Pflanzenzuchtg., 82*, 162-175.

Lukacsovich, T., & Waldman, A. S. (1999). Suppression of intrachromosomal gene conversion in mammalian cells by small degrees of sequence divergence. *Genetics, 151*(4), 1559-1568.

Ma, H., Moore, P. H., Liu, Z., Kim, M. S., Yu, Q., Fitch, M. M., . . . Ming, R. (2004). High-density linkage mapping revealed suppression of recombination at the sex determination locus in papaya. *Genetics, 166*(1), 419-436.

Ma, J., & Bennetzen, J. L. (2004). Rapid recent growth and divergence of rice nuclear genomes. *Proc Natl Acad Sci U S A, 101*(34), 12404-12410. doi: 10.1073/pnas.0403715101

Mackay, T. F. (2001). The genetic architecture of quantitative traits. *Annual review of genetics, 35*(1), 303-339.

Marcais, G., Delcher, A. L., Phillippy, A. M., Coston, R., Salzberg, S. L., & Zimin, A. (2018). MUMmer4: A fast and versatile genome alignment system. *PLoS Comput Biol, 14*(1), e1005944. doi: 10.1371/journal.pcbi.1005944

Matsunaga, S. (2006). Sex chromosome-linked genes in plants. *Genes & genetic systems, 81*(4), 219-226.

Matsunaga, S., Kawano, S., Michimoto, T., Higashiyama, T., Nakao, S., Sakai, A., & Kuroiwa, T. (1999). Semi-automatic laser beam microdissection of the Y chromosome and analysis of Y chromosome DNA in a dioecious plant, Silene latifolia. *Plant and Cell Physiology, 40*(1), 60-68.

Ming, R., Bendahmane, A., & Renner, S. S. (2011a). Sex chromosomes in land plants. *Annual Review of Plant Biology, 62*, 485-514.

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

Ming, R., VanBuren, R., Wai, C. M., Tang, H., Schatz, M. C., Bowers, J. E., . . . Biggers, E. (2015). The pineapple genome and the evolution of CAM photosynthesis. *Nature genetics, 47*(12), 1435.

Ming, R., Wang, J., Moore, P. H., & Paterson, A. H. (2007). Sex chromosomes in flowering plants. *Am J Bot, 94*(2), 141-150. doi: 10.3732/ajb.94.2.141

Mun, J.-H., Chung, H., Chung, W.-H., Oh, M., Jeong, Y.-M., Kim, N., . . . Lim, K.-B. (2015). Construction of a reference genetic map of Raphanus sativus based on genotyping by whole-genome resequencing. *Theoretical and applied genetics, 128*(2), 259-272.

Na, J.-K., Wang, J., & Ming, R. (2014). Accumulation of interspersed and sex-specific repeats in the non-recombining region of papaya sex chromosomes. *BMC Genomics, 15*(1), 335.

Neumann, P., Koblizkova, A., Navratilova, A., & Macas, J. (2006). Significant expansion of Vicia pannonica genome size mediated by amplification of a single type of giant retroelement. *Genetics, 173*(2), 1047-1056. doi: 10.1534/genetics.106.056259

Okazaki, Y., Takahata, S., Hirakawa, H., Suzuki, Y., & Onodera, Y. (2019). Molecular evidence for recent divergence of X- and Y-linked gene pairs in Spinacia oleracea L. *PLoS One, 14*(4), e0214949. doi: 10.1371/journal.pone.0214949

Onodera, Y., Yonaha, I., Masumo, H., Tanaka, A., Niikura, S., Yamazaki, S., & Mikami, T. (2011). Mapping of the genes for dioecism and monoecism in Spinacia oleracea L.: evidence that both genes are closely linked. *Plant cell reports, 30*(6), 965-971.

Otto, S. P., Pannell, J. R., Peichel, C. L., Ashman, T.-L., Charlesworth, D., Chippindale, A. K., . . . McAllister, B. F. (2011). About PAR: the distinct evolutionary dynamics of the pseudoautosomal region. *Trends in Genetics, 27*(9), 358-367.

Patterson, E. L., Saski, C. A., Sloan, D. B., Tranel, P. J., Westra, P., & Gaines, T. A. (2019). The Draft Genome of Kochia scoparia and the Mechanism of Glyphosate Resistance via Transposon-Mediated EPSPS Tandem Gene Duplication. *Genome Biol Evol, 11*(10), 2927-2940. doi: 10.1093/gbe/evz198

Perl-Treves, R. (1999). Male to female conversion along the cucumber shoot: approaches to studying sex genes and floral development in Cucumis sativus. *Sex determination in plants*, 189-215.

Peterson, B. K., Weber, J. N., Kay, E. H., Fisher, H. S., & Hoekstra, H. E. (2012). Double digest RADseq: an inexpensive method for de novo SNP discovery and genotyping in model and non-model species. *PloS one, 7*(5), e37135.

Piegu, B., Guyot, R., Picault, N., Roulin, A., Sanyal, A., Kim, H., . . . Panaud, O. (2006). Doubling genome size without polyploidization: dynamics of retrotransposition-driven genomic expansions in Oryza australiensis, a wild relative of rice. *Genome Res, 16*(10), 1262-1269. doi: 10.1101/gr.5290206

Poole, C. F., & GRIMBALL, P. C. (1939). Inheritance of new sex forms in Cucumis melo L. *Journal of Heredity, 30*(1), 21-25.

Pucholt, P., Ronnberg-Wastljung, A. C., & Berlin, S. (2015). Single locus sex determination and female heterogamety in the basket willow (Salix viminalis L.). *Heredity (Edinb), 114*(6), 575-583. doi: 10.1038/hdy.2014.125

Qian, W., Fan, G., Liu, D., Zhang, H., Wang, X., Wu, J., & Xu, Z. (2017). Construction of a high-density genetic map and the X/Y sex-determining gene mapping in spinach based on large-scale markers developed by specific-locus amplified fragment sequencing (SLAF-seq). *BMC genomics, 18*(1), 276.

Rastas, P. (2017). Lep-MAP3: robust linkage mapping even for low-coverage whole genome sequencing data. *Bioinformatics, 33*(23), 3726-3732.

Renner, S. S. (2014a). The relative and absolute frequencies of angiosperm sexual systems: dioecy, monoecy, gynodioecy, and an updated online database. *Am J Bot, 101*(10), 1588-1596. doi: 10.3732/ajb.1400196

Renner, S. S. (2014b). The relative and absolute frequencies of angiosperm sexual systems: dioecy, monoecy, gynodioecy, and an updated online database. *American Journal of botany, 101*(10), 1588-1596.

Renner, S. S., & Ricklefs, R. E. (1995). Dioecy and Its Correlates in the Flowering Plants. *American journal of Botany, 82*(5), 596-606. doi: Doi 10.2307/2445418

Ribera, A., Bai, Y., Wolters, A.-M. A., van Treuren, R., & Kik, C. (2020). A review on the genetic resources, domestication and breeding history of spinach (Spinacia oleracea L.). *Euphytica, 216*(3). doi: 10.1007/s10681-020-02585-y

Rice, P., Longden, I., & Bleasby, A. (2000). EMBOSS: the European molecular biology open software suite. *Trends in genetics, 16*(6), 276-277.

Rodriguez Del Rio, A., Minoche, A. E., Zwickl, N. F., Friedrich, A., Liedtke, S., Schmidt, T., . . . Dohm, J. C. (2019). Genomes of the wild beets Beta patula and Beta vulgaris ssp. maritima. *Plant J, 99*(6), 1242-1253. doi: 10.1111/tpj.14413

Ross, M. T., Grafham, D. V., Coffey, A. J., Scherer, S., McLay, K., Muzny, D., . . . Bentley, D. R. (2005). The DNA sequence of the human X chromosome. *Nature, 434*(7031), 325-337. doi: 10.1038/nature03440

Rubatzky, V. E., & Yamaguchi, M. (2012). *World vegetables: principles, production, and nutritive values*: Springer Science & Business Media.

Smit, A. F., & Hubley, R. (2010). RepeatModeler Open-1.0.

Sneep, J. (1982). The domestication of spinach and the breeding history of its varieties. *Euphytica (Suppl 2), 27*.

Spigler, R. B., Lewers, K. S., Johnson, A. L., & Ashman, T. L. (2010). Comparative mapping reveals autosomal origin of sex chromosome in octoploid Fragaria virginiana. *J Hered, 101 Suppl 1,* S107-117. doi: 10.1093/jhered/esq001

Spigler, R. B., Lewers, K. S., Main, D. S., & Ashman, T. L. (2008). Genetic mapping of sex determination in a wild strawberry, Fragaria virginiana, reveals earliest form of sex chromosome. *Heredity (Edinb), 101*(6), 507-517. doi: 10.1038/hdy.2008.100

Sun, X., Liu, D., Zhang, X., Li, W., Liu, H., Hong, W., . . . Zeng, H. (2013). SLAF-seq: an efficient method of large-scale de novo SNP discovery and genotyping using high-throughput sequencing. *PloS one, 8*(3), e58700.

Sunil, M., Hariharan, A. K., Nayak, S., Gupta, S., Nambisan, S. R., Gupta, R. P., . . . Srinivasan, S. (2014). The draft genome and transcriptome of Amaranthus hypochondriacus: a C4 dicot producing high-lysine edible pseudo-cereal. *DNA Res, 21*(6), 585-602. doi: 10.1093/dnares/dsu021

Takahata, S., Yago, T., Iwabuchi, K., Hirakawa, H., Suzuki, Y., & Onodera, Y. (2016). Comparison of spinach sex chromosomes with sugar beet autosomes reveals extensive synteny and low recombination at the male-determining locus. *Journal of Heredity, 107*(7), 679-685.

Tang, H., Zhang, X., Miao, C., Zhang, J., Ming, R., Schnable, J. C., . . . Lu, J. (2015). ALLMAPS: robust scaffold ordering based on multiple maps. *Genome biology, 16*(1), 3.

Tarailo-Graovac, M., & Chen, N. (2009). Using RepeatMasker to identify repetitive elements in genomic sequences. *Current protocols in bioinformatics, 25*(1), 4.10. 11-14.10. 14.

Temnykh, S., DeClerck, G., Lukashova, A., Lipovich, L., Cartinhour, S., & McCouch, S. (2001). Computational and experimental analysis of microsatellites in rice (Oryza sativa L.): frequency, length variation, transposon associations, and genetic marker potential. *Genome research, 11*(8), 1441-1452.

Tennessen, J. A., Govindarajulu, R., Liston, A., & Ashman, T. L. (2016). Homomorphic ZW chromosomes in a wild strawberry show distinctive recombination heterogeneity but a small sex-determining region. *New Phytol, 211*(4), 1412-1423. doi: 10.1111/nph.13983

Torres, M. F., Mathew, L. S., Ahmed, I., Al-Azwani, I. K., Krueger, R., Rivera-Nunez, D., . . . Malek, J. A. (2018). Genus-wide sequencing supports a two-locus model for sex-determination in Phoenix. *Nat Commun, 9*(1), 3969. doi: 10.1038/s41467-018-06375-y

Trebitsh, T., Staub, J. E., & O'Neill, S. D. (1997). Identification of a 1-aminocyclopropane-1-carboxylic acid synthase gene linked to the female (F) locus that enhances female sex expression in cucumber. *Plant Physiology, 113*(3), 987-995.

Vyskot, B., & Hobza, R. (2004). Gender in plants: sex chromosomes are emerging from the fog. *Trends Genet, 20*(9), 432-438. doi: 10.1016/j.tig.2004.06.006

Wadlington, W. H., & Ming, R. (2018a). Development of an X-specific marker and identification of YY individuals in spinach. *Theoretical and Applied Genetics, 131*(9), 1987-1994.

Wadlington, W. H., & Ming, R. (2018b). Development of an X-specific marker and identification of YY individuals in spinach. *Theor Appl Genet, 131*(9), 1987-1994. doi: 10.1007/s00122-018-3127-1

Walker, M. A., Pedamallu, C. S., Ojesina, A. I., Bullman, S., Sharpe, T., Whelan, C. W., & Meyerson, M. (2018). GATK PathSeq: a customizable computational tool for the discovery and identification of microbial sequences in libraries from eukaryotic hosts. *Bioinformatics, 34*(24), 4287-4289.

Wang, D., Zhang, Y., Zhang, Z., Zhu, J., & Yu, J. (2010). KaKs_Calculator 2.0: A Toolkit Incorporating Gamma-Series Methods and Sliding Window Strategies. *Genomics, Proteomics & Bioinformatics, 8*(1), 77-80. doi: 10.1016/s1672-0229(10)60008-3

Wang, J., Na, J.-K., Yu, Q., Gschwend, A. R., Han, J., Zeng, F., . . . Zhang, W. (2012). Sequencing papaya X and Yh chromosomes reveals molecular basis of incipient sex chromosome evolution. *Proceedings of the National Academy of Sciences, 109*(34), 13710-13715.

Wang, L., Ma, G., Wang, H., Cheng, C., Mu, S., Quan, W., . . . Zhang, Y. (2019). A draft genome assembly of halophyte Suaeda aralocaspica, a plant that performs C4 photosynthesis within individual cells. *Gigascience, 8*(9). doi: 10.1093/gigascience/giz116

Wang, Y., Wang, X., McCubbin, A. G., & Kao, T.-h. (2003). Genetic mapping and molecular characterization of the self-incompatibility (S) locus in Petunia inflata. *Plant molecular biology, 53*(4), 565-580.

Westergaard, M. (1958). The mechanism of sex determination in dioecious flowering plants *Advances in genetics* (Vol. 9, pp. 217-281): Elsevier.

Wu, J., Li, L.-T., Li, M., Khan, M. A., Li, X.-G., Chen, H., . . . Zhang, S.-L. (2014). High-density genetic linkage map construction and identification of fruit-related QTLs in pear using SNP and SSR markers. *Journal of experimental botany, 65*(20), 5771-5781.

Wu, X., Knapp, S., Stamp, A., Stammers, D. K., Jörnvall, H., Dellaporta, S. L., & Oppermann, U. (2007). Biochemical characterization of TASSELSEED 2, an essential plant short-chain dehydrogenase/reductase with broad spectrum activities. *The FEBS journal, 274*(5), 1172-1182.

Xie, W., Feng, Q., Yu, H., Huang, X., Zhao, Q., Xing, Y., . . . Zhang, Q. (2010). Parent-independent genotyping for constructing an ultrahigh-density linkage map based on population sequencing. *Proceedings of the National Academy of Sciences, 107*(23), 10578-10583.

Xu, C., Jiao, C., Sun, H., Cai, X., Wang, X., Ge, C., . . . Xu, Y. (2017a). Draft genome of spinach and transcriptome diversity of 120 Spinacia accessions. *Nature communications, 8*, 15275.

Xu, C., Jiao, C., Sun, H., Cai, X., Wang, X., Ge, C., . . . Wang, Q. (2017b). Draft genome of spinach and transcriptome diversity of 120 Spinacia accessions. *Nat Commun, 8*, 15275. doi: 10.1038/ncomms15275

Yamamoto, K., Oda, Y., Haseda, A., Fujito, S., Mikami, T., & Onodera, Y. (2014). Molecular evidence that the genes for dioecism and monoecism in Spinacia oleracea L. are located at different loci in a chromosomal region. *Heredity, 112*(3), 317.

Yang, Z., & Nielsen, R. (2000). Estimating synonymous and nonsynonymous substitution rates under realistic evolutionary models. *Molecular biology and evolution, 17*(1), 32-43.

Yin, T., DiFazio, S. P., Gunter, L. E., Zhang, X., Sewell, M. M., Woolbright, S. A., . . . Wang, M. (2008). Genome structure and emerging evidence of an incipient sex chromosome in Populus. *Genome research, 18*(3), 422-430.

Yu, Q., Hou, S., Feltus, F. A., Jones, M. R., Murray, J. E., Veatch, O., . . . Thimmapuram, J. (2008a). Low X/Y divergence in four pairs of papaya sex-linked genes. *The Plant Journal, 53*(1), 124-132.

Yu, Q., Navajas-Pérez, R., Tong, E., Robertson, J., Moore, P. H., Paterson, A. H., & Ming, R. (2008b). Recent origin of dioecious and gynodioecious Y chromosomes in papaya. *Tropical plant Biology, 1*(1), 49-57.

Zhang, H.-X., & Zeevaart, J. (1999). An efficient Agrobacterium tumefaciens-mediated transformation and regeneration system for cotyledons of spinach (Spinacia oleracea L.). *Plant cell reports, 18*(7-8), 640-645.

Zhang, J., Boualem, A., Bendahmane, A., & Ming, R. (2014). Genomics of sex determination. *Curr Opin Plant Biol, 18*, 110-116. doi: 10.1016/j.pbi.2014.02.012

Zhang, Q., Liu, C., Liu, Y., VanBuren, R., Yao, X., Zhong, C., & Huang, H. (2015). High-density interspecific genetic maps of kiwifruit and the identification of sex-specific markers. *DNA Research, 22*(5), 367-375.

Zhou, Q., Miao, H., Li, S., Zhang, S., Wang, Y., Weng, Y., . . . Gu, X. (2015). A sequencing-based linkage map of cucumber. *Molecular plant, 8*(6), 961-963.

Zhou, R., Macaya-Sanz, D., Carlson, C. H., Schmutz, J., Jenkins, J. W., Kudrna, D., . . . DiFazio, S. P. (2020). A willow sex chromosome reveals convergent evolution of complex palindromic repeats. *Genome Biology, 21*(1). doi: 10.1186/s13059-020-1952-4

Zuccolo, A., Sebastian, A., Talag, J., Yu, Y., Kim, H., Collura, K., . . . Wing, R. A. (2007). Transposable element distribution, abundance and role in genome size variation in the genus Oryza. *BMC Evol Biol, 7*, 152. doi: 10.1186/1471-2148-7-152