ANALYSIS OF NATURAL VARIATION IN RECOMBINATION IN MAIZE

A Dissertation

Presented to the Faculty of the Graduate School

of Cornell University

In Partial Fulfillment of the Requirements for the Degree of

Doctor of Philosophy

by

Gaganpreet Kaur Sidhu

January 2012

ANALYSIS OF NATURAL VARIATION IN RECOMBINATION IN MAIZE

Gaganpreet Kaur Sidhu, Ph. D.

Cornell University 2012

Meiotic recombination is a major source of genetic variation in plants. Understanding factors that affect the recombination rates in plants is important for general genetic studies on meiosis as well as for plant breeding efforts. To examine how maize evolution and domestication affected meiotic recombination genes, I examined evolution patterns in eleven genes controlling key recombination pathway steps in a diverse set of maize inbred lines and several teosinte accessions. Even though meiotic recombination genes generally exhibit high sequence conservation expected in a pathway controlling a key cellular process, I identified several different selection modes in the eleven genes. Adaptive evolution signatures were found in about half of the examined genes in maize. Interestingly, in Balsas teosinte, the closest wild relative of maize, different and fewer genes showed adaptive evolution patterns than in maize. Changes in relatively few amino acid residues were responsible for the adaptive evolution signatures in maize and teosintes. Through protein structure predictions, I found that several of the amino acid residues identified as targets of selection are likely to induce functional changes in their proteins. I hypothesize that the evolutionary changes in the recombination pathway may have contributed to the successful domestication of maize and its expansion to new cultivation areas. In another study to understand the factors that affect recombination rates in plants, I

analyzed natural variation in crossover rates in maize during meiosis. I analyzed the numbers of chiasmata in thirteen randomly chosen parent inbreds lines from the NAM (nested association mapping) population. I hypothesized that genetic differences in the inbreds may affect the number of crossovers, which directly translates to natural variation in recombination frequency in these inbreds. After analyzing chiasmata numbers in maize, I found that the inbreds were significantly different from each other with respect to the average chiasma numbers per cell. These findings indicate that there are genetic regulatory elements in these inbreds, which lead to the observed differences in chiasma counts. It is possible to identify these regulatory elements and understand their mode of action, which in future, would enable the manipulation of recombination frequencies.

BIOGRAPHICAL SKETCH

The author was born in Khai Pheme Ki, a small village in Punjab, in North Western part of India to S. Dilbag Singh and Smt. Sukhwinder kaur. Gagan attended Govt. high School in Khai Pheme Ki for her initial education. She later attended DAV College, Abohar for 11^{th} and 12^{th} grade. She took Biology, Physics and Chemistry as her subjects of concentration for these two years. This provided her with a great learning experience and built the foundation for a deep interest in genetics. She later completed her Bachelors with Honors in Plant breeding and Genetics from Punjab Agricultural University, Ludhiana. Before coming to Cornell for Ph.D., she completed a Master's degree in Washington State University.

Author feels that her experience at Cornell University has taught her a lot both personally and professionally. It has helped her realize her passions and dreams. She thanks everyone who have been a part of her life during this time.

Author is forever indebted to her parents, brother and sister for supporting her in her decisions to pursue her dreams. Her parents are very happy and proud that she has competed a Ph.D.

To my parents, brother and sister

With love

ACKNOWLEDGMENTS

My Ph.D. dissertation work would not have been possible without the support of my mentors, friends and family over the past five years. I express my sincere gratitude to all of them.

First and foremost, I am very thankful to Dr. Wojciech Pawlowksi for giving me an opportunity to work in his group. It was a great learning experience. This dissertation would not have been possible without his expert guidance. I am especially grateful for his patience and encouragement during the many ups and downs of my research work.

I would also like to thank my committee members Dr. Eric Alani, Dr. Susan McCouch and Dr. Jian Hua for many helpful discussions throughout my time at Cornell. As a committee, they provided me with a great support system, which helped me finish my research efficiently. Their insightful comments and guidance helped me improve this dissertation.

I am also very thankful to Dr. Teresa Pawlowska for her guidance and support throughout my degree. I am sincerely thankful to both Teresa and Wojciech for providing support and encouragement in many respects other than research.

I am thankful to past and current members of Pawlowski lab. I have had the great fortune to be working with a great group of people in the lab, who guided me throughout my degree.

A very special thanks to Celestia Fang for making the ever tedious work of counting chiasma easy for me. I wish her happiness and success in the future.

I am also thankful to my teachers in my high school back home especially Mrs. Chandra Prabha and Smt. Sukhwinder kaur for believing in me and guiding me.

I would also like to thank past and present members of Cornell Fashion Collective and Proff. Anita Racine for helping me get in touch with my artistic side. It was a great feeling to showcase my designs in the annual fashion shows here at Cornell. I will always be grateful for this experience.

While at Cornell, I have been surrounded by a great group of friends. I am thankful to all of you for being there for me in happy and difficult times. I am grateful for the moments I have spent with you all.

I am tremendously thankful to my family for their endless encouragement and support. I would not be where I am today without their love. I am thankful to my mom, dad, Aman and Beenu for providing me with all the support I asked for. I can never thank my family enough for their optimism, and for their belief and trust in me. This has made them very proud and I dedicate this research work to my parents, brother and sister.

TABLE OF CONTENTS

CHAPTER 1

INTRODUCTION

Meiosis is the single most important process to sexually reproducing organisms for both genetic diversity and the continuation of species. Meiotic cell division is hallmarked by three key processes: homologous chromosome pairing, chromosome synapsis, and homologous recombination. Most investigations into meiosis focus on the mechanisms of recombination since it is a powerful tool for generating genetic diversity.

1.1 Meiosis Overview

Meiosis is preceded by DNA replication, which leads to generating sister chromatids, just like mitosis. Unlike mitosis, meiosis distributes a single copy of every chromosome to four different nuclei at the end of the division. It is achieved through a single round of DNA replication followed by two nuclear divisions. The diploid chromosome complement is precisely halved via this process (*1*).

Meiosis consists of two consecutive nuclear divisions, meiosis-I and meiosis-II. Meiosis I is a reductional division, leading to halving of the chromosome number, while meiosis II is similar to a typical mitosis. Both meiosis-I and II are delineated by four substages prophase, metaphase, anaphase, and telophase. Prophase I is further broken down into five substages: leptotene, zygotene, pachytene, diplotene, and diakinesis.

Events taking place in leptotene include condensation of chromosomes and installation of axial element proteins along the length of each chromosome. Recombination is also initiated at this stage with the formation of double-strand breaks (DSBs) in chromosomal DNA (*2*).

The transition from leptotene to zygotene brings formation of the telomere bouquet, which is thought to facilitate homologous chromosome pairing (*3*). The telomere bouquet is a structure formed during early meiosis by congregation of telomeres on the nuclear envelope, which resembles a bouquet of flowers. The behavior of telomeres as they gather together fits the model of their random attachment to the nuclear envelope, followed by their movement along the inner surface of the nuclear envelope until they all converge (*4, 5 2001*). The bouquet is thought to contribute to chromosome pairing based on the observation that bouquet formation immediately precedes homologous chromosome pairing (*6*). The function of the bouquet, as assessed by analysis of bouquet-specific mutants, *pam1* in maize, *ndj1*/*tam1* in budding yeast, and several mutants in fission yeast, including *taz1*, and *rap1*, is required for efficient pairing as well as timely initiation of synapsis (*7-12*). However, since some mutants deficient in homologous pairing are able to form bouquets with few defects, and not all species have bouquets, it is clear that the bouquet may facilitate pairing of chromosomes but is not sufficient or universally necessary for the process.

During zygotene, a proteinaceous synaptonemal complex (SC) is laid between the homologous chromosomes and holds them together until late prophase. SCs are comprised of two lateral elements and a central region, which contains the central

element. The lateral (*13*)elements correspond to the chromosome axes. Transverse filaments lie across the central region to create a striated, zipper-like appearance. SCs are thought to assemble by a two step process: nucleation, by installing central region proteins at sites where homolog axes are very closely paired, followed by polymerization between and along the homolog axes. SCs are important for the normal formation of crossovers (COs) (*14*).

In budding yeast, it has been found that some SC proteins exhibit a peculiar pattern of distribution. The Zip1 protein, which forms the central region of SC, localizes along pachytene chromosomes with uneven abundance with regions of strong Zip1 staining interrupted by regions of weak Zip1 staining (*15, 16*). Axial element components Hop1 and Red1 also exhibit regions of intense staining interspersed by regions of less intense staining that are identical for the two proteins (*17*). These alternate intensity patterns occur all along the length of the SC but are more abundant in certain regions compared to others (*17*). Co-visualization of Zip1 and Hop1, at equivalent intensity levels, reveals that the two proteins exhibit distinct and often complementary staining patterns. (*17*). It has been speculated that the differential loading patterns of Zip1 and Hop1 might be involved in designating CO sites (*17*).

Also during zygotene, early recombination nodules (EN) become visible. ENs are electron dense cytological structures ~100 nm in diameter that contain recombination proteins. ENs are associated with axial elements and to a greater degree with the synaptonemal complex (*18*).

Pachytene displays completely homologously paired chromosomes. A subset of early nodules graduate to cytological structures called late nodules (LNs), which define crossover events (*19, 20*). LNs are approximately equal in number to crossovers. The progression of ENs to LNs is concluded by the appearance of chiasmata at diplotene. Chiasmata are the sites of actual physical connection between homologous chromosomes and can be easily visualized under the light microscope during the diplotene and diakinesis stages of meiosis. Chiasmata generally occur at a rate of 1 per chromosome arm.

At diplotene, chromosomes condense further and the central element of the synaptonemal complex is released except for the chiasmata.

1.2 Recombination

Meiotic recombination occurs during the prophase stage of meiosis I. It leads to formation of COs and non-crossovers (NCOs), creating genetic diversity (Fig. 1). Meiotic recombination also plays an important structural role in meiosis. Chiasmata, formed as a result of crossing-over ensure proper segregation of chromosomes into daughter cells. Genetic diversity generated as a result of meiotic recombination is a driving force of evolution.

Meiotic recombination starts with the formation of double-strand breaks (DSBs) in chromosomal DNA during the leptotene stage of prophase I (Fig. 1). Repair of DSBs coincides with SC formation and is required for normal pairing and synapsis in fungi, mammals, and plants (*21, 22 2001*), whereas in *Caenorhabditis elegans* and

Drosophila melanogaster, homologue pairing and SC formation are not dependent on meiotic recombination (*23, 24*).

The enzyme SPO11, which is a DNA topoisomerase-like protein, and is highly conserved across species, is responsible for the creation of DSBs. A failure to create DSBs results in absence of recombination. The SPO11 protein itself is not vital to the process beyond its mechanism, as *spo11*$^{-}$/*spo11*$^{-}$ mutants can be phenotypically restored by induction of DSBs by UV in mice (*25*). One SPO11 molecule is required per strand of DNA, thus two SPO11 molecules are involved in each DSB event. Topoisomerases change DNA by transiently breaking one or both strands, passing the unbroken DNA strand or strands through the break, and repairing the break. The broken ends of DNA are covalently linked to the enzyme. SPO11 is similarly attached to the DNA when it forms double-strand breaks during meiosis. The active site of SPO11 contains a tyrosine, which is responsible for the association of the protein with DNA to promote break formation (*26*).

SPO11 is required for initiating recombination as well as initiating the process of homologous chromosome pairing through DSB formation. *spo11* mutants do not undergo recombination or homologous pairing, however similar to the restoration of the recombination phenotype, DSBs genereated by irradiation can restore homologous pairing as well (*25*).

DSBs are repaired during zygotene. The repair process involves several recombination proteins. First, the MRN complex, which includes three proteins, MRE11, RAD50, and NBS1 resects the DSBs leading to formation of 3' overhangs (*27-30*) The MRN complex is highly conserved and plays a key role in sensing,

processing and repair of DSBs. MRE11 has a DNA binding domain and intrinsic endo and exonuclease activity. In vivo, MRE11 exists as a complex with RAD50 where both of these proteins form dimers and each molecule of MRE11 and RAD50 bind to each other. The NBS1 protein binds to this core complex via interactions with Mre11 to form a hexamer complex with two molecules each of MRE11, RAD50 and NBS1. MRE11 is responsible for bringing the two RAD50 molecules close together and it also promotes their ATPase activity (*31*). The MRN complex holds the MRE11 proteins in the center while the two Rad50 molecules are present on the periphery. The conformational changes in this complex are brought about by interactions between the MRE11 helix-loop-helix and the RAD50 coiled coil domains. These interactions create a clamp, which has very high DNA binding activity (*32*). While NBS1 stimulates the DNA binding and nuclease activities of the MRN complex, it does not itself possess a known enzymatic activity. Rather, NBS1 contributes to DSB repair primarily by mediating protein–protein interactions at DNA breakage sites (*33*). The ssDNA ends generated by the MRN complex are coated by two recA-like proteins, RAD51 and DMC1, catalyzing the invasion of homologous double-strand DNA (dsDNA) region resulting in single end invasion (SEI) events (Fig. 1). RecA proteins form right-handed helical filaments known as nucleoprotein filaments (*34*). The RAD51 protein forms a nucleofilament with ATP-dependent strand-exchange activity, which, has been shown to have roles in meiosis, somatic homologous recombination, and DSB repair. DMC1 also forms a nucleofilament, in cooperation with RAD51, but its function is confined to meiosis (*35*).

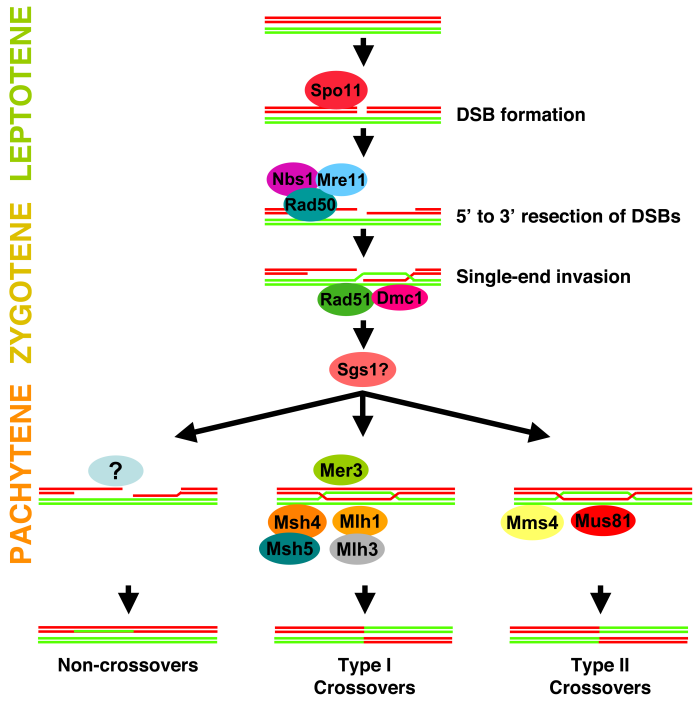


Figure 1.1: Diagram showing the main steps of the meiotic recombination pathway. Each chromosome is a different color. For simplicity, only one chromatid is shown for each chromosome. The diagram is based on data available in plants, as well as data from other systems that are likely applicable to plants. Question marks denote proteins whose identities or roles have not been yet unequivocally established.

At the end of recombination, DSBs are repaired either as COs or NCOs. In wild type maize, more than 500 DSBs are formed, leading to formation of nearly 20 COs, with the majority of DSBs being repaired as NCOs. Formation of COs involves resolution of double-Holliday junctions. Formation of NCOs is thought to occur through an independent pathway, which involves a synthesis-dependent strand annealing mechanism (*36*).

CO formation is regulated by a phenomenon known as CO interference, which prevents formation of multiple crossovers in the same chromosome region (*37, 38*). In plants, as well as yeast and mammals, there are two main classes of COs: type I COs that are subject to interference, and type II COs that are not (*39-45*). Type I CO events occur in a non-random pattern. About 85% of COs are Type I in *Arabidopsis thaliana*. The MSH4 and RCK/MER3 recombination proteins are involved in this pathway (*42, 46*). Type II COs occur with a random distribution. These COs are controlled by a pathway involving proteins MUS81 and MMS4/EME1 (*47, 48*).

1.3 Origin of Meiosis and meiotic Recombination

There are two different hypotheses on the origin of meiosis and recombination in eukaryotes. One of them states that meiosis originated from sexual transformation in bacteria (*49*). This hypothesis is supported by several pieces of evidence. First, RAD51 and DMC1 belong to a group of RecA-like proteins, which play crucial roles in recombination as well as during bacterial transformation. Bacterial transformation is the process by which bacterial cells take up foreign DNA. Second, the ancestors of eukaroytes were able to perform sexual transformation. Third, the core genes involved in meiotic recombination were present very early on during eukaryotic evolution.

The other hypothesis is that meiosis originated from mitosis (*50*). Meiosis and mitosis differ from each other by the presence of several key events that are exclusive to meiosis: chromosome pairing, recombination between homologous chromosomes, reductional division during first division of meiosis, and the absence of chromosome replication during second division of meiosis. It has been proposed that although the

simultaneous appearance of these novel steps in meiosis seems impossible, their step-by-step appearance as a result of natural selection of separate mutations also seems highly problematic. Therefore, a combination of mitosis and transformation leading to evolution of meiosis seems like a more probable scenario

It has been speculated that some gene products that evolved initially for use in mitotic cell division were later co-opted to be used in meiosis as well. Thus, the idea that meiosis arose from transformation and the idea that meiosis arose from mitosis are not necessarily mutually exclusive. However, the assumption that meiosis originated from mitosis or only after mitosis was fully established, would mean that there would have been an extended period when there was no meiosis, and therefore no sex, in eukaryotes. This would be the period when mitosis was still evolving. This assumption appears to be contradicted by evidence that the basic machinery for meiosis was present very early in eukaryote evolution (*13*).

Another discovery that suggests that meiotic recombination and thus meiosis itself emerged very early in eukaryotic evolution was the discovery of Dmc1 in several species of *Giardia*, one of the earliest forms of protists (basal eukaryotes) (*51*). Thus, it seems that the origin of meiosis is either based on bacterial transformation or a combination of mitosis and transformation.

1.4 Meiotic recombination versus somatic recombination

Meiotic recombination leads to segregation of homologous chromosomes by creating a physical connection between the chromosomes (*22, 52, 53*). The essential similarity between meiotic and somatic recombination processes is that they are both forms of

DSB repair. RAD51, a RecA protein is involved in the repair of DSBs in both meiosis and in somatic cells.

There are several differences between meiotic and somatic recombination. First, meiotic recombination occurs most of the time between homologous chromosomes while somatic recombination occurs mostly between sister chromatids. Second, meiotic recombination is specialized to generate COs and NCOs, leading to formation of recombinants, while mitotic recombination does not lead to formation of recombinants. Third, the programmed DSB formation process during meiosis at recombination hotspots induces 100–1000 fold more recombination relative to that in somatic cells (*54*). Fourth, crossover's distribution on chromosomes is non-random. Each bivalent receives at least one obligate crossover, which is essential for proper segregation of homologous chromosomes during first meiotic division.

1.5 Recombination Hotspots

Recombination does not occur uniformly on chromosomes. It has been well documented that recombination occurs predominantly in regions known as 'recombination hotspots' (*55*). Genome-wide patterns of DSB and CO distribution have been well studied in yeast and CO distribution patterns have been studied extensively in mammals (*56-59*). Recombination hotspots in mammals are typically 1-2 kb in size and are mostly located upstream of genic regions (*60, 61*). These hotspots are surrounded by much larger regions, which are devoid of recombination (*60, 61*). In humans, close to 80% of all recombination is believed to occur in as little as 10-20%

of the total genome space sequence (*62*). Recombination hotspots have also been observed in various plant species such as wheat, Arabidopsis, rice, and maize (*63*).

1.5.1 Factors affecting recombination event distribution

Data from a number of species have identified several factors that affect recombination rates, although so far no universal principles have emerged. Analyses of the *a1-sh2* region on maize chromosome 3 showed that most COs were resolved in or near genes. The 140-kb a1-sh2 interval of the maize genome contains at least four genes (a1, yz1, x1, and sh2). Physical positions of 101 meiotic recombination breakpoints were mapped in this interval and were found to be non-uniformly distributed across the interval (*64*). These breakpoints were concentrated within three recombination hotspots. Two of these recombination hotspots are in genic regions (a1 and yz1) and one is in a non-genic region. Results from this study suggested that not all hotspots are genes and indicate that not all genes are hotspots (*64*).

No correlation between CO hotspot location and genes has been observed in a survey of CO frequency on Arabidopsis chromosome 4 (*65*). 71 single nucleotide polymorphisms (SNPs) covering the entire chromosome 4 of Arabidopsis thaliana were examined using a population of 702 F2 plants. Recombination rates varied along the chromosome from 0 cM/Mb near the centromere to 20 cM/Mb on the short arm next to the NOR region, with chromosome average of 4.6 cM/Mb. CO rates were negatively correlated with the GC content. This is in contrast to what has been reported in other eukaryotes (*56, 59, 66-68*). COs also significantly correlate with the

density of single repeats and the CG ratio, but not with genes, pseudogenes, transposable elements, or dispersed repeats.

In mammals, DNA sequence polymorphisms at several recombination hotspots have been shown to affect the overall recombination rates (*61*) (*60*) (*62, 69*). Analyses of mammalian recombination hotspots led to identification of DNA sequence motifs that may cause specific regions to become hotspots. A 13 bp degenerative sequence (CCNCCNTNNCCNC) has been found in at least 41% of recombination hotspots in humans. However, this sequence does not directly control the activity of hotspots, although it is crucial for binding of a protein PRDM9, which acts as a trans-acting factor that controls hotspot activation in mice and humans (*58, 59*). PRDM9 contains a protein-protein binding domain, a PR/SET domain that can trimethylate histone 3 lysine 4 (H3K4), and has an array of 8-16 zinc fingers (*70*). These studies showed that specific patterns of histone tri-methylation that are known marks of open (active) chromatin are very strongly associated with the locations of meiotic DSB hotspots (*58, 59*).

The overall pattern of recombination in an organism reflects the aggregate behavior of its individual hotspots. In addition to this, recombination is affected by trans-regulating factors, which operate at the chromosomal level, including sex, and the genetic background. It has been reported in several species that sexes differ from each other with respect to their overall recombination rates. Genetic maps are much longer in females than males in humans and mice, while genetic maps are smaller for females in Arabidopsis (*71-74*).

There is also strong evidence of an effect of the genetic background on the rates of genome-wide recombination rates. A study on four strains of mouse reported that the average number of crossovers (studied by analyzing chromosomal foci of the MLH1 recombination protein during meiosis) per male meiosis vary from 21.5 to 24.9. (*75*). Natural variation of crossover number has also been studied in Arabidopsis (*76*). In this study, chiasma numbers were analyzed for eight Arabidopsis accessions. There was no significant variation in mean chiasma frequency between plants within accessions, but there was significant variation between accessions with mean CO frequency per cell ranging from 7.9 to 9.24 (*76*). The analysis also revealed that the pattern of chiasma distribution between arms and among chromosomes is different among the accessions (*76*).

1.5.2 Inheritance and evolution of recombination of hotspots

There has been lot of speculation on the stability of hotspots through generations. Recombination hotspots are thought to be evolutionary unstable since alleles with high recombination frequency are constantly being replaced by alleles with low recombination frequency through recombination events (a phenomenon known as a 'hotspot paradox'). Comparisons of hotspot locations between humans and chimpanzees, has confirmed this hypothesis (*77 2005*). In this study, fine-scale comparison of hotspot conservation was made in populations of humans and chimpanzees by studying two regions spanning a total of 14 Mb. Despite 99% DNA sequence level similarity, only 3 out of 39 inferred hotspots matched between the two

species (*77, 78*). There was no correlation between fine-scale recombination rates between the two species (*77, 78*).

On the other hand, an analysis of two yeast species, *Saccharomyces paradoxus* and *Saccharomyces cerevisiae*, showed considerable overlap of CO hotspot locations between the two species, despite the fact that they are at least 10 times more divergent from each other than humans are from chimpanzees. This unexpected result was speculated to be caused by the low frequency of sex and outcrossing in these species. The low level of sexual reproduction frequency might have acted to reduce the population genetic effect of biased gene conversion (*57*).

The identification of PRDM9 as a trans regulator of hotspot activity in mice and humans has shed new light on the phenomenon of 'hotspot paradox'. Studies have shown that the PRDM9 sequence is highly polymorphic in mice and mammals and several other species. The variation in PRDM9 is concentrated in the three zinc finger domains that specify its DNA sequence binding activity. It has also been noticed that the 13 bp degenerative DNA sequence motif in human hotspots is present in several other regions in the genome that are not recombination hotspots. Together, these findings indicate that hotspots are constantly evolving.

1.6 Natural and artificial selection and the recombination pathway

The main steps and events in the meiotic recombination pathway, as well as proteins that facilitate them, are highly conserved across species. The overall conservation of the recombination pathway suggests that its components are subject to predominantly purifying selection. Purifying selection works to stabilize alleles in a population by

removing deleterious mutations. This expectation was partially confirmed by a study of a set of meiotic genes in two species of Drosophila: *Drosophila melanogaster* and *D. simulans* (*79*). In this study, 33 genes involved in meiosis and meiosis-related processes, such as, chromosome segregation, achiasmate segregation, crossover regulation, double-strand-break formation, heterochromatin binding, recombination and/or repair, sister-chromatid cohesion, spindle assembly, and telomere maintenance, were examined for patterns of selection. Most genes were found to be subject to purifying selection. However, in addition to evidence of purifying selection, this study also revealed patterns of polymorphisms indicative of positive selection in six out of thirty three genes examined (*79*).

The best examples of selection acting on whole recombination pathway in plants could be found in the plant species, which have been domesticated. There have been several speculations on how domestication would affect recombination rates but there are two main hypotheses, which predict recombination rates in populations undergoing domestication. The first hypothesis predicts an increase in recombination rates through domestication (*80, 81*). Theoretical predictions (*82*)and empirical studies (*83*) also indicate that populations experiencing directional or strong selection pressures are likely to evolve increased recombination rates. The second hypothesis argues that increased recombination rates serve as a pre-adaptation to domestication (*84*) by increasing the response to strong selection pressure. These two hypotheses were tested in a study by Ross-Ibarra (*85*). This study used chiasma frequencies available from cytogenetic literature from both wild and domesticated species and rejected the second hypothesis. The chiasma numbers were higher in cultivated

species when compared to wild types (*85*), proving the hypothesis that selection acts on recombination pathway so that the populations experiencing selection evolve higher recombination rates.

A further confirmation of this hypothesis would be identifying recombination genes that are subject to selection pressure during domestication. Several methods have been developed over the years to detect presence of selection based on DNA sequence variation (*86*). One group of methods, which includes the often-used Tajima's D (*87*) and Fu and Li's D and F tests (*88*), is based on assessing the frequency spectrum of allelic polymorphisms in the population. Another group, which includes the McDonald-Kreitman (*89*) and Hudson-Kreitman-Aguade (*90*) tests, compares the patterns of sequence polymorphisms within the species with the patterns of sequence divergence between species. Finally, a number of methods that are used specifically to detect selection within the gene-coding region, assess the ratio of non-synonymous to synonymous nucleotide substitution rates (*91-93*). These methods are based on the premise that under neutral evolution non-synonymous and synonymous substitutions should occur at the same frequencies, while under selection these rates should differ from each other (*86*). This diversity of methods is a testament to the complexity of the processes that contribute to adaptive evolution. The different methods for identifying selection patterns often produce conflicting results from the same data sets; in some studies the overlap between different methods have been shown to be minimal or none (*94-96*). These discrepancies are likely caused by the fact that different methods are more sensitive to different types of selection events. The ability to detect selection patterns in DNA sequence data depends on the strength,

timing and frequency of the selection sweeps, constraints to nucleotide substitutions within the locus, the type of selection, the allele frequency before the selection sweep, and finally, on whether the locus is itself the target of selection or is "hitchhiked" because of selection in neighboring genes (*86, 95-97*)

1.7 Maize as a model system for genetic research

Maize is an excellent system to study natural variation for complex traits because of its high level of intra-species diversity. Maize exhibits more genetic diversity than any other genetic model system. Two average inbreds of maize are as different from each other at the DNA sequence level as chimpanzees are from humans (*98*).

Several studies genetic polymorphism in maize focused on examining DNA single nucleotide polymorphism (SNPs). A study on SNP variation on maize chromosome 1 using 16 maize landraces and nine inbred lines revealed that a SNP is present in the maize genome every 104 base pairs (*99*). In a more extensive study, a haplotype map of maize genome was generated using several million sequence polymorphisms in 27 diverse inbred lines (*100*). This study reported presence of a polymorphism every 44 base pairs (*100*).

Maize genomic diversity has also been explored in detail using array-based genomic hybridization between maize inbreds B73 and Mo17 (*101*). In this study, extensive structural variations were found between the two inbred lines. There were several hundred variations based on copy number differences between B73 and Mo17 and several thousand variations were present-absent variations (*101*). The most notable one is a ~2 Mb region on chromosome 6, which is present in B73 but absent in

Mo17. This region contains 24 annotated single copy genes (*101*). It was proposed that these haplotype variations might contribute to heterosis and the observed phenotypic variation among inbreds in maize (*101*).

The maize genome is 2500 Mb in size and is highly repetitive. Maize is an ancient allotetraploid, which arose through hybridization of two ancestors about 5 million years ago and these ancestors had diverged from a common ancestor about 12 million years ago (*102, 103*). This hybridization event lead to a whole genome duplication in maize. Then maize gradually became diploid by losing nearly 50% of one of the progenitor's genes (*104-108*). Analysis of genome architecture revealed that over 85% of the maize genome consists of repetitive DNA, mostly transposable elements (*109*). The most abundant form of transposons in maize are LTR (Long Terminal Repeat) retrotansposons, which, range from ~100 bp to over 5 kb in size (*110*). Maize chromosomes have large repetitive peri-centromeric regions, ranging from 60 to 113 Mbp in size, which are relatively devoid of recombination (*100, 111*).

1.8 Maize Domestication

Modern maize is a product of a single domestication event from Balsas teosinte (*Zea mays* ssp. *parviglumis*) that occurred about 8700 years ago in the Balsas River valley in southern Mexico (*112, 113*). Maize has maintained a substantial proportion (60–70%) of the genetic variation found in *Z. mays* ssp. *parviglumis*, and is more diverse than its more distantly related wild relative *Z. luxurians* (*100, 114-116*). However, during domestication and subsequent breeding, a small fraction of maize genes ("domestication genes") have been subject to very strong selective sweeps and, as a

result, have lost most if not all of their pre-domestication diversity (*115-118*)]. Maize eventually spread to other regions out of Mexico, owing to diffusion of seeds along trade networks. The earliest use of Maize is documented to be about 3200 years ago in Southwestern United States and about 2100 years ago in Eastern United States.

REFERENCES

1. G. V. Borner, N. Kleckner, N. Hunter, Crossover/noncrossover differentiation, synaptonemal complex formation, and regulatory surveillance at the leptotene/zygotene transition of meiosis. *Cell* 117, 29 (Apr 2, 2004).
2. S. Keeney, C. N. Giroux, N. Kleckner, Meiosis-specific DNA double-strand breaks are catalyzed by Spo11, a member of a widely conserved protein family. *Cell* 88, 375 (Feb 7, 1997).
3. O. Niwa, M. Shimanuki, F. Miki, Telomere-led bouquet formation facilitates homologous chromosome pairing and restricts ectopic interaction in fission yeast meiosis. *The EMBO journal* 19, 3831 (Jul 17, 2000).
4. H. W. Bass, W. F. Marshall, J. W. Sedat, D. A. Agard, W. Z. Cande, Telomeres cluster de novo before the initiation of synapsis: a three-dimensional spatial analysis of telomere positions before and during meiotic prophase. *The Journal of cell biology* 137, 5 (Apr 7, 1997).
5. H. Scherthan, J. Bahler, J. Kohli, Dynamics of chromosome organization and pairing during meiotic prophase in fission yeast. *The Journal of cell biology* 127, 273 (Oct, 1994).
6. L. Harper, I. Golubovskaya, W. Z. Cande, A bouquet of chromosomes. *Journal of cell science* 117, 4025 (Aug 15, 2004).
7. I. N. Golubovskaya, L. C. Harper, W. P. Pawlowski, D. Schichnes, W. Z. Cande, The pam1 gene is required for meiotic bouquet formation and efficient

homologous synapsis in maize (Zea mays L.). *Genetics* 162, 1979 (Dec, 2002).

8. P. R. Chua, G. S. Roeder, Tam1, a telomere-associated meiotic protein, functions in chromosome synapsis and crossover interference. *Genes & development* 11, 1786 (Jul 15, 1997).

9. M. N. Conrad, A. M. Dominguez, M. E. Dresser, Ndj1p, a meiotic telomere protein required for normal chromosome synapsis and segregation in yeast. *Science* 276, 1252 (May 23, 1997).

10. D. Q. Ding, A. Yamamoto, T. Haraguchi, Y. Hiraoka, Dynamics of homologous chromosome pairing during meiotic prophase in fission yeast. *Developmental cell* 6, 329 (Mar, 2004).

11. E. R. Nimmo, A. L. Pidoux, P. E. Perry, R. C. Allshire, Defective meiosis in telomere-silencing mutants of Schizosaccharomyces pombe. *Nature* 392, 825 (Apr 23, 1998).

12. E. Trelles-Sticken, M. E. Dresser, H. Scherthan, Meiotic telomere protein Ndj1p is required for meiosis-specific telomere distribution, bouquet formation and efficient homologue pairing. *The Journal of cell biology* 151, 95 (Oct 2, 2000).

13. Malik S-B, Pightling AW, Stefaniak LM, Schurko AM, L. J. Jr, An Expanded Inventory of Conserved Meiotic Genes Provides Evidence for Sex in Trichomonas vaginalis. *PLoS ONE* 3(8), (2008).

14. J. D. Higgins, E. Sanchez-Moran, S. J. Armstrong, G. H. Jones, F. C. Franklin, The Arabidopsis synaptonemal complex protein ZYP1 is required for

chromosome synapsis and normal fidelity of crossing over. *Genes & development* 19, 2488 (Oct 15, 2005).

15. M. Sym, J. A. Engebrecht, G. S. Roeder, ZIP1 is a synaptonemal complex protein required for meiotic chromosome synapsis. *Cell* 72, 365 (Feb 12, 1993).

16. H. Dong, G. S. Roeder, Organization of the yeast Zip1 protein within the central region of the synaptonemal complex. *The Journal of cell biology* 148, 417 (Feb 7, 2000).

17. G. V. Borner, A. Barot, N. Kleckner, Yeast Pch2 promotes domainal axis organization, timely recombination progression, and arrest of defective recombinosomes during meiosis. *Proceedings of the National Academy of Sciences of the United States of America* 105, 3327 (Mar 4, 2008).

18. L. K. Anderson, K. D. Hooker, S. M. Stack, The distribution of early recombination nodules on zygotene bivalents from plants. *Genetics* 159, 1259 (Nov, 2001).

19. S. M. Stack, L. K. Anderson, Crossing over as assessed by late recombination nodules is related to the pattern of synapsis and the distribution of early recombination nodules in maize. *Chromosome research : an international journal on the molecular, supramolecular and evolutionary aspects of chromosome biology* 10, 329 (2002).

20. L. K. Anderson *et al.*, High-resolution crossover maps for each bivalent of Zea mays using recombination nodules. *Genetics* 165, 849 (Oct, 2003).

21. F. Baudat, K. Manova, J. P. Yuen, M. Jasin, S. Keeney, Chromosome synapsis defects and sexually dimorphic meiotic progression in mice lacking Spo11. *Molecular cell* 6, 989 (Nov, 2000).

22. G. S. Roeder, Meiotic chromosomes: it takes two to tango. *Genes & development* 11, 2600 (Oct 15, 1997).

23. K. S. McKim *et al.*, Meiotic synapsis in the absence of recombination. *Science* 279, 876 (Feb 6, 1998).

24. A. F. Dernburg *et al.*, Meiotic recombination in C. elegans initiates by a conserved mechanism and is dispensable for homologous chromosome synapsis. *Cell* 94, 387 (Aug 7, 1998).

25. M. A. Bellani, P. J. Romanienko, D. A. Cairatti, R. D. Camerini-Otero, SPO11 is required for sex-body formation, and Spo11 heterozygosity rescues the prophase arrest of Atm-/- spermatocytes. *Journal of cell science* 118, 3233 (Aug 1, 2005).

26. F. Hartung *et al.*, The catalytically active tyrosine residues of both SPO11-1 and SPO11-2 are required for meiotic double-strand break induction in Arabidopsis. *The Plant cell* 19, 3090 (Oct, 2007).

27. S. P. Jackson, Sensing and repairing DNA double-strand breaks. *Carcinogenesis* 23, 687 (May, 2002).

28. K. Valerie, L. F. Povirk, Regulation and mechanisms of mammalian double-strand break repair. *Oncogene* 22, 5792 (Sep 1, 2003).

29. M. F. Lavin, The Mre11 complex and ATM: a two-way functional interaction in recognising and signaling DNA double strand breaks. *DNA repair* 3, 1515 (Nov 2, 2004).

30. A. Sancar, L. A. Lindsey-Boltz, K. Unsal-Kacmaz, S. Linn, Molecular mechanisms of mammalian DNA repair and the DNA damage checkpoints. *Annual review of biochemistry* 73, 39 (2004).

31. H. S. Lim, J. S. Kim, Y. B. Park, G. H. Gwon, Y. Cho, Crystal structure of the Mre11-Rad50-ATPgammaS complex: understanding the interplay between Mre11 and Rad50. *Genes & development* 25, 1091 (May 15, 2011).

32. K. Lammens *et al.*, The Mre11:Rad50 structure shows an ATP-dependent molecular clamp in DNA double-strand break repair. *Cell* 145, 54 (Apr 1, 2011).

33. T. T. Paull, M. Gellert, Nbs1 potentiates ATP-driven DNA unwinding and endonuclease cleavage by the Mre11/Rad50 complex. *Genes & development* 13, 1276 (May 15, 1999).

34. A. I. Roca, M. M. Cox, RecA protein: structure, function, and role in recombinational DNA repair. *Progress in nucleic acid research and molecular biology* 56, 129 (1997).

35. A. Shinohara, M. Shinohara, Roles of RecA homologues Rad51 and Dmc1 during meiotic recombination. *Cytogenetic and genome research* 107, 201 (2004).

36. K. Mitchel, H. Zhang, C. Welz-Voegele, S. Jinks-Robertson, Molecular structures of crossover and noncrossover intermediates during gap repair in yeast: implications for recombination. *Molecular cell* 38, 211 (Apr 23, 2010).

37. H. J. Muller, The mechanism of crossing-over. *Am. Nat.* 50, 193 (1916).

38. G. H. Jones, The control of chiasma distribution. *Symposia of the Society for Experimental Biology* 38, 293 (1984).

39. N. M. Hollingsworth, S. J. Brill, The Mus81 solution to resolution: generating meiotic crossovers without Holliday junctions. *Genes & development* 18, 117 (Jan 15, 2004).

40. F. W. Stahl *et al.*, Does crossover interference count in Saccharomyces cerevisiae? *Genetics* 168, 35 (Sep, 2004).

41. H. Guillon, F. Baudat, C. Grey, R. M. Liskay, B. de Massy, Crossover and noncrossover pathways in mouse meiosis. *Molecular cell* 20, 563 (Nov 23, 2005).

42. F. G. Lhuissier, H. H. Offenberg, P. E. Wittich, N. O. Vischer, C. Heyting, The mismatch repair protein MLH1 marks a subset of strongly interfering crossovers in tomato. *The Plant cell* 19, 862 (Mar, 2007).

43. J. D. Higgins, S. J. Armstrong, F. C. Franklin, G. H. Jones, The Arabidopsis MutS homolog AtMSH4 functions at an early step in recombination: evidence for two classes of recombination in Arabidopsis. *Genes & development* 18, 2557 (Oct 15, 2004).

44. R. Mercier *et al.*, Two meiotic crossover classes cohabit in Arabidopsis: one is dependent on MER3,whereas the other one is not. *Current biology : CB* 15, 692 (Apr 26, 2005).

45. M. Falque, L. K. Anderson, S. M. Stack, F. Gauthier, O. C. Martin, Two types of meiotic crossovers coexist in maize. *The Plant cell* 21, 3915 (Dec, 2009).

46. A. Lynn, R. Soucek, G. V. Borner, ZMM proteins during meiosis: crossover artists at work. *Chromosome research : an international journal on the molecular, supramolecular and evolutionary aspects of chromosome biology* 15, 591 (2007).

47. T. de los Santos *et al.*, The Mus81/Mms4 endonuclease acts independently of double-Holliday junction resolution to promote a distinct subset of crossovers during meiosis in budding yeast. *Genetics* 164, 81 (May, 2003).

48. M. C. Whitby, Making crossovers during meiosis. *Biochemical Society transactions* 33, 1451 (Dec, 2005).

49. T. Cavalier-Smith, Origins of the machinery of recombination and sex. *Heredity* 88, 125 (Feb, 2002).

50. A. S. Wilkins, R. Holliday, The evolution of meiosis from mitosis. *Genetics* 181, 3 (Jan, 2009).

51. M. A. Ramesh, S. B. Malik, J. M. Logsdon, Jr., A phylogenomic inventory of meiotic genes; evidence for sex in Giardia and an early eukaryotic origin of meiosis. *Current biology : CB* 15, 185 (Jan 26, 2005).

52. K. N. Smith, A. Nicolas, Recombination at work for meiosis. *Current opinion in genetics & development* 8, 200 (Apr, 1998).

53. D. Zickler, N. Kleckner, Meiotic chromosomes: integrating structure and function. *Annual review of genetics* 33, 603 (1999).

54. S. Keeney, Mechanism and control of meiotic recombination initiation. *Current topics in developmental biology* 52, 1 (2001).

55. M. Lichten, A. S. Goldman, Meiotic recombination hotspots. *Annual review of genetics* 29, 423 (1995).

56. J. L. Gerton *et al.*, Global mapping of meiotic recombination hotspots and coldspots in the yeast Saccharomyces cerevisiae. *Proceedings of the National Academy of Sciences of the United States of America* 97, 11383 (Oct 10, 2000).

57. I. J. Tsai, A. Burt, V. Koufopanou, Conservation of recombination hotspots in yeast. *Proceedings of the National Academy of Sciences of the United States of America* 107, 7847 (Apr 27, 2010).

58. F. Baudat *et al.*, PRDM9 is a major determinant of meiotic recombination hotspots in humans and mice. *Science* 327, 836 (Feb 12, 2010).

59. S. Myers *et al.*, Drive against hotspot motifs in primates implicates the PRDM9 gene in meiotic recombination. *Science* 327, 876 (Feb 12, 2010).

60. K. Paigen *et al.*, The recombinational anatomy of a mouse chromosome. *PLoS genetics* 4, e1000119 (Jul, 2008).

61. F. Baudat, B. de Massy, Cis- and trans-acting elements regulate the mouse Psmb9 meiotic recombination hotspot. *PLoS genetics* 3, e100 (Jun, 2007).

62. S. Myers, L. Bottolo, C. Freeman, G. McVean, P. Donnelly, A fine-scale map of recombination rates and hotspots across the human genome. *Science* 310, 321 (Oct 14, 2005).

63. C. Mezard, Meiotic recombination hotspots in plants. *Biochemical Society transactions* 34, 531 (Aug, 2006).

64. H. Yao *et al.*, Molecular characterization of meiotic recombination across the 140-kb multigenic a1-sh2 interval of maize. *Proceedings of the National Academy of Sciences of the United States of America* 99, 6157 (Apr 30, 2002).

65. J. Drouaud *et al.*, Variation in crossing-over rates across chromosome 4 of Arabidopsis thaliana reveals the presence of meiotic recombination "hot spots". *Genome research* 16, 106 (Jan, 2006).

66. Y. Blat, R. U. Protacio, N. Hunter, N. Kleckner, Physical and functional interactions among basic chromosome organizational features govern early steps of meiotic chiasma formation. *Cell* 111, 791 (Dec 13, 2002).

67. T. D. Petes, J. D. Merker, Context dependence of meiotic recombination hotspots in yeast: the relationship between recombination activity of a reporter construct and base composition. *Genetics* 162, 2049 (Dec, 2002).

68. J. Meunier, L. Duret, Recombination drives the evolution of GC-content in the human genome. *Molecular biology and evolution* 21, 984 (Jun, 2004).

69. S. Myers, C. Freeman, A. Auton, P. Donnelly, G. McVean, A common sequence motif associated with recombination hot spots and genome instability in humans. *Nature genetics* 40, 1124 (Sep, 2008).

70. K. Hayashi, K. Yoshida, Y. Matsui, A histone H3 methyltransferase controls epigenetic events required for meiotic prophase. *Nature* 438, 374 (Nov 17, 2005).

71. H. Donis-Keller *et al.*, A genetic linkage map of the human genome. *Cell* 51, 319 (Oct 23, 1987).

72. K. W. Broman, J. C. Murray, V. C. Sheffield, R. L. White, J. L. Weber, Comprehensive human genetic maps: individual and sex-specific variation in recombination. *American journal of human genetics* 63, 861 (Sep, 1998).

73. S. Shifman *et al.*, A high-resolution single nucleotide polymorphism genetic map of the mouse genome. *PLoS biology* 4, e395 (Nov, 2006).

74. J. Drouaud *et al.*, Sex-specific crossover distributions and variations in interference level along Arabidopsis thaliana chromosome 4. *PLoS genetics* 3, e106 (Jun, 2007).

75. K. E. Koehler, J. P. Cherry, A. Lynn, P. A. Hunt, T. J. Hassold, Genetic control of mammalian meiotic recombination. I. Variation in exchange frequencies among males from inbred mouse strains. *Genetics* 162, 297 (Sep, 2002).

76. E. Sanchez-Moran, S. J. Armstrong, J. L. Santos, F. C. Franklin, G. H. Jones, Variation in chiasma frequency among eight accessions of Arabidopsis thaliana. *Genetics* 162, 1415 (Nov, 2002).

77. S. E. Ptak *et al.*, Fine-scale recombination patterns differ between chimpanzees and humans. *Nature genetics* 37, 429 (Apr, 2005).

78. W. Winckler *et al.*, Comparison of fine-scale recombination rates in humans and chimpanzees. *Science* 308, 107 (Apr 1, 2005).

79. J. A. Anderson, W. D. Gilliland, C. H. Langley, Molecular population genetics and evolution of Drosophila meiosis genes. *Genetics* 181, 177 (Jan, 2009).

80. H. Rees, B chromosomes. *Science progress* 61, 535 (Winter, 1974).

81. A. Burt, G. Bell, Mammalian chiasma frequencies as a test of two theories of recombination. *Nature* 326, 803 (Apr 23-29, 1987).

82. S. P. Otto, Y. Michalakis, The evolution of recombination in changing environments. *Trends Ecol. Evol.* 13, 145 (Apr, 1998).

83. M. Saleem, B. C. Lamb, E. Nevo, Inherited differences in crossing over and gene conversion frequencies between wild strains of Sordaria fimicola from "Evolution Canyon". *Genetics* 159, 1573 (Dec, 2001).

84. Gornall, Recombination systems and plant domestication. *Biological Journal of the Linnean Society* 20, 375 (1983).

85. J. Ross-Ibarra, The evolution of recombination under domestication: a test of two hypotheses. *The American naturalist* 163, 105 (Jan, 2004).

86. R. Nielsen, Molecular signatures of natural selection. *Annu. Rev. Genet.* 39, 197 (2005).

87. F. Tajima, Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. *Genetics* 123, 585 (Nov, 1989).

88. Y. X. Fu, W. H. Li, Statistical tests of neutrality of mutations. *Genetics* 133, 693 (Mar, 1993).

89. J. H. McDonald, M. Kreitman, Adaptive protein evolution at the Adh locus in Drosophila. *Nature* 351, 652 (Jun 20, 1991).

90. R. R. Hudson, M. Kreitman, M. Aguade, A test of neutral molecular evolution based on nucleotide data. *Genetics* 116, 153 (May, 1987).

91. R. Nielsen, Z. Yang, Likelihood models for detecting positively selected amino acid sites and applications to the HIV-1 envelope gene. *Genetics* 148, 929 (Mar, 1998).

92. S. V. Muse, B. S. Gaut, A likelihood approach for comparing synonymous and nonsynonymous nucleotide substitution rates, with application to the chloroplast genome. *Mol. Biol. Evol.* 11, 715 (Sep, 1994).

93. A. L. Hughes, M. Nei, Pattern of nucleotide substitution at major histocompatibility complex class I loci reveals overdominant selection. *Nature* 335, 167 (Sep 8, 1988).

94. D. Vermaak, S. Henikoff, H. S. Malik, Positive selection drives the evolution of *rhino*, a member of the heterochromatin protein 1 family in *Drosophila*. *PLoS Genet.* 1, 96 (Jul, 2005).

95. W. Zhai, R. Nielsen, M. Slatkin, An investigation of the statistical power of neutrality tests based on comparative and population genetic data. *Mol. Biol. Evol.* 26, 273 (Feb, 2009).

96. S. Biswas, J. M. Akey, Genomic insights into positive selection. *Trends Genet.* 22, 437 (Aug, 2006).

97. H. Innan, Y. Kim, Pattern of polymorphism after strong artificial selection in a domestication event. *Proc. Natl. Acad. Sci. USA* 101, 10667 (Jul 20, 2004).

98. E. S. Buckler, B. S. Gaut, M. D. McMullen, Molecular and functional diversity of maize. *Current opinion in plant biology* 9, 172 (Apr, 2006).

99. M. I. Tenaillon *et al.*, Patterns of DNA sequence polymorphism along chromosome 1 of maize (Zea mays ssp. mays L.). *Proceedings of the National Academy of Sciences of the United States of America* 98, 9161 (Jul 31, 2001).

100. M. A. Gore *et al.*, A first-generation haplotype map of maize. *Science* 326, 1115 (Nov 20, 2009).

101. N. M. Springer *et al.*, Maize inbreds exhibit high levels of copy number variation (CNV) and presence/absence variation (PAV) in genome content. *PLoS genetics* 5, e1000734 (Nov, 2009).

102. B. S. Gaut, J. F. Doebley, DNA sequence evidence for the segmental allotetraploid origin of maize. *Proceedings of the National Academy of Sciences of the United States of America* 94, 6809 (Jun 24, 1997).

103. Z. Swigonova *et al.*, On the tetraploid origin of the maize genome. *Comparative and functional genomics* 5, 281 (2004).

104. B. S. Gaut, M. Le Thierry d'Ennequin, A. S. Peek, M. C. Sawkins, Maize as a model for the evolution of plant nuclear genomes. *Proceedings of the National Academy of Sciences of the United States of America* 97, 7008 (Jun 20, 2000).

105. B. S. Gaut, Patterns of chromosomal duplication in maize and their implications for comparative maps of the grasses. *Genome research* 11, 55 (Jan, 2001).

106. K. Ilic, P. J. SanMiguel, J. L. Bennetzen, A complex history of rearrangement in an orthologous region of the maize, sorghum, and rice genomes.

Proceedings of the National Academy of Sciences of the United States of America 100, 12265 (Oct 14, 2003).

107. J. Lai *et al.*, Gene loss and movement in the maize genome. *Genome research* 14, 1924 (Oct, 2004).

108. R. J. Langham *et al.*, Genomic duplication, fractionation and the origin of regulatory novelty. *Genetics* 166, 935 (Feb, 2004).

109. P. S. Schnable *et al.*, The B73 maize genome: complexity, diversity, and dynamics. *Science* 326, 1112 (Nov 20, 2009).

110. P. SanMiguel, B. S. Gaut, A. Tikhonov, Y. Nakajima, J. L. Bennetzen, The paleontology of intergene retrotransposons of maize. *Nature genetics* 20, 43 (Sep, 1998).

111. A. H. Paterson *et al.*, The Sorghum bicolor genome and the diversification of grasses. *Nature* 457, 551 (Jan 29, 2009).

112. D. R. Piperno, A. J. Ranere, I. Holst, J. Iriarte, R. Dickau, Starch grain and phytolith evidence for early ninth millennium B.P. maize from the Central Balsas River Valley, Mexico. *Proc. Natl. Acad. Sci. USA* 106, 5019 (Mar 31, 2009).

113. Y. Matsuoka *et al.*, A single domestication for maize shown by multilocus microsatellite genotyping. *Proc. Natl. Acad. Sci. USA* 99, 6080 (Apr 30, 2002).

114. A. Eyre-Walker, R. L. Gaut, H. Hilton, D. L. Feldman, B. S. Gaut, Investigation of the bottleneck leading to the domestication of maize. *Proc. Nat. Acad. Sci. USA* 95, 4441 (Apr 14, 1998).

115. M. I. Tenaillon, J. U'Ren, O. Tenaillon, B. S. Gaut, Selection versus demography: A multilocus investigation of the domestication process in maize. *Mol. Biol. Evol.* 21, 1214 (Jul, 2004).

116. F. Tian, N. M. Stevens, E. S. t. Buckler, Tracking footprints of maize domestication and evidence for a massive selective sweep on chromosome 10. *Proc. Natl. Acad. Sci. USA* 106, 9979 (Jun 16, 2009).

117. S. I. Wright *et al.*, The effects of artificial selection on the maize genome. *Science* 308, 1310 (May 27, 2005).

118. M. Yamasaki *et al.*, A large-scale screen for artificial selection in maize identifies candidate agronomic loci for domestication and crop improvement. *Plant Cell* 17, 2859 (Nov, 2005).

CHAPTER 2

EVOLUTION OF MEIOTIC RECOMBINATION GENES IN MAIZE AND TEOSINTE

2.1 Introduction

Meiotic recombination produces genetic variation by creating new combinations of alleles, and facilitates purging of deleterious mutations from genomes and populations (Gaut et al. 2007). While the role of recombination in evolution is well recognized, the evolution of the recombination pathway itself has received little attention. The goal of this study was to investigate the evolution of the meiotic recombination pathway during the domestication and diversification of maize. We examined the patterns of polymorphism and divergence in maize and teosinte genes encoding proteins that control key events in meiotic recombination: SPO11, MRE11, DMC1, SGS1, RAD51A, MSH4, MLH1, and MUS81.

The major steps of meiotic recombination are highly conserved among eukaryotes, although some variation exists (Villeneuve and Hillers 2001). Recombination is initiated by formation of double-strand breaks (DSBs) in chromosomal DNA by Spo11, a protein belonging to the topoisomerase family (Keeney, Giroux, and Kleckner 1997; Grelon et al. 2001; Stacey et al. 2006). The DSBs are subsequently resected from 5' to 3' by the MRN protein complex in plants and animals to generate single-stranded DNA (ssDNA) overhangs (Borde 2007; Waterworth et al. 2007). MRE11 is a key component of this complex, as it possesses endonuclease, exonuclease, and helicase activities that directly facilitate the ssDNA

overhang formation (Borde 2007; Buis et al. 2008). The ssDNA ends are then coated by two DNA strand-exchange proteins, Rad51A and Dmc1, and invade homologous double-stranded DNA regions (Neale and Keeney 2006). Eventually, meiotic DSBs are repaired into either crossovers (COs) or non-crossovers (NCOs; gene conversions). In budding yeast, a RecQ helicase Sgs1 has been suggested to prevent a subset of recombination intermediates from becoming COs (Louis and Borts 2003; Rockmill et al. 2003). CO formation is regulated by a phenomenon known as CO interference, which prevents formation of multiple crossovers in the same chromosome region (Copenhaver 2005; Jones and Franklin 2006). In plants, as well as yeast and mammals, there are two main classes of COs: type I COs that are subject to interference, and type II COs that are not (Higgins et al. 2004; Chen et al. 2005; Mercier et al. 2005; Jones and Franklin 2006; Berchowitz et al. 2007). The two CO classes are outcomes of parallel pathways, which are facilitated by different complexes of recombination proteins. MSH4 and MLH1 act in formation of the type I COs (Higgins et al. 2004; Dion et al. 2007) while MUS81 is involved in type II CO formation (Berchowitz et al. 2007).

The overall conservation of the meiotic recombination pathway suggests that its components are subject to predominantly purifying selection. This expectation was partially confirmed by a recent study of a set of meiotic genes in *Drosophila melanogaster* and *D. simulans* (Anderson, Gilliland, and Langley 2009). However, in addition to evidence of purifying selection, the *Drosophila* study also revealed patterns of polymorphisms indicative of positive selection in a small fraction of the genes.

Theoretical predictions (Otto and Michalakis 1998) and empirical studies (Saleem, Lamb, and Nevo 2001) indicate that populations experiencing directional or strong selection pressures are likely to evolve increased recombination rates. Increase in meiotic recombination rates also has been suggested to accompany domestication (Ross-Ibarra 2004), although its has also been argued that high recombination rates are a pre-adaptation to domestication rather than an effect of domestication (Gornall 1983). Based on these predictions, we hypothesized that recombination rate increases are likely to be associated with adaptive evolution patterns in genes involved in the meiotic recombination pathway.

Various methods have been devised to detect patterns of selection acting on a locus based on DNA sequence (Nielsen 2005). One group of methods, which includes the often-used Tajima's D (Tajima 1989) and Fu and Li's D and F tests (Fu and Li 1993), is based on assessing the frequency spectrum of allelic polymorphisms in the population. Another group, which includes the McDonald-Kreitman (McDonald and Kreitman 1991) and Hudson-Kreitman-Aguade (Hudson, Kreitman, and Aguade 1987) tests, compares the patterns of sequence polymorphisms within the species with the patterns of sequence divergence between species. Finally, a number of methods that are used specifically to detect selection within the gene coding region, assess the ratio of non-synonymous to synonymous nucleotide substitution rates (Hughes and Nei 1988; Muse and Gaut 1994; Nielsen and Yang 1998). These methods are based on the premise that under neutral evolution non-synonymous and synonymous substitutions should occur at the same frequencies, while under selection these rates should differ from each other (Nielsen 2005). This diversity of methods is

a testament to the complexity of the processes that contribute to adaptive evolution. The different methods for identifying selection patterns often produce conflicting results from the same data sets; in some studies the overlap between different methods have been shown to be minimal or none (Vermaak, Henikoff, and Malik 2005; Biswas and Akey 2006; Zhai, Nielsen, and Slatkin 2009). These discrepancies are likely caused by the fact that different methods are more sensitive to different types of selection events. The ability of to detect selection patterns in DNA sequence data depends on the strength, timing and frequency of the selection sweeps, constraints to nucleotide substitutions within the locus, the type of selection, the allele frequency before the selection sweep, and finally, on whether the locus is itself the target of selection or is "hitchhiked" because of selection in neighboring genes (Innan and Kim 2004; Nielsen 2005; Biswas and Akey 2006; Zhai, Nielsen, and Slatkin 2009)

Modern maize is a product of a single domestication event from Balsas teosinte (*Zea mays* ssp. *parviglumis*) that occurred about 8700 years ago in the Balsas River valley in southern Mexico (Matsuoka et al. 2002; Piperno et al. 2009). Maize has maintained a substantial proportion (60–70%) of the genetic variation found in *Z. mays* ssp. *parviglumis*, and is more diverse than its more distantly related wild relative *Z. luxurians* (Eyre-Walker et al. 1998; Tenaillon et al. 2004; Gore et al. 2009; Tian, Stevens, and Buckler 2009). However, during domestication and subsequent breeding, a small fraction of maize genes ("domestication genes") have been subject to very strong selective sweeps and, as a result, have lost most if not all of their pre-domestication diversity (Tenaillon et al. 2004; Wright et al. 2005; Yamasaki et al. 2005; Tian, Stevens, and Buckler 2009).

In our analyses of recombination genes, we found that although these genes were not subject to as strong selective sweeps as the domestication genes, nearly half of the recombination genes examined in maize showed signatures of adaptive evolution. Interestingly, in *Z. mays* ssp. *parviglumis* different and fewer genes than in maize exhibited patterns consistent with positive selection. We identified several sequence polymorphisms that have the potential to considerably affect the protein function. These data suggest that changes in recombination genes may have contributed to the successful domestication of maize and its expansion to new cultivation areas.

2.2 Materials and methods

2.2.1 Plant material

Sequence diversity in recombination pathway genes was examined in a set of 31 maize inbreds and 14 teosinte accessions. The maize inbreds were selected to maximize the genetic diversity of maize (Liu et al. 2003) and included 25 of the 26 founders of the Nested Association Mapping (NAM) population (McMullen et al. 2009): B73, B97, CML103, CML228, CML247, CML277, CML322, CML333, CML52, CML69, Hp301, Il14H, Ki11, Ki3, Ky21, M162W, M37W, Mo18W, MS71, NC350, NC358, Oh43, Oh7B, P39, and Tx303. We supplemented this set with six other inbred lines: A188, A344, CO106, CO125, CO255, and Mo17. The teosinte lines included nine *Z. mays* ssp. *parviglumis* lines, eight of which came from a set developed by John Doebley (University of Wisconsin, Madison) (TI01, TI02, TI05, TI07, TI11, TI15, TI16, and TI17), and one was from the CIMMYT collection

(TL74A J2 K67-5). The other teosinte lines used in this study were *Z. mays* ssp. *mexicana* (K69-7 and BA93 WST 85-2), *Z. mays* ssp. *huehuetenangensis* (TL93B Teo Huehue), *Z. diploperennis* (JAL87 Las Joyas), and *Z. luxurians* (TL92B TEO-Guate). Seeds and/or tissue samples for all lines except A188, A344, and Mo17 were kindly provided by Ed Buckler (USDA-ARS and Cornell University, Ithaca, NY).

2.2.2 Gene sequences

Genomic regions of eleven recombination genes (*Dmc1*, *Mlh1*, *Mre11A*, *Mre11B*, *Msh4*, *Mus81-1*, *Rad51A1*, *Rad51A2*, *Sgs1*, *Spo11-1*, and *Spo11-2*) were identified in the whole-genome sequence of the maize B73 inbred (Schnable et al. 2009) (last accessed on May 28, 2009) using sequences of known Arabidopsis and maize recombination genes as BLAST queries. Gene coding regions were delineated using full-length cDNA and EST sequences available in the GenBank (last accessed on March 23, 2009). For *Mlh1*, *Msh4*, *Mus81*, and *Sgs1*, only partial EST sequences were present in the GenBank. We determined the full coding regions of these genes using RT-PCR, which was performed as previously described (Pawlowski et al. 2004).

To obtain sequences of the eleven recombination genes from the set of maize inbreds and teosinte lines, PCR primers (Table 2.1) were designed to amplify full-length genomic regions, excluding large introns, in four to eleven fragments for each gene. Nearly all primers were designed to anneal in introns to obtain entire coding regions and ensure that only orthologus sequences are amplified. When selecting

primer sites, we avoided regions present in multiple copies in the maize genome sequence.

Sequencing reactions were performed directly on PCR products in both orientations with BigDye v3.1 (Applied Biosystems, Foster City, CA), and analyzed using the Applied Biosystems 3730 automated sequence analyzer. Manual sequence editing was conducted using Sequencher (Gene Codes Corp., Ann Arbor, MI).

Table 2.1: Sequences of PCR primers used to amplify meiotic recombination gene homologs from maize and teosintes.

Gene	Primer pair	Forward primer	Reverse primer
Dmc1	1	CTCAGCACCCTCAAGTACACAG	CACTAATGCGGGAGGAGAGA
	2	TAATCCACCTTTGCAGTTTTGA	AGTACCCTTGTCCTTCTGCGT
	3	AACACTTTGCATCACAGAGGC	GTCGGCTGGAAACTTCATTAA
	4	GGTTGCCTACATTGACACTGA	CCACTGAAATCAACACGGAAC
	5	GTTCCGTGTTGATTTCAGTGG	CCGAATCAGTCTTTTACGTCC
Mlh1	1	ACGGAGGGAGTAGTTGTTTTTATAT	CTCGACGCTAGAAAAGTGGG
	2	ATATTTTCGTGGTTTCACCTTAA	TATTCGTTGCCTAAAAAAATCACT
	3	CCGCTCCTTTTTGTTTTCTACT	ACCACGTAAGATCAATGTTCATAAA
	4	CTTTGATGTGAACTTCTGTTTCTG	AATAACCTCTAAAGACCTGCAGC
	5	TGCTGCAGGTCTTTAGAGGTTA	TATTTTCGAGTTTGCCTGCA
	6	TGCAGAGGTAAGTTATTTTTGGAA	AGGAGTGGTTTCAGGGTTGTC
	7	ATTGCTATTGGTGCTAAAAGGTT	TATAGGGTGTCTTTCTTGGCTTC
Mre11A	1	CTTCTCTTCCCTTGCTGCTG	TTTACACATTTTCACTAGTCCACAA
	2	TGCAGCAATGTTGTTCACT	CAAAATGAAGGTCCAAA

		TTT	AACAA
	3	TTGTTTTTGGACCTTCATT TTG	TTTTTCCTTTCTTGATGA TTTTTC
	4	GAAAAATCATCAAGAAAG GAAAAA	CAGGGAGCAAATGGACA CAG
	5	GGGTTTGCACTCCTATACT TCTTT	CAGCCCCAGGAAAAATA CAA
	6	TTGTATTTTTCCTGGGGCT G	TCGGCTCTTCCAACTTTA CAA
Mre11B	1	TGAGGCTCGTTAATTTGTG TTC	GCCAGTCGCAAAATCAA AAC
	2	CACATTTTCATTCCTTTTG GG	TTAACAGAGAAACGTGA AGGAAAC
	3	AGCATTTTTTTGTACGTTC ACTTAT	ACCTCTTGCCCTCTTTTT AAATA
	4	TGTTCTGTCTTTCCGCTTTT G	CCATATGAGCGTACCCT AGCA
	5	CATTCCATCTTTTTCTAGT ATTGCT	TGGTGGTAGTTAATTTGT GATGC
	6	CCGCTGAAGAAAAAGTAG GATAT	CAACTAATCGATCAGGG TCCTC
Msh4	1	GGTCATTCCATTCCGTCAC AT	TCCGACGAAATAACAGC TCAC
	2	TGATTTGGTCATCCTTTTT TGA	ACCGTAAAAGTTGAAAT CCTGAA
	3	GTCATATCACCTTTTTTCT GGAAT	GAAACGACACGCAGGTA TCC
	4	CTGGGAATTTGAGTTGTTG ATCT	CTACGAGTAAAGACAAT GGGACC
	5	CACAGGCCCAAGTCCATA AC	TTATGAAATTATGAAGG GAACCAA
	6	CTACGAGTTCCTGGTCATG CC	CACAGCGCTACGGATGA CTAA
	7	TTGCTTTTTTGTTGATGTC CT	AGTTTTCGTGACCCTTCT CG
	8	GCTTTGAGTCTGCCATTTT TTA	TTTAGTCATTTTCAGCTC TTCTCG
	9	TTGCCATGGTTAGATATCA GAATTA	AAATGATCAGCAGTTGG GTACA
	10	CTAAAACCCCCCATTCTAT TCA	AACATTTATTAAGGGCC AGAGC
Mus81-1	1	CCGCAAAATTCAAATCAG TTC	CAGACAGTTAAGCATGA AAATCG
	2	TTTTCCACACCAATTATCT TTTCT	GCAAGGTTCCTCATAAT TTTAAGA
	3	TTGATATTTTTGTGTTGTC TTCCAG	GAAAAAGGTATTACTGT CGAGGC
	4	TGATCTCTTAGTTGTGCCA TGTG	ATATGCCTCGAGATCAC CAAGT

Rad51A1	1	TTGCCTATTGGTGTTGGTTGT	CAAGAAGCCGGTTAGGGTTA
	2	AGCTCCTATGGTTGCATTTTG	CCAACAGAACAATCAAGGAATG
	3	ATCCACCCTCTTTTTCCATTT	CACCTTTGATTCGTAAGCAGAA
	4	TCTGGCTGCTGTTCTGAATCT	CCAAGGAATGACAAAGCAAAT
Rad51A2	1	CAAAGCCAAAGTGAGGATGC	AACGGGACCTAAGATATTGGC
	2	TTGGGGAATTTTCATGTTAGAGA	GTGGTGCCATATCTTAGAATTTGT
	3	ACAAATTCTAAGATATGGCACCAC	AAGTAACGTTTGCATGAGGACA
	4	CTGCTATGAAATTACCGGACAA	ACGAGTGCAGCAGATCAGAGA
Sgs1	1	CCAATCCCTTCTGTTCCCTT	AAGTCTCGACAAATGTGCTAAGC
	2	ATCCCTACCTTCTCTTTTACCTACA	TAGCCCCTCAGTATAGAACAAAGAT
	3	GGCTATTTCCGCATGTTATCT	AGCTGCTTAGAACCCATGACC
	4	TTTTCCTTTGCCAATCACTGA	TGTGGCGTGTCAGTAAAATAGAA
	5	CTGTTGAGACTGTTGGAAAATGTAT	GCTCCCAATCTATATCCAGGTAAT
	6	TGGCGTACAGATTTTGAACAGT	GTCCACGTCATTTTCACAGTAACT
	7	GCACTGAAGACACATAAGGACAAT	TGGAAACAAAAAATAAATAGAAACG
	8	CGTTTCTATTTATTTTTTGTTTCCA	AATACAATGACAGCCACACACAA
	9	CCCCCCCTCAAACTAACATAT	CTCATTTCCTTCAACGTGTCTG
	10	TTGTTGCACCTTGTCATTCG	AACCCTTGACATCACTGCATC
Spo11-1	1	TTGTTAATTTTTTGTCTTGAGAGGA	CCCTCAGGAGCACATCTACG
	2	CGTCCTCTCTCTGCTCCACA	AGCACACAGGATGATTAAGGAAC
	3	TTTGCTAGGTTTGCGATGTTATA	AAGAGCAAAAAATGATCAGTGACT
	4	GCATTATAGTGAGCTGAAGTTGTCT	GCATAGCAGCATCAGATAAATCA
	5	TGCTACTTGCGTAATTTCATTCT	ACTCTATGACACATCCTGGACG
Spo11-2	1	ACGCAGACCACTACTGGACTC	GAGGAAGTTGAGCACCGCTAC
	2	GTACACCAGTTGCCAAAC	ATAATTCCGTCCCAGGG

	CTC	TACT
3	TATCCTCGCCCATTTGTCTG	ATGCTTGAAATGGAGGTACACA
4	TTGCATCCAAATTCTTTATGATAA	TCGTCAGATGTAATCGCCCT

2.2.3 Phylogenetic analyses

To understand the phylogenetic history of maize and teosinte recombination genes within the context of eukaryortic evolution, alignments of recombination protein sequences from

several species of eukaryotes were performed using ClustalX (Larkin et al. 2007) and adjusted manually. Alignment gaps were excluded from analyses. Maximum parsimony analyses of protein sequences were conducted using PAUP 4.0 (Swofford 2003). Bayesian analyses were performed using MrBayes 3.1.2 (Huelsenbeck and Ronquist 2001). TreeviewPPC (Page 1996) was used to display phylogenetic trees. To compare evolution rates between different branches of phylogenetic trees, we used the Tajima's 1D relative rate test (Tajima 1993) implemented in MEGA4 (Tamura et al. 2007).

2.2.4 Gene genealogy analyses

To examine evolution of the recombination genes within the maize and teosinte lineages, alignments of maize and teosinte nucleotide sequences of the eleven recombination genes were done and manually adjusted in Sequencher. Gaps in sequence alignments were coded manually using the simple indel coding approach (Simmons and Ochoterena 2000). To construct gene genealogies, Bayesian analyses were performed using MrBayes 3.1.2 (Huelsenbeck and Ronquist 2001). Prior to

these analyses, the best-fit models for nucleotide substitutions (GTR+I for *Msh4*, GTR+I+Γ for *Mlh1* and *Mre11A*, HKY+I for *Mre11B*, *Mus81-1*, *Spo11-1*, *Rad51A1*, and *Sgs1*, HKY+I+Γ for *Dmc1*, *Rad51A2*, and *Spo11-1*) were determined from the Akaike Information Criterion (AIC) implemented in MrModeltest 2.3 (Nylander 2004). Mesquite (Maddison and Maddison 2009) was used to display gene genealogy trees.

In addition to the genealogy trees, we constructed a parsimony split network for each of the genes using SplitsTree (Huson and Bryant 2006). Phylogenetic networks can better than trees reflect phylogenetic uncertainties in the gene evolutionary history that are caused, for example, by hybridization and recombination (Huson and Bryant 2006).

2.2.5 Sequence divergence and diversity analyses

To assess the levels of sequence diversity in the recombination genes within maize, nucleotide diversity measures, π (Nei and Li 1979) and θ_W (Watterson 1975) were calculated using DNAsp v. 5 (Librado and Rozas 2009).

To examine the rates of divergence of recombination genes across eukaryotes, we used the K tree analysis (Soria-Carrasco et al. 2007), which compares the overall sizes of phylogenetic trees. The trees were constructed based on amino acid sequences for the same set of species for each of the genes. The K tree analysis was conducted using the Ktreedist_v1 program (Soria-Carrasco et al. 2007).

2.2.6 Selection analyses

We used several methods to examine the DNA sequences of the recombination genes in maize and teosintes for presence of selection signatures, including Tajima's D (Tajima 1989), Fu and Li's D (Fu and Li 1993), and Fu and Li's F (Fu and Li 1993), Hudson-Kreitman-Aguade (HKA) (Hudson, Kreitman, and Aguade 1987), McDonald-Kreitman (MK) (McDonald and Kreitman 1991), and the likelihood ratio (LRT) (Nielsen and Yang 1998) tests. Tajima's D and Fu and Li's D and F are frequency spectrum-based tests and were conducted using DNAsp v. 5 (Librado and Rozas 2009). HKA and MK compare within-species diversity to between-species divergence and are also implemented in DNAsp v. 5 (Librado and Rozas 2009). LRT examines the ratios of non-synonymous to synonymous nucleotide substitution rates and conducts pair-wise comparisons between several models that describe different selection patterns defined by the ratio of non-synonymous (dN) to synonymous (dS) substitution rates, including purifying selection ($dN/dS < 1$), neutral evolution ($dN/dS = 1$), positive selection ($dN/dS > 1$), to identify the best-fitting model for each gene. The analysis was performed using the codeml program in the PAML package (Yang 2007). The LRT method is not reliable when recombination is frequent among the examined haplotypes (Anisimova, Nielsen, and Yang 2003). Therefore, prior to the LRT analysis, we established that recombination frequencies in the coding region of each gene did not exceed the acceptable limits (Anisimova, Nielsen, and Yang 2003) using the Genetic Algorithm Recombination Detection (GARD) method (Kosakovsky Pond et al. 2006) conducted using a web interface

(http://www.datamonkey.org/GARD/). The GARD method identifies recombination breakpoints in sequence alignments by searching for phylogenetic incongruence.

Coalescent simulations were conducted to examine if specific values of the Tajima's D statistic were likely to be obtained by chance alone under neutral evolution. We used Hudson's ms program (Hudson 2002) to generate 10,000 coalescent simulations using previously described parameters (Wright et al. 2005). A conservative assumption of no intra-locus recombination was used in the simulations (Hudson 1990). The population mutation parameter θ was estimated from teosinte data. To simulate the domestication bottleneck, we used 2.45 as the value of the bottleneck severity (*k*), the ratio of population size during bottleneck to bottleneck duration.

2.2.7 Protein structure predictions

To determine the locations of polymorphic amino acid residues in three-dimensional protein structures, we conducted protein structure prediction analyses.

Maize protein sequences were threaded using Cn3D on the available empirical (www.ncbi.nlm.nih.gov/Structure/CN3D/cn3d.shtml) three-dimensional structures of the MutL transducer domain (MMDB ID: 10447) in MLH1, the phosphoesterase domain (MMDB ID: 34451) present in MRE11, the ERCC domain (MMDB ID: 52594) present in MSH4, the ERCC domain (MMDB ID: 52594) present in MUS81, the BRC repeat (MMDB ID: 21264) present in RAD51, the DEXDc domain (MMDB ID: 13107), the HELICc domain (MMDB ID: 12961), the REQC domain (MMDB

ID: 36694), and the HRDC domain (MMDB ID: 34680) in SGS1, and the TOPRIM domain (MMDB ID: 11634) in SPO11.

2.3 Results

2.3.1 Genomic organization of maize meiotic recombination genes

Of the eight recombination proteins that we set to investigate (DMC1, MLH1, MRE11, MSH4, MUS81, RAD51A, and SPO11), maize homologs of two, RAD51A and MRE11, have been previously identified and examined at the functional level (Franklin et al. 1999; Altun 2007; Li et al. 2007; Waterworth et al. 2007). These studies showed that the maize genome contains duplicated copies of *Mre11* (*Mre11A* and *Mre11B*) and *Rad51A* (*Rad51A1* and *Rad51A2*). We confirmed these findings (Table 2.2) using the complete maize genome sequence now available (Schnable et al. 2009).

We identified maize homologs of genes encoding DMC1, MLH1, MSH4, MUS81, SGS1, and SPO11 by searching the publicly available maize genomic and EST sequence resources using TBLASTN with Arabidopsis protein sequences as queries. All these proteins have been previously examined in Arabidopsis at the functional level (Couteau et al. 1999; Grelon et al. 2001; Higgins et al. 2004; Stacey et al. 2006; Berchowitz et al. 2007; Dion et al. 2007; Hartung, Suer, and Puchta 2007; Sanchez-Moran et al. 2008).

Table 2.2: Maize genes encoding key meiotic recombination proteins.

Gene	Coding sequence	ORF	Chromosome /bin	Number of pseudogenes containing truncated gene sequence
Dmc1	1035bp	4059bp	3/09	12
Mlh1	2178bp	6869bp	8/06	4
Mre11A	2121bp	4867bp	2/02	-
Mre11B	2019bp	5389bp	4/04	-
Msh4	2415bp	51074bp	2/06	-
Mus81-1	1353bp	3409bp	3/05	-
Rad51A1	1023bp	3589bp	7/04	-
Rad51A2	1023bp	2815bp	3/05	-
Sgs1	3528bp	12455bp	5/04	-
Spo11-1	1152bp	4316bp	5/01	-
Spo11-2	1149bp	>2513bp[a]	4/04	1

[a] Actual ORF length unknown because of a gap in intron 7.

In Arabidopsis, DMC1, MLH1, MSH4, and MUS81 are encoded by single genes. Our investigation of the maize genome sequence revealed that *Dmc1*, *Mlh1*, and *Msh4* are also present as single full-length genes in maize (Table 2.2). In contrast, we found two sequence homologs of *Mus81*, *Mus81-1* and *Mus81-2*, which shared a rather limited 46% identity and 62% similarity at the amino acid level. Arabidopsis has multiple genes encoding RecQ helicases but only one, RECQ4A, is a likely functional homolog of Sgs1 (Hartung, Suer, and Puchta 2007). A single homolog of *RECQ4A* is present in maize (Table 2.1). SPO11 in Arabidopsis is represented by three isoforms resulting from an ancient gene duplication in now extinct basal eukaryotes (Hartung and Puchta 2000; Malik et al. 2007). However, only two of the three *SPO11* genes, *SPO11-1* and *SPO11-2*, function in meiotic recombination (Grelon et al. 2001; Stacey

et al. 2006). The maize genome contains single homologs of all three Arabidopsis *SPO11* genes.

In addition to full-length gene copies, when searching the maize genome sequence, we discovered truncated copies of *Dmc1*, *Mlh1*, and *Spo11-2* (Table 2.2). These partial sequences were flanked by DNA that showed no similarities to the corresponding full-length genes. The partial copies represented different fragments of the full-length genes, always spanning several exons and introns. They showed nearly 100% sequence identity to the corresponding fragments of the full-length paralogs, suggesting that they are relatively recent pseudogenes.

2.3.2 Origin of the duplicated recombination genes in maize

The maize genome is most likely a product of an allopolyploidization event that occurred roughly 11.4 Mya between two ancestors that diverged from each other about 20.5 Mya (Gaut and Doebley 1997). The presence of two *Rad51A* copies in maize was interpreted to be a result of this duplication (Franklin et al. 1999). To examine the evolutionary ancestry of maize recombination genes, we conducted phylogenetic analyses of amino acid sequences from several representative species of eukaryotes using Bayesian as well as maximum parsimony (MP) methods (Figure 2.1a, 2.1b, 2.1c. 2.1d, 2.1e). Both methods produced essentially identical trees, except for SPO11, where the Bayesian tree better resolved the phylogenetic relationships than the MP tree. The phylogeny reconstructions revealed that the presence of duplicated copies of *Mre11*, *Rad51A*, and *Mus81* preceded the polyploidization event that gave rise to the tetraploid maize genome (Figure 2.1a, 2.1d, 2.1b). *Rad51A* and

Mre11 are duplicated in other grasses, including rice, and the duplication events pre-date the divergence of the maize and rice lineages. *Mus81* also has undergone duplication before the maize-rice divergence but in the rice lineage, the copy corresponding to maize *Mus81-2* was subsequently lost. According to Tajima's 1D relative rate test, maize *Mus81-2* shows a much accelerated evolution rate compared to *Mus81-1* ($P = 0.00465$; Arabidopsis was used the outgroup). These data create significant uncertainty whether *Mus81-2* retained the same function in meiosis as *Mus81-1*. The rice *Mus81-1* gene has been shown to play a role in recombination (Mimida et al. 2007). In contrast, no functional information exists for *Mus81-2* in any species. Consequently, we decided to only use *Mus81-1* in further analyses.

Our reconstruction of the *Spo11* phylogeny (Figure 2.1e) showed that the three *Spo11* genes present in maize, *Spo11-1*, *Spo11-2*, and *Spo11-3*, are orthologs of the three Arabidopsis *SPO11* genes. Interestingly, we found that the rice genome contains also a fourth homolog of *SPO11*, *SPO11-4*, which is not present in the maize or sorghum genomes. Moreover, we did not find any potential orthologs of *SPO11-4* among plant *SPO11* sequences available in the GenBank. These data suggests that *SPO11-4* originated after the maize-rice divergence but the exact origin of the gene is not clear.

Figure 2.1a

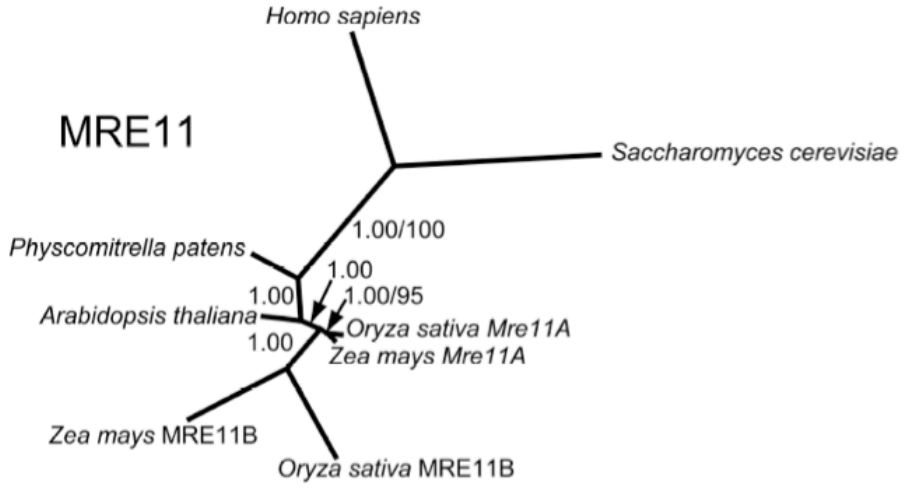


Figure 2.1b

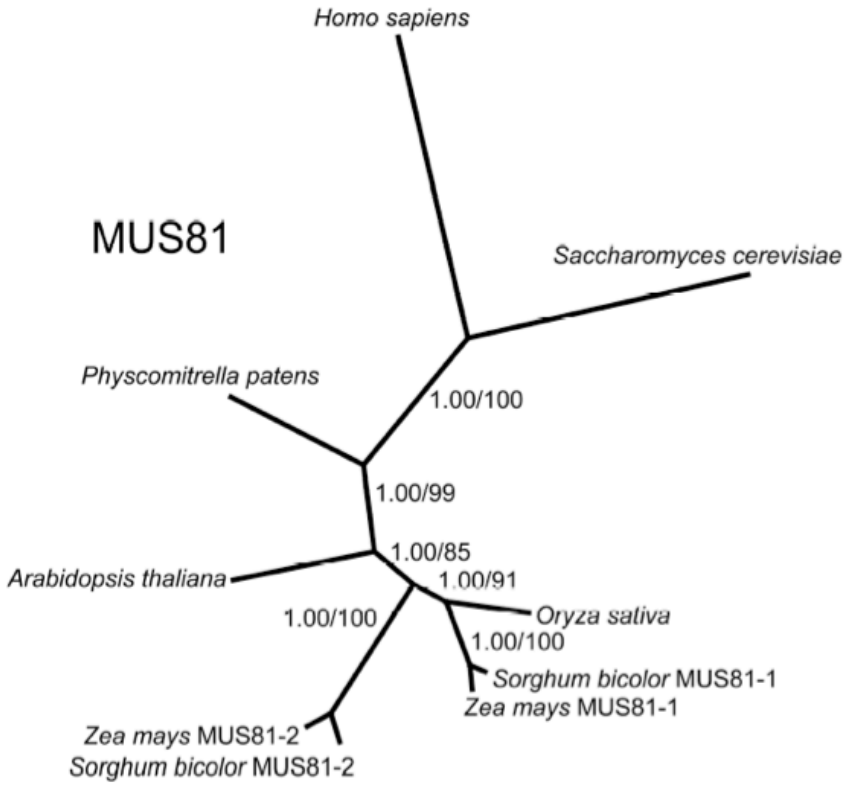


Figure 2.1c

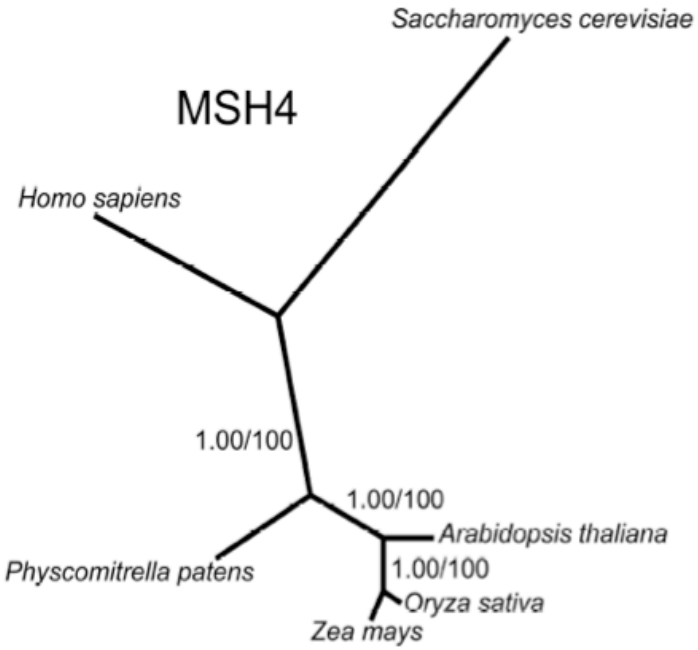


Figure 2.1d

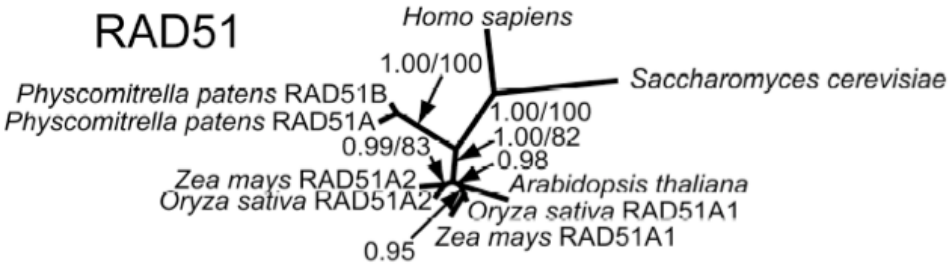


Figure 2.1e

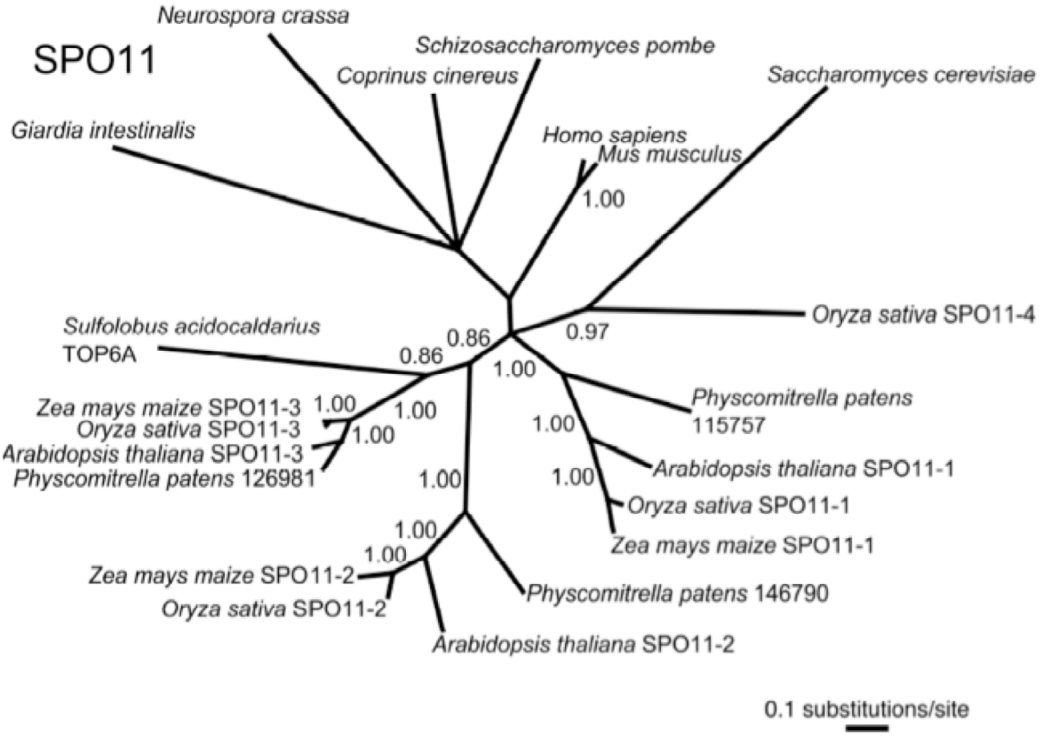


Figure 2.1(a-e): Phylogeny reconstructions of the meiotic recombination proteins in eukaryotes based on the Bayesian and maximum parsimony methods. For all proteins, except SPO11 (2.1e), the Bayesian trees and the maximum parsimony trees were identical. For SPO11, the Bayesian tree is shown. Numbers next to branches are posterior probabilities and bootstrap support values. Only posterior probability values of 0.8 and higher and bootstrap support values of 70% and higher are reported.

2.3.3 Evolutionary histories of recombination genes in maize and teosintes

To study the evolution patterns of recombination genes in maize and teosinte, we sequenced eleven genes, *Dmc1*, *Mlh1*, *Mre11A*, *Mre11B*, *Msh4*, *Mus81-1*, *Rad51A1*, *Rad51A2*, *Sgs1*, *Spo11-1*, and *Spo11-2*, from a set of 31 diverse maize inbred lines, which together represent more than 85% of the allelic diversity in maize (see Materials and Methods). We also included nine lines of *Zea mays* ssp. *parviglumis* (Balsas teosinte), the direct wild ancestor of cultivated maize (Matsuoka et al. 2002; Piperno et al. 2009). In addition, we used several more distantly related teosinte accessions representing *Z. mays* ssp. *mexicana*, *Z. mays* ssp. *huehuetenangensis*, *Z. diploperennis*, and *Z. luxurians* (see Materials and Methods). For each gene, we sequenced the entire coding region (Table 2.1), up to 240 bp of the 5' region, and between 826 bp and 3605 bp of intron fragments. For a few genes, we also obtained short 3' fragments.

Just like phylogenies are used to elucidate gene evolution patterns in different species, gene genealogies are helpful tools for understanding gene evolution within species. We used the Bayesian analysis method to reconstruct genealogies of genomic sequences of the eleven recombination genes in maize and teosintes (Figure 2.2a-2.2f).

Figure 2.2a:

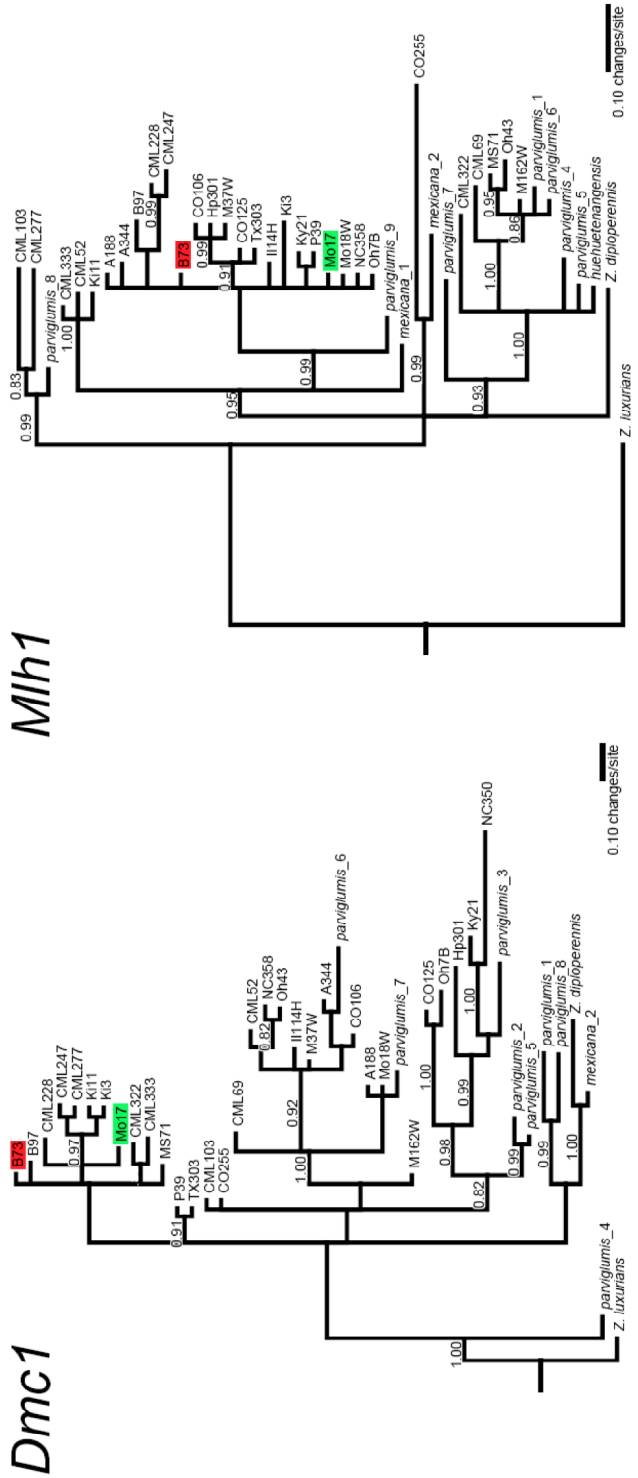

Mlh1
CML103
CML277
parviglumis_8
CML333
CML52
Ki11
A188
A344
B97
CML228
CML247
B73
CO106
Hp301
M37W
CO125
Tx303
Il14H
Ki3
Ky21
P39
Mo17
Mo18W
NC358
Oh7B
parviglumis_9
mexicana_1
CO255
mexicana_2
parviglumis_7
CML322
CML69
MS71
Oh43
M162W
parviglumis_1
parviglumis_6
parviglumis_4
parviglumis_5
huehuetenangensis
Z. diploperennis
Z. luxurians
0.10 changes/site
Dmc1
B73
B97
CML228
CML247
CML277
Ki11
Ki3
Mo17
CML322
CML333
MS71
CML69
CML103
CO255
P39
Tx303
CML52
NC358
Oh43
Il14H
M37W
A344
CO106
parviglumis_6
A188
Mo18W
parviglumis_7
M162W
CO125
Oh7B
Hp301
Ky21
NC350
parviglumis_3
parviglumis_2
parviglumis_5
parviglumis_1
parviglumis_8
Z. diploperennis
mexicana_2
parviglumis_4
Z. luxurians
0.10 changes/site

Figure 2.2b:

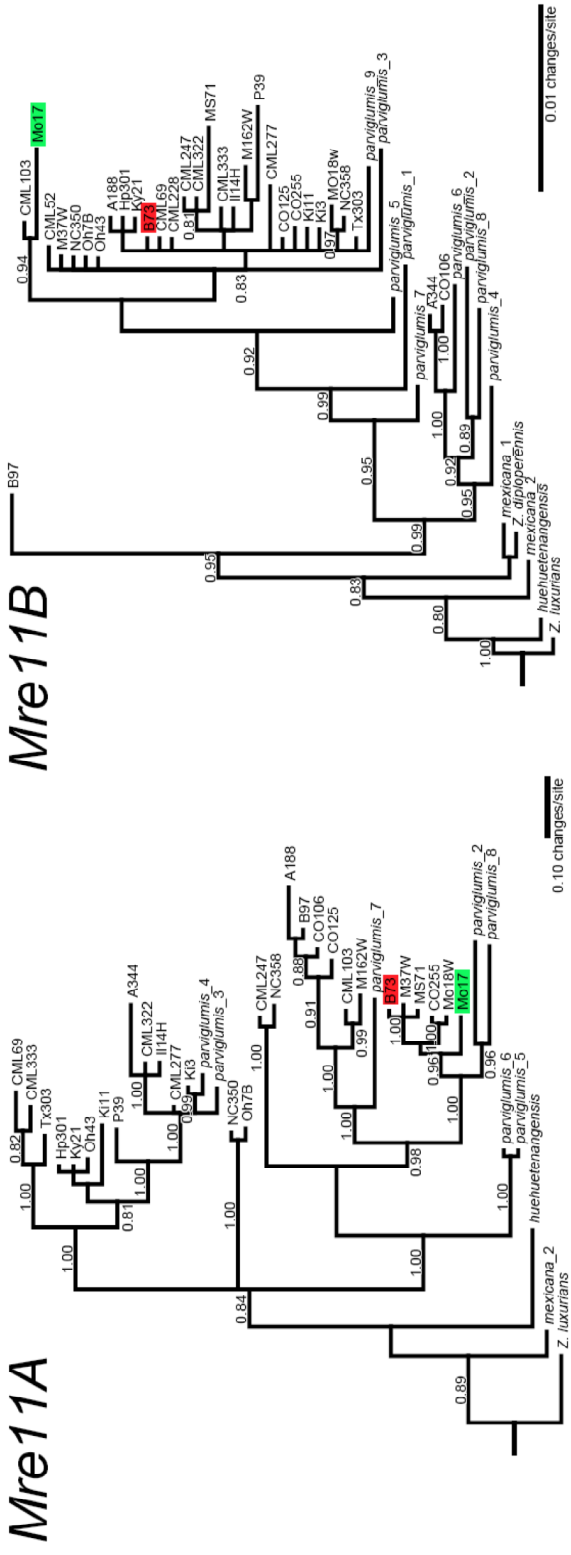

Mre11B
B97
CML103
Mo17
CML52
M37W
MS71
Oh7B
Oh43
A188
Hp301
Ky21
B73
CML69
CML228
CML247
CML322
MS71
CML333
Il14H
M162W
P39
CML277
CO125
CO255
Ki11
Ki3
Mo18w
NC358
Tx303
parviglumis_9
parviglumis_3
parviglumis_5
parviglumis_1
parviglumis_7
A344
CO106
parviglumis_6
parviglumis_2
parviglumis_8
parviglumis_4
mexicana_1
Z. diploperennis
mexicana_2
huehuetenangensis
Z. luxurians
0.01 changes/site
Mre11A
CML69
CML333
Tx303
Hp301
Ky21
Oh43
Ki11
P39
A344
CML322
Il14H
CML277
Ki3
parviglumis_4
parviglumis_3
NC350
Oh7B
CML247
NC358
A188
B97
CO106
CO125
CML103
M162W
parviglumis_7
B73
M37W
MS71
CO255
Mo18W
Mo17
parviglumis_2
parviglumis_8
parviglumis_6
parviglumis_5
huehuetenangensis
mexicana_2
Z. luxurians
0.10 changes/site

Figure 2.2c:

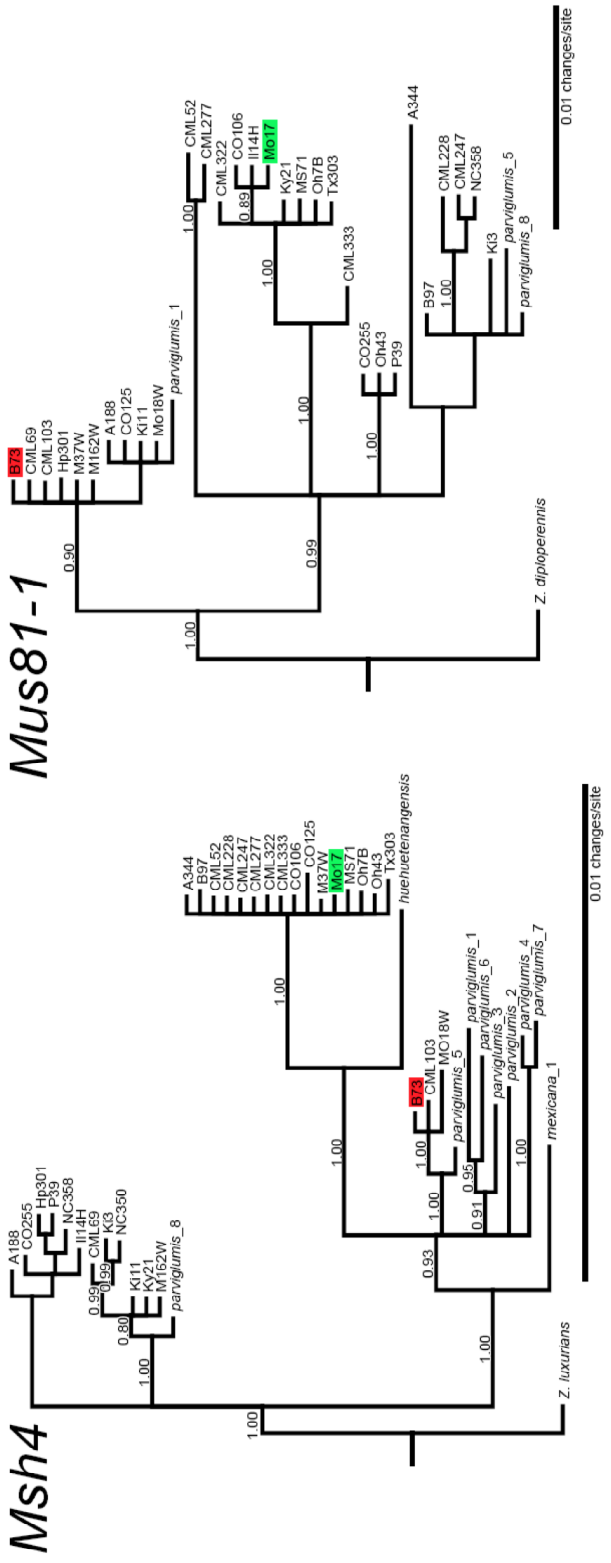

Mus81-1
B73
CML69
CML103
Hp301
M37W
M162W
A188
CO125
Ki11
Mo18W
parviglumis_1
CML52
CML277
CML322
CO106
Il14H
Mo17
Ky21
MS71
Oh7B
Tx303
CML333
CO255
Oh43
P39
A344
B97
CML228
CML247
NC358
Ki3
parviglumis_5
parviglumis_8
Z. diploperennis
0.01 changes/site
Msh4
A188
CO255
Hp301
P39
NC358
Il14H
CML69
Ki3
NC350
Ki11
Ky21
M162W
parviglumis_8
A344
B97
CML52
CML228
CML247
CML277
CML322
CML333
CO106
CO125
M37W
Mo17
MS71
Oh7B
Oh43
Tx303
huehuetenangensis
B73
CML103
MO18W
parviglumis_5
parviglumis_1
parviglumis_6
parviglumis_3
parviglumis_2
parviglumis_4
parviglumis_7
mexicana_1
Z. luxurians
0.01 changes/site

Figure 2.2d:

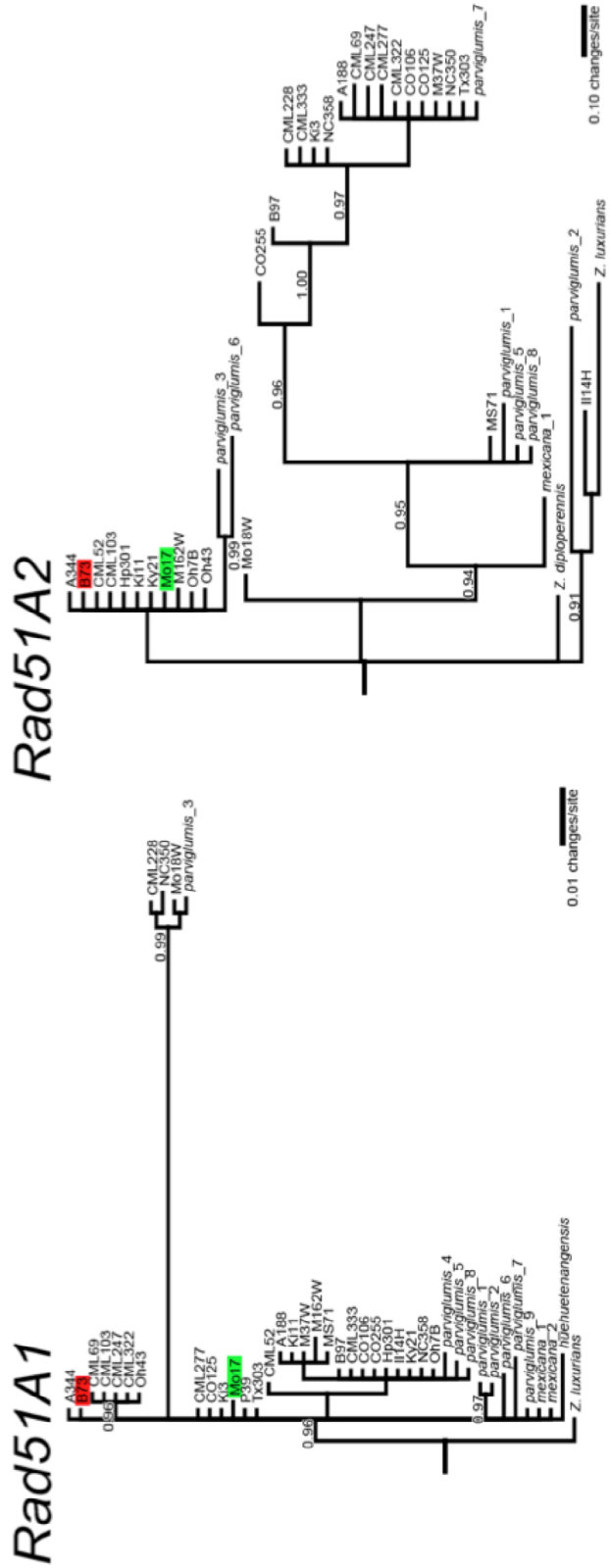

Rad51A2
A344
B73
CML52
CML103
Hp301
KI11
KY21
M017
M162W
Oh7B
Oh43
parviglumis_3
parviglumis_6
Mo18W
CO255
B97
CML228
CML333
Ki3
NC358
A188
CML69
CML247
CML277
CML322
CO106
CO125
M37W
NC350
Tx303
parviglumis_7
MS71
parviglumis_1
parviglumis_5
parviglumis_8
mexicana_1
Z. diploperennis
parviglumis_2
Il14H
Z. luxurians
0.10 changes/site
0.99
0.96
1.00
0.97
0.95
0.94
0.91
Rad51A1
A344
B73
CML69
CML103
CML247
CML322
Oh43
CML277
CO125
Ki3
M017
B97
Tx303
CML52
A188
KI11
M37W
M162W
MS71
B97
CML333
CO106
CO255
Hp301
Il14H
KY21
NC358
Oh7B
parviglumis_4
parviglumis_5
parviglumis_8
parviglumis_1
parviglumis_2
parviglumis_6
parviglumis_7
parviglumis_9
mexicana_1
mexicana_2
Z. luxurians
huehuetenangensis
CML228
NC350
Mo18W
parviglumis_3
0.01 changes/site
0.99
0.96
0.97

Figure 2.2e:


Spo11-2
CML52
CML247
CML322
CO255
parviglumis_5
A188
Ky21
B73
B97
CML69
CML103
CML228
Hp301
Ki3
Ki11
Mo17
Mo18W
M37W
MS71
NC350
NC358
Oh7B
Tx303
A344
CML333
Il14H
P39
parviglumis_4
CO106
M162W
Oh43
parviglumis_2
parviglumis_7
parviglumis_8
CML277
parviglumis_1
parviglumis_6
parviglumis_3
parviglumis_9
mexicana_2
Z. diploperennis
huehuetenangensis
Z. luxurians
0.10 changes/site
Spo11-1
A188
A344
B73
Hp301
Il14H
M37W
Tx303
B97
CML52
CML69
CML103
CML228
CML247
CML277
CML322
CO106
CO125
CO255
Ki3
Ki11
Ky21
M162W
MS71
Mo18W
Oh7B
NC350
NC358
parviglumis_1
P39
Mo17
Oh43
parviglumis_8
parviglumis_3
parviglumis_7
parviglumis_6
parviglumis_9
mexicana_1
mexicana_2
parviglumis_2
parviglumis_5
parviglumis_4
huehuetenangensis
Z. luxurians
0.01 changes/site

Figure 2.2f:

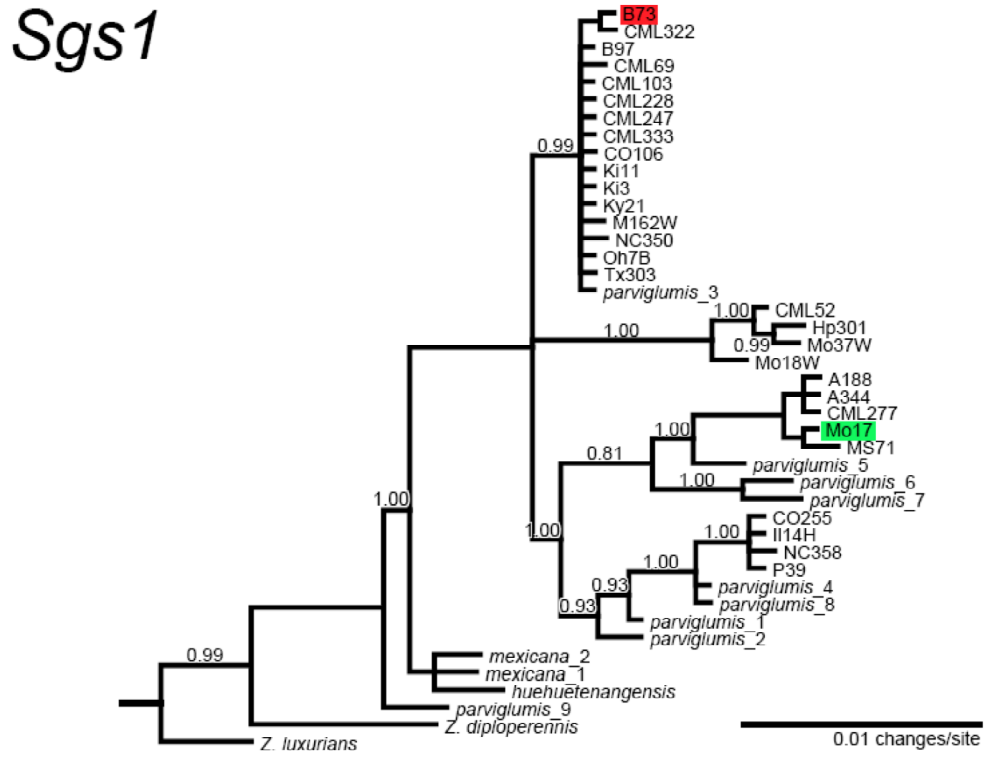


Figure 2.2a-2.2f: Gene genealogy reconstructions of genomic sequences of the eleven recombination genes in maize and teosintes using Bayesian analysis method.

Genealogies of different recombination genes exhibit highly conflicting topologies (Figure 2.2a-2.2f). Considerable differences were found not only between genes encoding different proteins but also between both homologs of the three duplicated genes (*Mre11*, *Rad51A*, and *Spo11*). Some genes showed relatively simple genealogies with a few groups of similar and/or identical haplotypes, while other genes exhibited fairly complex genealogy patterns. For example, most maize inbreds exhibit identical or nearly identical *Spo11-1* haplotypes (alleles) (Figure 2.2e). In contrast, the *Spo11-2* genealogy tree shows several distinct major clades. We traced the haplotypes of the recombination genes carried by the B73 and Mo17 inbreds (Figure 2.2e), which belong to different heterotic groups (i.e. groups of inbreds that exhibit strong heterosis in the progeny when crossed to inbreds from other heterotic groups but not when crossed among themselves) and exhibit substantial allelic differences for genes across the genome (Liu et al. 2003). We found that B73 and Mo17 exhibit similar or identical alleles of *Dmc1*, *Mlh1*, *Mre11A*, *Rad51A1*, *Rad51A2*, and *Spo11-2*, but diverged alleles of *Mre11B*, *Msh4*, *Mus81-1*, *Sgs1*, and *Spo11-1* (Figure 2.2).

Although gene genealogy trees are useful for understanding the overall relatedness of different haplotypes, the evolutionary patterns displayed by trees are often serious oversimplifications. In addition to mutation accumulation and selection, gene evolution patterns are likely to be affected by conflicting forces of recombination and gene flow, which cannot be accurately portrayed using phylogenetic trees (Posada and Crandall 2001). Also, in intra-specific gene

genealogies, ancestral haplotypes persisting in the population together with their descendants may be difficult to be properly position within the tree. Unlike trees, phylogenetic networks can reflect the phylogenetic uncertainties caused by hybridization and recombination events, and more accurately depict the patterns of gene evolution (Huson and Bryant 2006). Consequently, to better understand the genealogy of the eleven recombination genes in maize and teosintes, we constructed parsimony split networks (Figure 2.3a-2.3h). Examination of these networks showed that phylogenetic uncertainties, which could represent potential recombination events in the ancestry of the extant alleles, could indeed be found within the genealogies of the genomic regions of the eleven genes. Nevertheless, the overall haplotype grouping of each gene in the network analyses was usually in a good agreement with the haplotype grouping in the gene genealogy trees.

Figure 2.3 (a-h): Parsimony split networks for eleven recombination genes in maize and teosinte. Gene names are mentioned on top of each corresponding parsimony network tree.

Figure 2.3a:

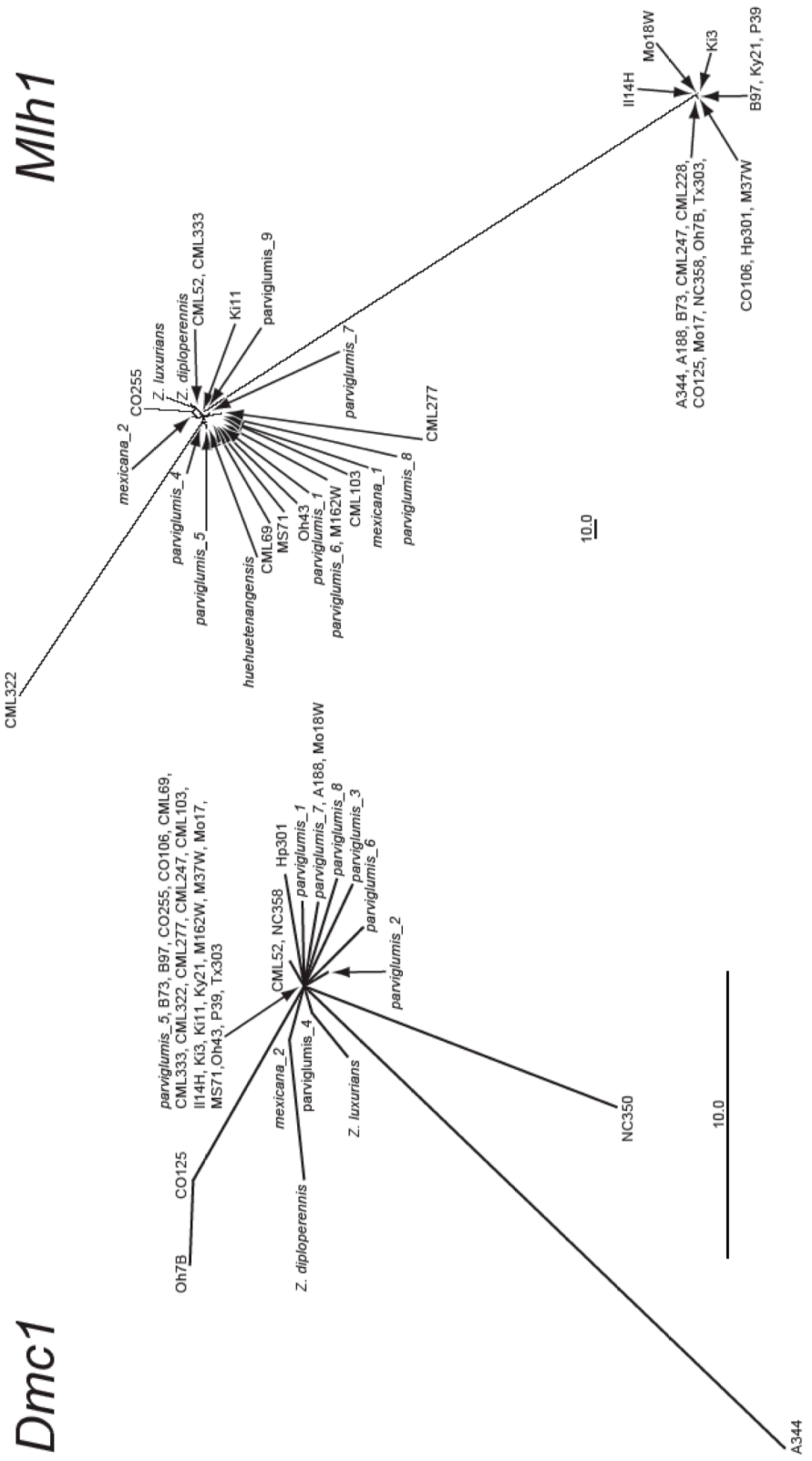

Mlh1
Dmc1
CML322
mexicana_2
CO255
Z. luxurians
Z. diploperennis
CML52, CML333
Ki11
parviglumis_9
parviglumis_7
CML277
parviglumis_4
parviglumis_5
huehuetenangensis
CML69
MS71
Oh43
parviglumis_1
parviglumis_6, M162W
CML103
mexicana_1
parviglumis_8
10.0
Il14H
Mo18W
Ki3
B97, Ky21, P39
A344, A188, B73, CML247, CML228, CO125, Mo17, NC358, Oh7B, Tx303
CO106, Hp301, M37W
Oh7B
CO125
parviglumis_5, B73, B97, CO255, CO106, CML69, CML333, CML322, CML277, CML247, CML103, Il14H, Ki3, Ki11, Ky21, M162W, M37W, Mo17, MS71,Oh43, P39, Tx303
mexicana_2
Z. diploperennis
parviglumis_4
CML52, NC358
Hp301
parviglumis_1
parviglumis_7, A188, Mo18W
parviglumis_8
parviglumis_3
parviglumis_6
parviglumis_2
Z. luxurians
NC350
10.0
A344

Figure 2.3b:

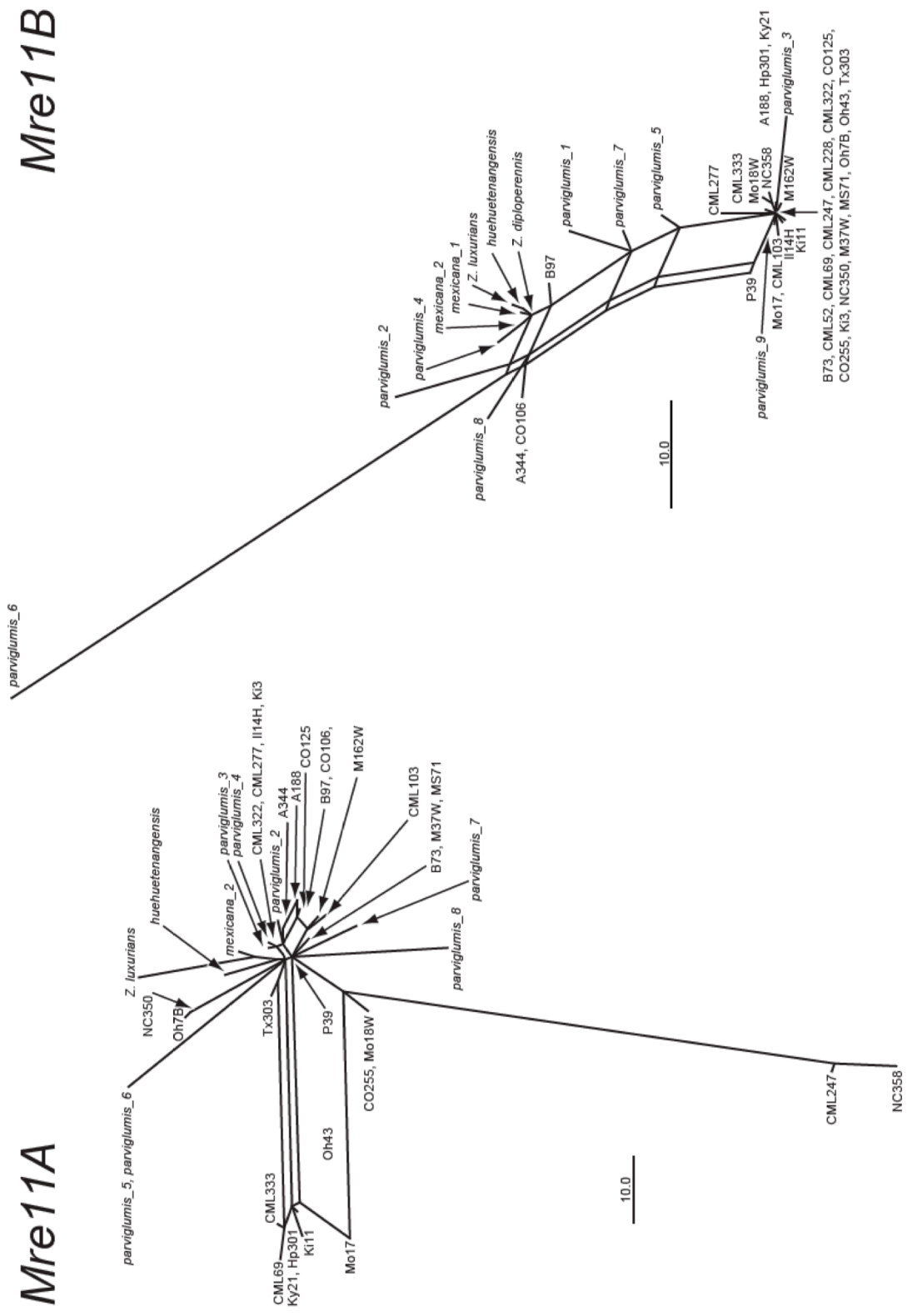

Mre11B
parviglumis_6
parviglumis_2
parviglumis_4
mexicana_2
mexicana_1
Z. luxurians
huehuetenangensis
Z. diploperennis
B97
parviglumis_1
parviglumis_7
parviglumis_5
CML277
CML333
Mo18W
NC358
M162W
A188, Hp301, Ky21
parviglumis_3
P39
Mo17, CML103
Il14H
Ki11
parviglumis_9
B73, CML52, CML69, CML247, CML228, CML322, CO125,
CO255, Ki3, NC350, M37W, MS71, Oh7B, Oh43, Tx303
parviglumis_8
A344, CO106
10.0
Mre11A
parviglumis_5, parviglumis_6
Z. luxurians
huehuetenangensis
mexicana_2
parviglumis_3
parviglumis_4
CML322, CML277, Il14H, Ki3
parviglumis_2
A344
A188
CO125
B97, CO106,
M162W
CML103
B73, M37W, MS71
parviglumis_7
parviglumis_8
NC350
Oh7B
Tx303
P39
CO255, Mo18W
CML247
NC358
CML333
Oh43
CML69
Ky21, Hp301
Ki11
Mo17
10.0

Figure 2.3c:

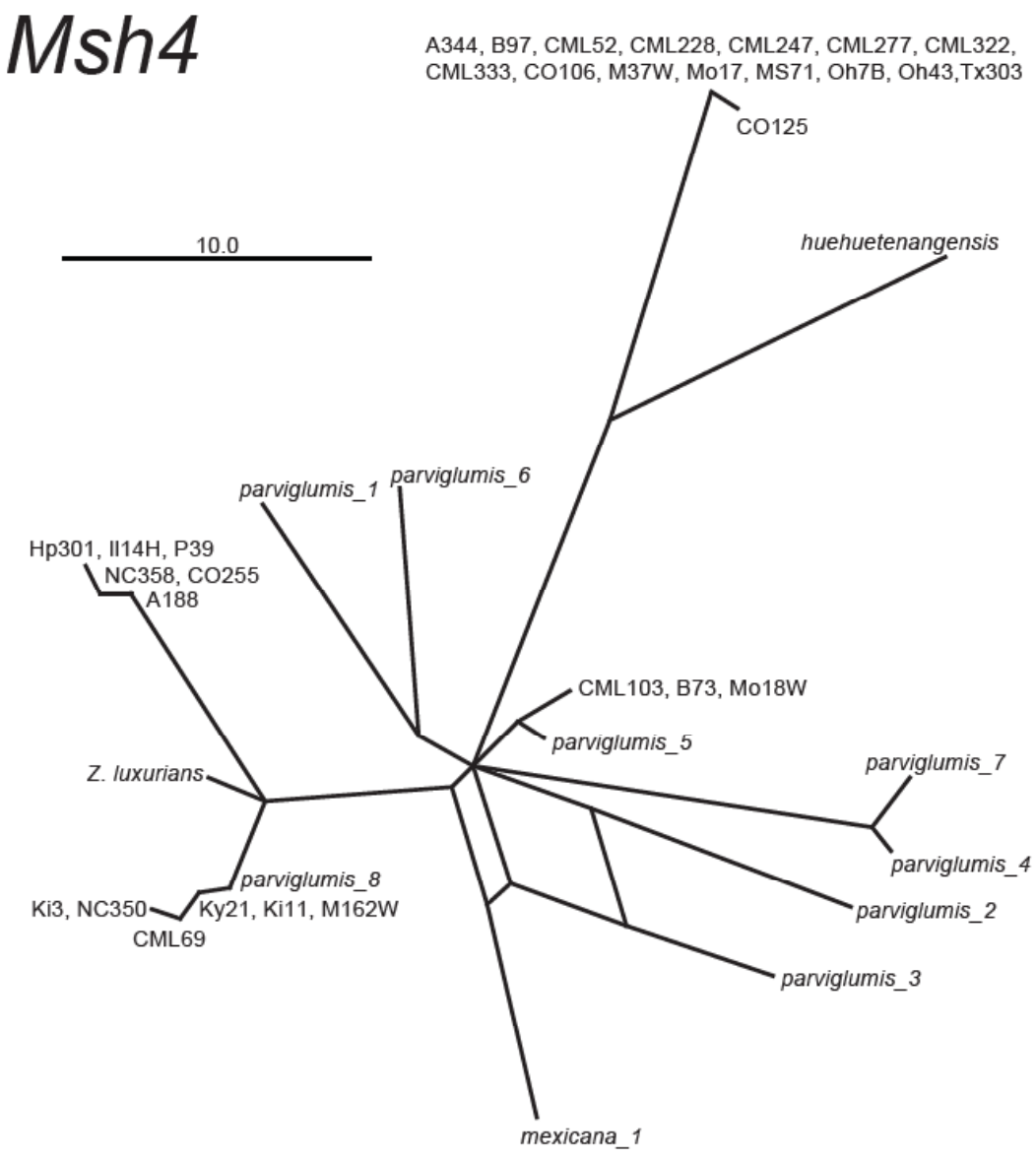

Msh4
A344, B97, CML52, CML228, CML247, CML277, CML322,
CML333, CO106, M37W, Mo17, MS71, Oh7B, Oh43,Tx303
CO125
10.0
huehuetenangensis
parviglumis_1
parviglumis_6
Hp301, Il14H, P39
NC358, CO255
A188
CML103, B73, Mo18W
parviglumis_5
Z. luxurians
parviglumis_7
parviglumis_4
parviglumis_8
Ki3, NC350
Ky21, Ki11, M162W
CML69
parviglumis_2
parviglumis_3
mexicana_1

Figure 2.3d

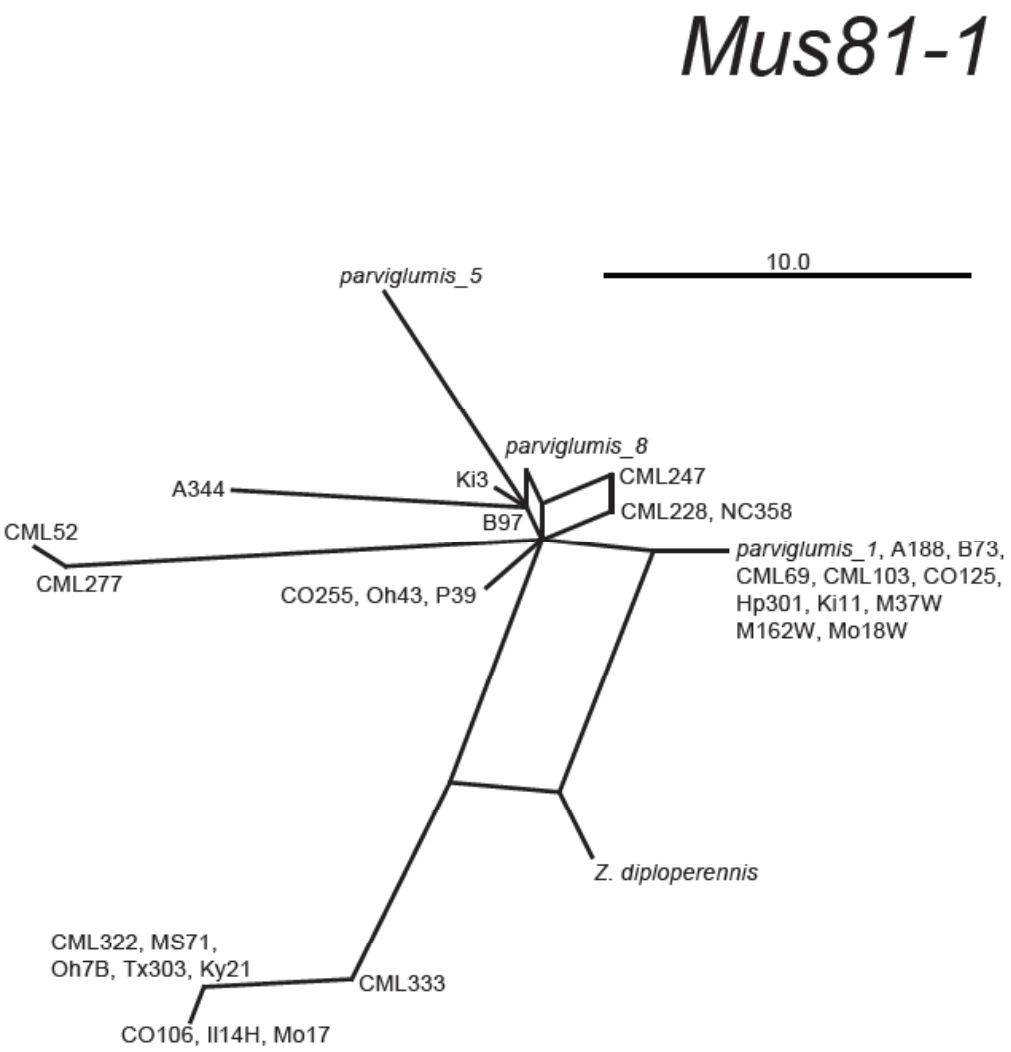

Mus81-1
10.0
parviglumis_5
parviglumis_8
Ki3
CML247
A344
CML228, NC358
B97
CML52
parviglumis_1, A188, B73,
CML69, CML103, CO125,
Hp301, Ki11, M37W
M162W, Mo18W
CML277
CO255, Oh43, P39
Z. diploperennis
CML322, MS71,
Oh7B, Tx303, Ky21
CML333
CO106, Il14H, Mo17

Figure 2.3e:

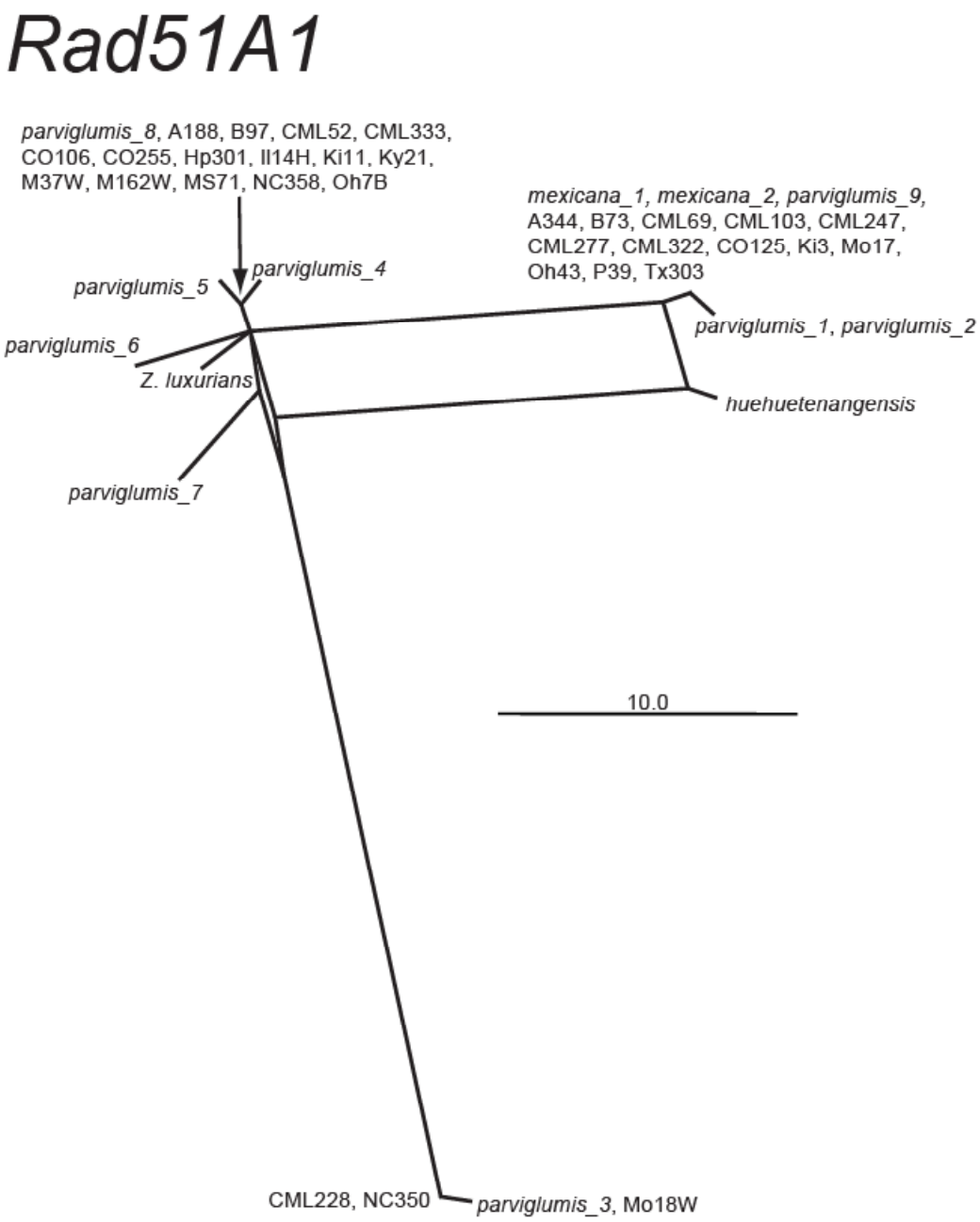

Rad51A1
parviglumis_8, A188, B97, CML52, CML333, CO106, CO255, Hp301, Il14H, Ki11, Ky21, M37W, M162W, MS71, NC358, Oh7B
mexicana_1, mexicana_2, parviglumis_9, A344, B73, CML69, CML103, CML247, CML277, CML322, CO125, Ki3, Mo17, Oh43, P39, Tx303
parviglumis_4
parviglumis_5
parviglumis_1, parviglumis_2
parviglumis_6
Z. luxurians
huehuetenangensis
parviglumis_7
10.0
CML228, NC350
parviglumis_3, Mo18W

Figure 2.3f:

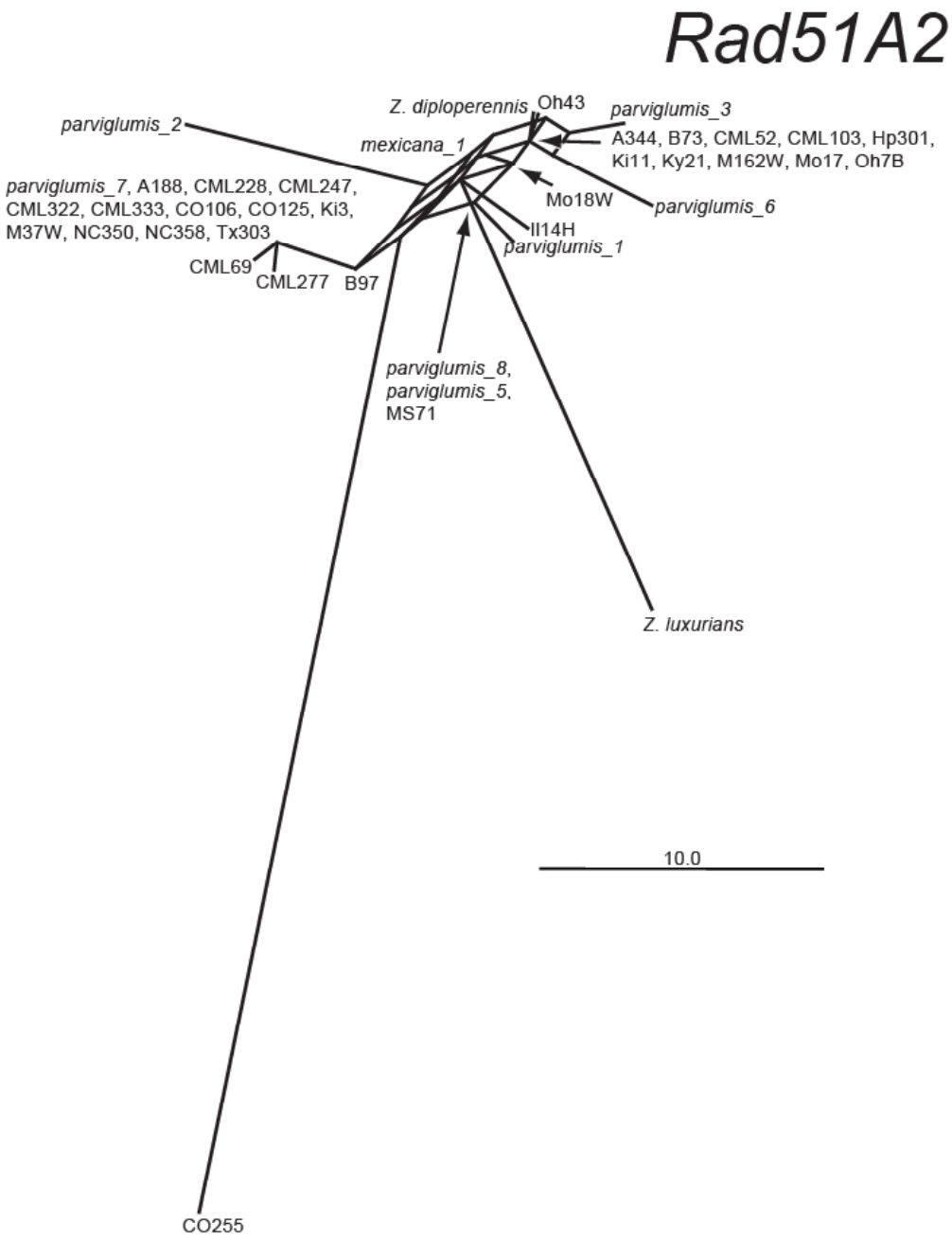

Rad51A2
Z. diploperennis
Oh43
parviglumis_3
parviglumis_2
mexicana_1
A344, B73, CML52, CML103, Hp301,
Ki11, Ky21, M162W, Mo17, Oh7B
parviglumis_7, A188, CML228, CML247,
CML322, CML333, CO106, CO125, Ki3,
M37W, NC350, NC358, Tx303
Mo18W
parviglumis_6
Il14H
parviglumis_1
CML69
CML277
B97
parviglumis_8,
parviglumis_5,
MS71
Z. luxurians
10.0
CO255

Figure 2.3g:

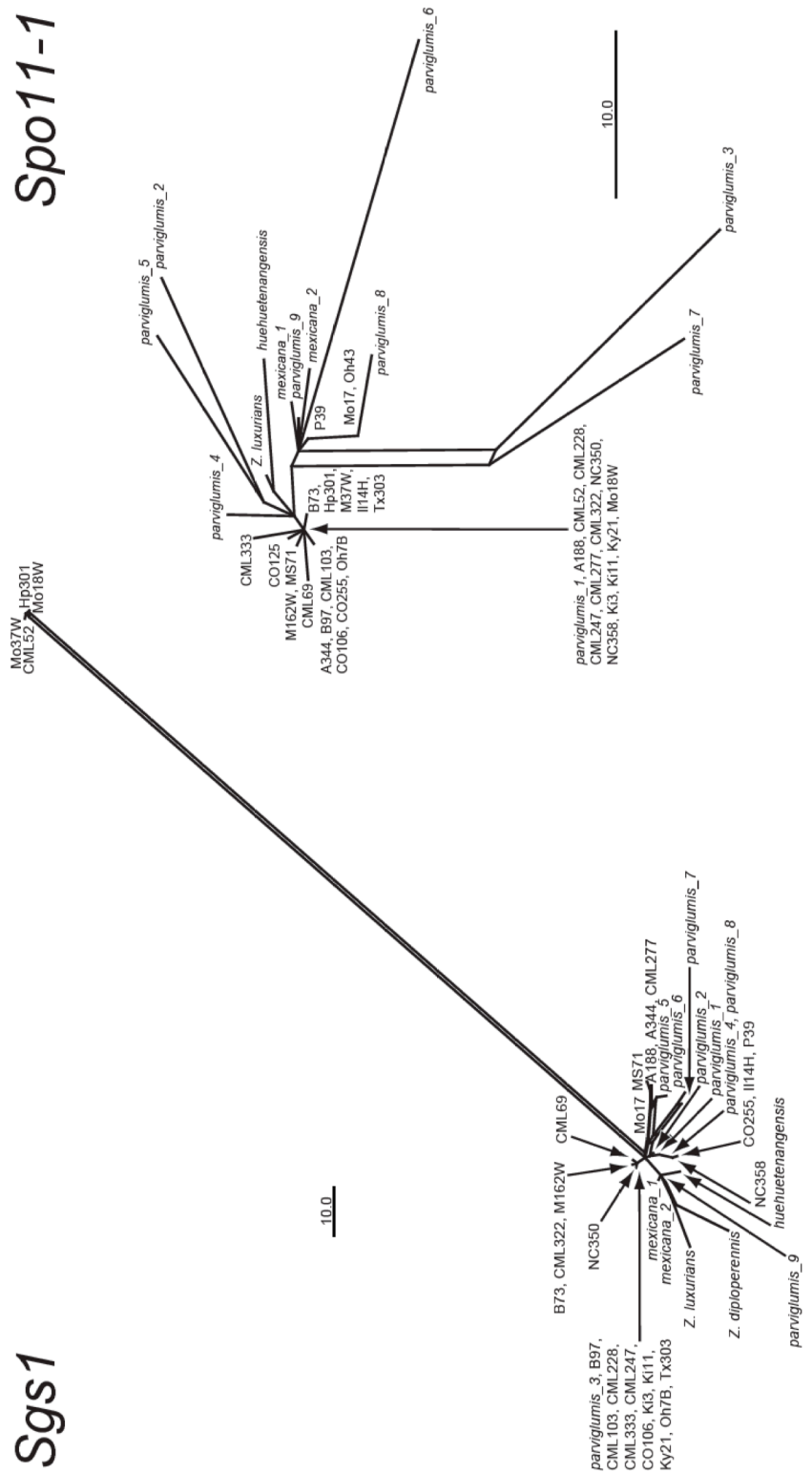
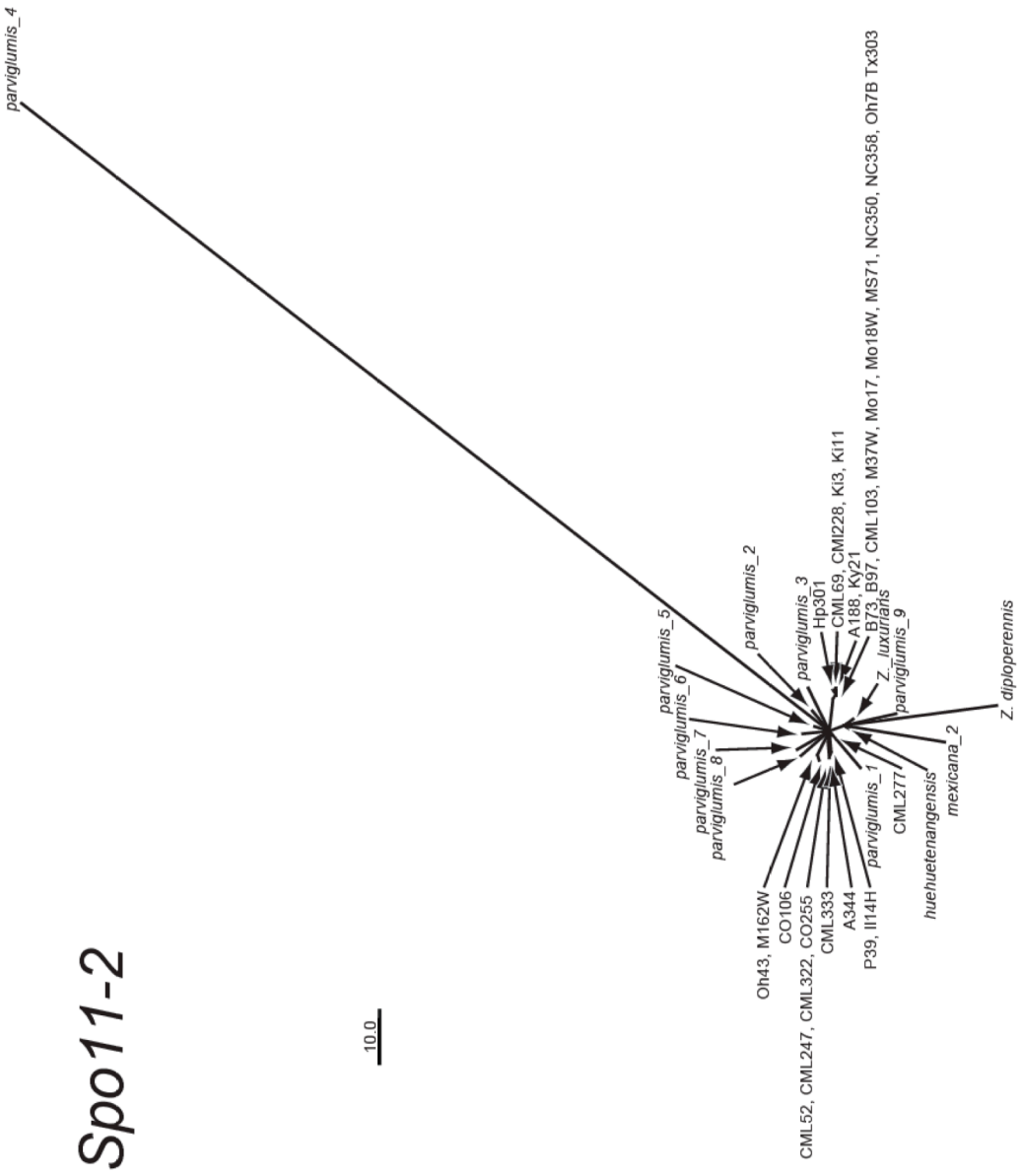

Spo11-1
Sgs1
10.0
10.0
Mo37W
CML52, Hp301
Mo18W
parviglumis_5
parviglumis_2
huehuetenangensis
mexicana_1
parviglumis_9
mexicana_2
parviglumis_6
parviglumis_8
Z. luxurians
P39
Mo17, Oh43
parviglumis_4
CML333
CO125
M162W, MS71
CML69
A344, B97, CML103,
CO106, CO255, Oh7B
B73,
Hp301,
M37W,
II14H,
Tx303
parviglumis_1, A188, CML52, CML228,
CML247, CML277, CML322, NC350,
NC358, Ki3, Ki11, Ky21, Mo18W
parviglumis_7
parviglumis_3
B73, CML322, M162W
CML69
NC350
Mo17, MS71
A188, A344, CML277
parviglumis_5
parviglumis_6
parviglumis_7
parviglumis_2
parviglumis_1
parviglumis_4, parviglumis_8
CO255, II14H, P39
NC358
huehuetenangensis
mexicana_1
mexicana_2
Z. luxurians
Z. diploperennis
parviglumis_9
parviglumis_3, B97,
CML103, CML228,
CML333, CML247,
CO106, Ki3, Ki11,
Ky21, Oh7B, Tx303

Figure 2.3h:


Spo11-2
10.0
parviglumis_4
parviglumis_5
parviglumis_6
parviglumis_7
parviglumis_8
parviglumis_2
parviglumis_3
Hp301
CML69, CML228, Ki3, Ki11
A188, Ky21
B73, B97, CML103, M37W, Mo17, Mo18W, MS71, NC350, NC358, Oh7B Tx303
Z. luxurians
parviglumis_9
Z. diploperennis
mexicana_2
huehuetenangensis
CML277
parviglumis_1
P39, Il14H
A344
CML333
CML52, CML247, CML322, CO255
CO106
Oh43, M162W

2.3.4 Sequence diversity and divergence rates

In addition to examining the evolutionary histories of the eleven genes, we also investigated their rates of sequence diversity at the nucleotide and amino acid levels (Table 2.3). For each gene, we examined the nucleotide sequence diversity using the Nei's π statistic (average number of nucleotide differences between any two sequences in the sample (Nei and Li 1979), and θ_W (the number of polymorphic nucleotide sites per nuclotide (Watterson 1975). We also calculated the number of polymorphic amino acid residues per residue. To calculate the π and θ_W values, we only considered the gene coding sequence rather then the entire genomic regions. Nucleotide diversity was significantly higher in non-coding regions of the genes (data not shown), and because each of the eleven genes contains a different number of introns, the genomic regions would include variable fractions of coding and non-coding sequences. We also confined these calculations to a set of 25 maize inbreds that was common for all eleven genes.

The nucleotide and amino acid sequence diversity measures showed substantial difference between different genes and also between paralogs of the three duplicated genes, *Mre11*, *Rad51A*, and *Spo11*. *Dmc1*, *Rad51A2*, and *Spo11-2* showed the most nucleotide sequence diversity, while *Rad51A1* and *Spo11-1* consistently exhibited the least nucleotide diversity. MRE11B, MUS81, and MLH1 showed the highest rate of amino acid sequence polymorphism, while RAD51A1, SPO11-1, and MSH4 showed the least amino acid polymorphism.

Table 2.3: Overall rates of nucleotide and amino acid sequence diversity in recombination genes.

Gene	Maize (set of 25 lines)				Eukaryote-wide
	Nucleotide sequence diversity in the coding region			Amino acid sequence diversity	Amino acid sequence divergence
	Length	π	θ_W	Polymorphic amino acid residues per residue	K scale factor[a]
Dmc1	1032	0.00332	0.00436	0.0087	1.74
Mlh1	1416	0.00199	0.00299	0.0166	0.43
Mre11A	2118	0.00229	0.00288	0.0085	1.20
Mre11B	2019	0.00129	0.00330	0.0223	1.00
Msh4	2412	0.00321	0.00231	0.0062	0.82
Mus81	1089	0.00217	0.00244	0.0200	0.76
Rad51A1	873	0.00162	0.00212	0.0000	2.18
Rad51A2	666	0.00692	0.00398	0.0118	2.22
Sgs1	1878	0.00273	0.00268	0.0145	0.42
Spo11-1	954	0.00096	0.00167	0.0052	0.73
Spo11-2	984	0.00444	0.00458	0.0131	-[b]

[a] The K scale factor describes the sequence divergence rate by measuring the overall branch length in the gene phylogenetic tree (Soria-Carrasco et al., 2007). Smaller K values indicate more divergence.

[b] The eukaryote-wide rate of sequence divergence could not be calculated as *Spo11-2* forms a separate lineage in plants and is absent from other extant groups of eukaryotes

To examine whether the differences in the diversity rates among maize recombination genes are simply a reflection of the fact that some recombination genes tend to evolve faster than others overall in all eukaryotes, we compared the sequence diversity measures for the set of 25 maize inbreds to the divergence rates of the recombination genes among eukaryotes. To measure the divergence rates across eukaryotes, we used the K tree score approach (Soria-Carrasco et al. 2007). This method compares

different phylogenetic tress by calculating the K scale factors that scales all trees to the size of an arbitrarily selected reference tree. For the comparison, we constructed Bayesian trees based on protein sequences from *S. cerevisiae*, human, Arabidopsis, rice, and the B73 inbred of maize. As the reference tree, we used a tree based on concatenated sequences of all of the examined recombination proteins. This analysis showed that SGS1, MLH1, and SPO11-1 exhibited the highest divergence rates overall in eukaryotes while RAD51A and DMC1 evolved at the slowest pace (Table 2.3).

A comparison of the divergence rates in maize and all eukaryotes indicated that MRE11B, which shows the most amino acid sequence diversity in maize, exhibits only a moderate rate of sequence evolution overall in eukaryotes. RAD51A2, which shows the slowest divergence rate across eukaryotes shows moderate amino acid sequence diversity in maize compared to the other recombination genes. These data may suggest that MRE11B and RAD51A2 experienced a relative acceleration of amino acid sequence evolution in maize. In contrast, SPO11-1, which shows the fastest sequence divergence rate across all eukaryotes, displays one of the lowest amino acid sequence diversity rates in maize, suggesting that it may have experienced a relative deceleration of the amino acid sequence change in maize compared to the other recombination genes.

2.3.5 Selection patterns in recombination genes

To gain a more complete understanding of molecular evolution of the maize and teosinte recombination pathway, we examined the eleven recombination genes for signatures of selection. Because our goal was to examine selection patterns that relate to the protein functions, we limited the examination of selection signatures to the coding regions of the genes. Selection in regulatory gene regions is likely to be associated with adaptive changes during maize domestication in many classes of genes, for example in genes controlling plant architecture, such as *teosinte branched 1* (*tb1*) (Wang et al. 1999). However, such situation is unlikely to be the case for meiotic recombination genes, since there is little evidence for tight transcriptional regulation of these genes in plants. Even genes that are highly regulated in other species, for example *Spo11* or *Hop2*, are ubiquitously expressed in plants (Grelon et al. 2001; Schommer et al. 2003; Pawlowski et al. 2009).

To examine the patterns of selection in the recombination genes we took several complementary approaches. First, we searched for evidence of selection in maize and Balsas teosinte by examining the distribution of alleles in the population using frequency spectrum-based statistics Tajima's D (Tajima 1989), and Fu and Li's D and F (Fu and Li 1993). The Tajima's D test calculates the average number nucleotide differences between pairs of sequences as a fraction of the total number of segregating sites in the sequence. The Fu and Li's D and F tests are based on a similar algorithm but compare the polarity of nucleotide changes to an outgroup. We used *Z. luxurians* as the outgroup for both maize and *Z. mays* ssp. *parviglumis* for all

genes, except *Mus81-1*, for which we used *Z. diploperennis*, as we could not generate a sufficiently long sequence fragment for *Z. luxurians Mus81-1*.

Using the Tajima's D and Fu and Li's D and F tests, we found significant departures from the neutral evolution pattern in six genes in maize, *Dmc1*, *Mre11B*, *Msh4*, *Sgs1*, and *Spo11-2* (Table 2.4). *Dmc1*, *Mre11B*, *Msh4*, *Sgs1*, and *Spo11-2*, showed significant positive values of Fu and Li's D. *Mre11B* and *Msh4* also exhibited significant positive values of Fu and Li's F. *Rad51A2* showed a significant positive value of Tajima's D. In contrast, significant negative values for Fu and Li's D and F were found for *Mre11B*. For *Mlh1*, we also obtained strongly negative values of Fu and Li's D and F. However, these values were not statistically significant although they were nearly significant ($P < 0.10$).

We hypothesized that they may reflect selection acting only on a relatively small region of the *Mlh1* gene, while the other gene regions do not exhibit significant departures from neutrality. To test this hypothesis, we separately examined nucleotide sequences corresponding to four domains in the MLH1 protein, the N-terminal ATPase domain, the transducer domain located in the center of the protein, a linker region located between the ATPase and the transducer domains, and the C-terminal protein region. We found that the linker region showed statistically significant negative values of Fu and Li's F and D (Table 2.4).

Table 2.4: Results of frequency spectrum-based tests to detect selection patterns in coding regions of recombination genes in maize.

Gene	Length	N	Fu & Li D		Fu & Li F		Tajima's D		
			Value	Percentile of CS-generated distribution[a]	Value	Percentile of CS-generated distribution[a]	Value	Percentile of empirical distribution[b]	Percentile of CS-generated distribution[a]
Dmc1	1035	30	1.73842*	94.0	1.8524*	84.3	-0.50340	32.6	14.2
Mlh1	1589	26	-1.98286#	1.3	-2.14871#	1.5	-1.36469	9.7	3.7
linker domain	279	26	-2.74114**[c]	0.1	-2.87808**[c]	0.1	-1.74323**[c]	2.4	0.5
Mre11A	2121	29	-0.78960	7.1	-0.86882	6.5	-0.57980	29.0	12.6
Mre11B	1792	31	-2.54110*	0.1	-2.74082*	0.4	-1.74966*	2.1	1.5
Msh4	2415	31	1.82793*	91.0	2.14002*	87.5	1.65970	95.4	74.4
Mus81	1353	27	-0.45244	17.0	-0.51995	12.1	-0.39346	35.4	17.3
Rad51A1	1023	30	1.39229	91.5	1.22880	67.7	0.24754	57.6	33.3
Rad51A2	658	30	0.90418	40.0	1.73702#	75.7	2.61602**	99.9	96.5
Sgs1	3528	22	1.54931#	64.5	1.13756	42.5	-0.27752	38.5	18.9
Spo11-1	1158	30	0.70282	35.1	0.07890	18.3	-0.72935	26.5	10.0
Spo11-2	854	29	1.71266*	90.0	1.38134	60.2	-0.01630	47.1	23.8

Significance levels: ** = P < 0.01; * = P < 0.05; # = P < 0.1.

[a] Percentile relative to a distribution of test values generated by coalescent simulations (CS) assuming neutral evolution and a domestication bottleneck.

[b] Percentile relative to the genome-wide distribution of Tajima's D values from a survey of 703 random polymorphic loci in maize and *Z. mays ssp. parviglumis* (Wright et al. 2005).

[c] Calculated with Bonferroni correction for multiple testing.

Significant positive values for Fu and Li's D and F and Tajima's D suggest that the locus is under balancing (diversifying) selection (Tajima 1989; Fu and Li 1993). Significant negative values suggest a selective sweep. However, all three statistics are sensitive to demographic factors. A population bottleneck results in strongly positive Fu and Li's D and F and Tajima's D values (Nielsen 2005). Conversely, a population expansion may cause negative values of these statistics. To test the impact of selection vs. demographics, we used two available tools to assess the potential significance of the Tajima's D values: (i) we compared Tajima's D values for the eleven recombination genes to a genome-wide distribution of Tajima's D values using an approach similar to the one recently employed to examine selection patterns in cell cycle genes in Arabidopsis (Sterken et al. 2009), and (ii) we compared the Tajima's D values for the recombination genes to critical test values derived from coalescent simulations (CS) (Hudson 1990).

(i) A comparison of a statistic for a specific locus to a genome-wide distribution of the statistic is based on a tenet that while selection acts on individual loci, demographics is likely to have genome-wide effect. We weighed our Tajima's D values against the genome-wide distribution of Tajima's D values based on a survey of 703 random polymorphic loci (Wright et al. 2005). This survey was conducted on a set of 14 maize inbreds, all of which were included in our set. Loci that showed no polymorphism were excluded from the survey data, as Tajima's D statistic cannot be calculated for them. We found that the Tajima's D values for *Mre11B*, *Rad51A2*, and the linker region of *Mlh1* fall into the extreme 2.5% fractions of the genome-wide Tajima's D distribution (Table 2.4). We also conducted frequency spectrum-based

tests for the *Z. mays* ssp. *parviglumis* accessions in the same manner as we did in maize. However, we did not find significant departures from neutrality for any of the genes (Table 2.4), although it should be noted that the number of accessions that we used for *Z. mays* ssp. *parviglumis* was smaller than the number of inbred lines used in maize, which could have affected our results. As we did not have a genome-wide outgroup sequence data we could not make a similar comparison for the Fu and Li's D and F tests.

(ii) To further evaluate the significance of the frequency-spectrum statistics, we used Hudson's ms program (Hudson 2002) to examine the probability of obtaining the empirical Tajima's D for the recombination genes in maize values under neutral evolution in coalescent simulations. The simulations incorporated a population bottleneck under the parameters proposed for the maize domestication bottleneck by Wright (Wright et al. 2005). The results of this analysis (Table 2.4) were similar to those obtained from testing our Tajima's D values against the genome-wide Tajima's D distribution. Our empirical Tajima's D values for *Mre11B* and the linker region of *Mlh1* fell into the low extreme 2.5% of simulated Tajima's D values. *Rad51A2* was slightly less significant and located in the high extreme 3.5% of the simulated values.

Overall, the results of the two approaches to test the significance of our Tajima's D values suggest that the values for maize *Mre11B*, *Rad51A2*, and the linker region of *Mlh1* are unlikely to result from neutral evolution alone, even under a domestication bottleneck affecting population demographics.

To complement the frequency spectrum-based tests, we examined our data set using the Hudson-Kreitman-Aguade (HKA) test, which compares the patterns of

intraspecies diversity to interspecies divergence relative to a neutrally-evolving locus (Hudson, Kreitman, and Aguade 1987) and the McDonald-Kreitman (MK) test, which compares the intraspecies diversity to interspecies divergence at synonymous vs. non-synonymous sites in the gene coding region (McDonald and Kreitman 1991). *Z. luxurians* was used as the outgroup (*Z. diploperennis* for Mus81-1). In the HKA test, we compared the evolution patterns in the recombination genes to a set of previously used neutral loci *adh1*, *bz2*, *csu1138*, *csu1177*, and *fus6* (Tenaillon et al. 2001). Neither HKA nor MK showed significant deviations from neutrality in any of the genes in maize and *Z. mays* ssp. *parviglumis*. We hypothesize that it may be difficult to find patterns of selection in recombination genes using these two tests, as the vast majority of amino acid residues in these proteins are likely to be functionally constrained and under purifying selection. Both tests are known to exhibit reduced power in such situations (Zhai, Nielsen, and Slatkin 2009). HKA have also been shown to have reduced power of detecting artificial selection sweeps if the initial frequency of the beneficial allele in the population is relatively high (Innan and Kim 2004).

As another approach to identify signatures of selection in recombination genes, we used the likelihood ratio test (LRT) approach (Nielsen and Yang 1998), which examines ω, the ratio of the non-synonymous substitution rate (dN) to the synonymous substitution rate (dS) in the gene's coding region. Because different regions in the protein sequence maybe under very different selection pressures and constraints, the method allows different ω values for individual codons, which makes it highly sensitive in detecting adaptive selection signatures. LRT can be used for

within-species comparisons as long as sequence diversity is high and the level of intra-genic recombination is low (Anisimova, Nielsen, and Yang 2003). Therefore, prior to the analyses, we examined the data set for presence of intra-genic recombination using the Genetic Algorithm Recombination detection method (GARD) (Kosakovsky Pond et al. 2006). We found no or low-level recombination frequencies that did not exceed the rates acceptable for the LRT analyses in any of the eleven genes. To further ensure that recombination did not affect the results of the test, for the genes where GARD detected recombination breakpoints, we individually tested the fragments separated by the recombination sites. In each case, the results of the LRT analysis were identical to the results of the analyses using the entire gene-coding region.

For the LRT analysis, we only used lines in which the entire coding region of the gene was available. To examine the patterns of selection, for each gene, several models were tested that assume different selection patterns, including positive selection, purifying selection, and neutral evolution, to identify which model fits best the empirical data (Table 2.5). In the analysis using maize inbreds and *Z. luxurians* as an outgroup, we found that coding regions of five genes *Mlh1*, *Mre11B*, *Mus81-1, Rad51A2,* and *Spo11-2* showed statistically significant signatures of positive selection (Table 2.5). For the six other genes, the models assuming positive selection showed low or zero probability, except for *Dmc1*, where the probability was slightly higher but still not within the statistically significant range (Table 2.5).

Table 2.5: Selection patterns in recombination genes in maize detected using the likelihood ratio test based on ω (dN/dS).

	Selection models[a]			% of codons			
Gene	M0 vs. M3 ($2\Delta l$)[b]	M1 vs. M2 ($2\Delta l$)[c]	M7 vs. M8 ($2\Delta l$)[d]	under purifying selection	under positive selection	evolving neutrally	positively selected sites[e]
Maize							
Dmc1	15.32**	4.12	6.14	100	-	-	-
Mlh1	50.48**	19.46**	20.54**	93.3	6.7	-	9 R*, 137 H, 339 V**, 393 A, 539 I*
Mre11A	1.06	0.46	0.12	100.0	-	-	-
Mre11B	24.64**	26.46**	14.23**	96.5	3.5	-	227 T*, 421 D*, 635 A*, 642 V**
Msh4	0.00	0.00	0.00	100.0	-	-	-
Mus81	22.64**	15.62**	15.74**	97.3	2.7	-	238 L**, 403 A*, 411 I*
Rad51A1	0.00	0.00	0.00	100.0	-	-	-
Rad51A2	43.76**	18.94**	20.62**	98.7	1.3	-	11 S**, 15 E*, 110 L**
Sgs1	0.00	0.00	0.00	100.0	-	-	-
Spo11-1	0.00	0.00	0.00	100.0	-		-
Spo11-2	80.52**	33.78**	35.18**	98.2	1.8	-	16 G*, 315 D**, 330 S**, 350 E**
Teosintes[f]							
Dmc1	0.00	0.00	0.00	100	-	-	-
Mlh1	0.00	0.00	0.00	100	-	-	-
Mre11A	28.90**	12.24**	12.34**	98.6	1.4	-	236 N, 314 R, 410 R, 657 R
Mre11B	0.00	0.00	0.00	100	-	-	-
Msh4	0.00	0.00	0.00	100	-	-	-
Rad51A1	0.00	0.00	0.00	100	-	-	-
Rad51A2	10.72	4.23	4.54	84.3	-	15.7	-
Sgs1	24.70**	10.36**	10.78**	94.5	5.5	-	6 P*, 9 V*, 332 Q*, 367 V*, 880 V*, 904 R*
Spo11-1	15.88**	4.98	5.04	91.8	-	8.2	-
Spo11-2	12.52	3.18	3.82	84.8	-	15.2	-

[a] Likelihood ratio tests of selection models. Ratios statistically significant at $P = 0.01$ are denoted with *

[b] Model M0 (one ratio) assumes a constant dN/dS ratio across all sites in the gene's coding region. Model M3 (discrete) assumes different proportions of discrete classes of sites with different ω values. df = 4.

[c] Model M1 (neutral model) assumes all sites are either under purifying selection ($\omega = 0$) or evolving neutrally ($\omega = 1$). Model M2 adds a third category of sites under positive selection ($\omega > 1$). df = 2.

[d] Model M7 (beta model) assumes that ω ranges from 0 (strong negative selection) to 1 (neutral evolution) and varies among sites in the gene's coding region according to a beta distribution. Model M8 (beta& ω) similarly to M7 assumes that ω varies among sites but allows that, in addition to ranging from 0 to 1, ω may take values > 1 (positive selection). df = 2.

[e] All sites listed show statistically significant evidence of positive selection with posterior probabilities posterior probabilities >0.90. Sites showing statistically significant evidence of positive selection with posterior probabilities > 0.95 are marked *. Sites with statistically significant evidence of positive selection with posterior probabilities > 0.99 are marked **.

[f] Analysis was not conducted for teosintes because sequences from too few lines were available.

We also conducted the LRT analyses using the sequences of the various teosinte accessions. In teosintes, we found selection patterns that were different from those in the maize inbred analysis (Table 2.5). *Mre11A* and *Sgs1* showed evidence of positive selection, while *Rad51A2* and *Spo11-2* exhibited neutral evolution patterns. None of the genes showing positive selection in the maize analysis exhibited positive selection patterns in teosintes.

To further understand the patterns of selection, we used the Bayes empirical Bayes method (BEB) (Yang, Wong, and Nielsen 2005), which can identify specific amino acid residues that are under positive selections. An additional benefit of BEB is that it is less sensitive to recombination than the LRT method (Anisimova, Nielsen,

and Yang 2003). We found that even in the genes that were identified as being under positive selection by LRT, the majority of amino acid residues were under purifying selection. Overall, in the genes found by LRT to experience positive selection, only 6% of all amino acid residues and about one-fifth of the polymorphic amino acid residues were targets of positive selection.

2.3.6 Patterns of amino acid substitutions

The ultimate test of adaptive evolution is identifying functional differences between polymorphic alleles of the gene. However, because of the complexities of protein-protein and protein-DNA interactions in the recombination pathway, this kind of analysis is not currently feasible. Nevertheless, numerous functional domains have been identified in recombination proteins and their three-dimensional structures have been elucidated. About one-quarter of amino acid residues that we found to be under positive selection in maize and teosinte are highly conserved among eukaryotes and about half of them are in known functional domains. To investigate whether the amino acid polymorphisms at residues targeted by positive selection are likely to be associated with functional changes in the proteins, we examined the positions of the polymorphic residues within the predicted three-dimensional proteins structures. These analyses identified several polymorphisms at amino acid residues found to be under positive selection that are likely to result in considerable protein structure changes. One of the notable polymorphisms was an alanine to phenylalanine change in RAD51A2 at residue 110 (Table 2.5), which is located in a conserved ATP-binding site. We found that this residue flanks a small loop on the protein surface that is

adjacent to the Mg^{2+}-binding pocket in the ATPase domain (Figure 4). Mg^{2+} binding is known to induce a conformational change in the RAD51 protein, which is required for DNA binding (Namsaraev and Berg 1998). Alanine is present at this residue in all eukaryotes that we examined, except for the two unusual RAD51A proteins in *Physcomitrella* (Markmann-Mulisch et al. 2002), which instead contain phenylalanine. Another example of a polymorphism likely to result in a functional change was a substitution of highly conserved hydrophilic serine by hydrophobic glycine at position 330 in SPO11-2 (Table 2.5). This residue is located on the surface of the TOPRIM domain in the SPO11 protein, on the face of the protein that touches the DNA (Nichols et al. 1999).

In addition to polymorphisms resulting in amino acid substitutions, we found a sequence polymorphism resulting in a premature stop codon in the *Mre11B* gene in *Z. mays* ssp. *parviglumis* line 5. This stop codon is expected to produce a non-functional protein truncated after the first eleven amino acids.

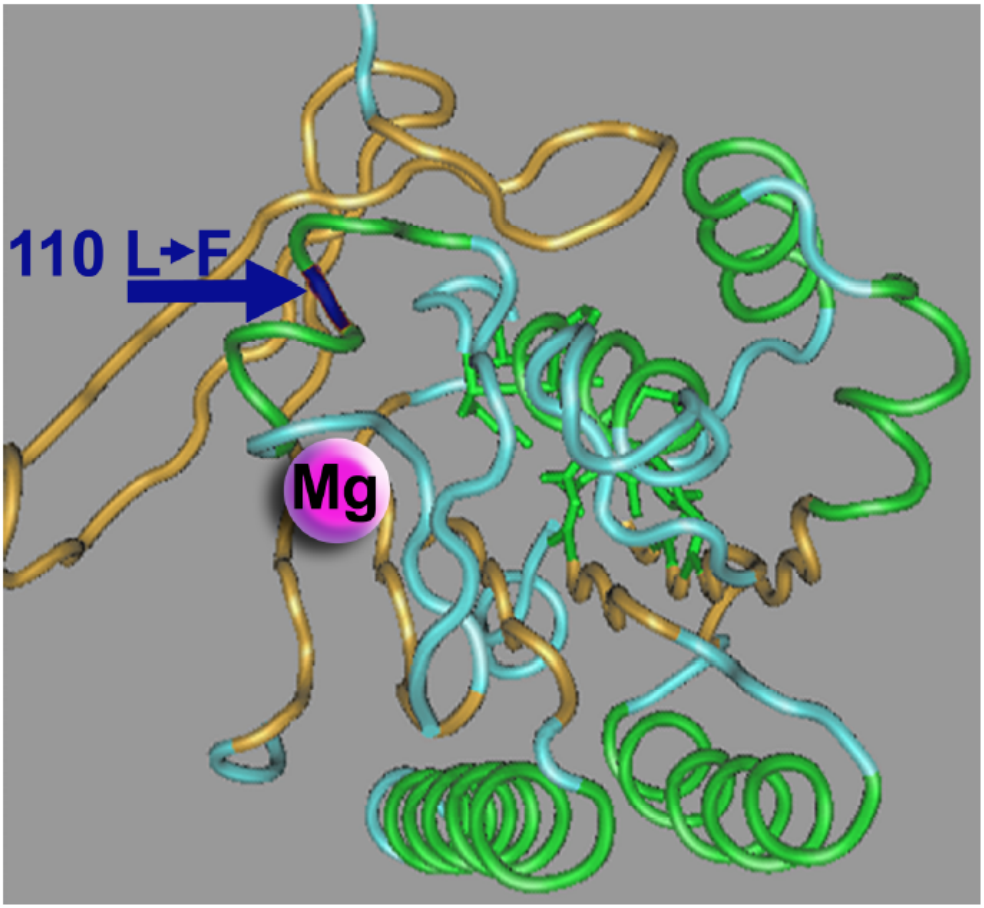


Figure 2.4: Flat projection of the three-dimensional structure of the BRC repeat in the RAD51A protein. The position corresponding to the polymorphic residue 110 in maize RAD51A2 is marked with an arrow. This amino acid is located at the base of a small loop on the protein surface that is adjacent to the Mg^{2+}-binding pocket in the ATPase domain.

2.4 Discussion

2.4.1 Evolution patterns in recombination genes

Even though meiotic recombination genes exhibit high sequence conservation across eukaryotes, our analysis of eleven recombination genes in maize and teosinte identified a variety of selection patterns. Not surprisingly, using different tests to detect selection signatures produced somewhat different results. Frequency spectrum-based tests indicated departure from natural evolution patterns in seven genes in maize (*Dmc1, Mlh1*, *Mre11B*, *Msh4*, *Rad51A2*, *Sgs1*, and *Spo11-2*), although we only found evidence in three of them (*Mlh1*, *Mre11B*, and *Rad51A2*) that this departure was caused by adaptive evolution rather than demographics. With the LRT approach, we discovered positive selection patterns in *Mlh1*, *Mre11B*, *Mus81*, *Rad51A2*, and *Spo11-2* and in *Mre11A* and *Spo11-1* in teosinte. Overall, these data showed that a large fraction of recombination genes have been subject to non-neutral evolution.

Analyses of genome-wide diversity patterns in maize have shown that a relatively small number of maize genes, about 2 – 4 %, have experienced extremely strong selective sweeps during maize domestication (Tenaillon et al. 2004; Wright et al. 2005; Tian, Stevens, and Buckler 2009). These sweeps led to very severe diversity losses at the affected loci and their targets were mostly genes controlling plant architecture and critical agronomic traits (Wang et al. 1999; Wright et al. 2005). The recombination genes examined in this study do not appear to be in the same category of selection targets and exhibit higher levels of polymorphism in maize than the domestication-related genes affecting the agronomic and plant architecture traits. Consequently, the selective pressures experienced by the recombination genes are

likely to be significantly lower than those experiences by the domestication genes. It is also quite possible that the selective pressure on the recombination genes was episodic and not uniform across all maize lineages. For example, in some situations increased meiotic recombination rates may have been favored as they would facilitate formation of new gene combinations. In other lineages, lower recombination rates may have helped preserve linkage blocks of advantageous alleles. The fact that QTLs for recombination frequencies are not widely shared among diverse maize inbreds (McMullen et al. 2009), although they are present in individual mapping populations (Esch et al. 2007), provides some credibility to this claim. Future analyses of larger sets of maize inbreds may help discern whether lineage-specific selection patterns in recombination genes indeed exists.

Although several proteins involved in the meiotic recombination pathway, including Mlh1, Mre11, Mus81, Rad51, and Sgs1, are also known to function in somatic DNA repair, we did not detect obvious differences between the evolution patterns of genes encoding these proteins compared to genes that encode proteins with exclusively meiotic functions, such as Dmc1, Msh4, or Spo11. We speculate that the meiotic function may be predominant in the dual-function genes. Mutants in most meiotic recombination genes in plants do not show somatic defects, unless artificially exposed to genotoxic stress (Li et al. 2004; Puizina et al. 2004; Berchowitz et al. 2007; Dion et al. 2007; Hartung, Suer, and Puchta 2007), suggesting that these genes are not absolutely required for somatic growth and development under many normal growth conditions.

We found evidence that several sequence polymorphisms in the maize and teosinte recombination genes were likely to have functional consequences, suggesting that they indeed may have provided fitness benefits and/or benefits to the breeder and/or farmer. A number of the polymorphic residues were located in known functional domains of the proteins, and in some of these cases we were able to speculate on the potential impact of the amino acid changes on the protein structure by analyzing the positions of the polymorphic residues within three-dimensional protein models. Such protein changes could, for example, be associated with increased recombination rates. They could also affect the distribution of crossovers across the genome or affect the frequency of ectopic recombination between repetitive or homeologous genome regions. We found genes exhibiting adaptive selection patterns at all major steps of the recombination pathway, which suggests that the recombination pathway's adapting abilities are not limited to changing the crossover rates.

Interestingly, we found that in teosintes there were fewer genes exhibiting adaptive selection patterns and these genes were different from those found in maize. This fact suggests that different patterns of adaptation in the recombination pathway are beneficial to wild teosintes growing in a fairly small region in Mexico than to maize inbreds cultivated in a much wider area around the globe. Increased meiotic recombination rates are predicted to be beneficial in situations of strong or changing selection pressures (Otto and Michalakis 1998; Saleem, Lamb, and Nevo 2001), such as those present during domestication. It is, therefore, conceivable that adaptive changes in the recombination pathway may be of more importance in maize, and that

they may have contributed to the successful domestication of maize and its expansion to new growing areas.

2.4.2 Duplicated recombination genes

Three of the recombination genes examined in this study, *Mre11*, *Rad51*, and *Spo11*, are present in the maize genome as duplicated copies. However, all three of these duplications predate the allopolyploidization event that took place in the direct ancestry of maize (Gaut and Doebley 1997). Two (in *Rad51* and *Mre11*) likely took place at the base of the grass lineage and one (in *Spo11*) in the ancestor of all extant eukaryotes. These data imply that duplicated copies of the recombination genes that must have been formed as a result of the maize-specific polyploidization event have been subsequently lost from the genome. This observation is consistent with previous predictions, based on analyses of gene duplications in Arabidopsis, that DNA metabolism genes are preferentially subject to gene loss following whole-genome duplication (Blanc and Wolfe 2004).

We observed substantial differences in the evolution patterns between the paralogs of *Mre11*, *Rad51A*, and *Spo11*. While *Mre11B*, *Rad51A2*, and *Spo11-2* exhibited signatures of adaptive selection, *Mre11A*, *Rad51A1*, and *Spo11-1* did not. We also found that each of the copies of the duplicated genes exhibited very different rates of sequence diversity in maize. Overall, these data suggest that the duplicated gene copies have acquired distinct functions, which is a process known to follow gene duplications (Adams and Wendel 2005). Although the functions of the two *Spo11* genes have not been studied in maize yet, in Arabidopsis both *Spo11-1* and *Spo11-2*

are required for proper progression of recombination, which is consistent with each of the genes having a distinct function. In contrast, the two *Rad51A* genes in maize were found to be completely redundant in mutant analyses (Li et al. 2007). Our data showing different selection patterns in *Rad51A1* and *Rad51A2* suggests that differences in the functions of the two genes exist after all. The functions of the two *Mre11* genes in maize have not been studied, although the presence of a premature stop codon in *Mre11B* in one of the *Z. mays* ssp. *parviglumis* accessions suggests that *Mre11A* may substitute for *Mre11B*, which implies that the two genes are at least to some degree functionally redundant. Collectively, the results of our study suggest that evolutionary analyses are useful in complementing conventional genetic analyses in functional studies of duplicated genes.

REFERENCES

Adams, K. L., and J. F. Wendel. 2005. Polyploidy and genome evolution in plants. Curr. Opin. Plant Biol. **8**:135-141.

Altun, C. 2007. Maize Mre11 DNA repair and recombination complex. PhD thesis. Purdue University. Pp. 127.

Anderson, J. A., W. D. Gilliland, and C. H. Langley. 2009. Molecular population genetics and evolution of Drosophila meiosis genes. Genetics **181**:177-185.

Anisimova, M., R. Nielsen, and Z. Yang. 2003. Effect of recombination on the accuracy of the likelihood method for detecting positive selection at amino acid sites. Genetics **164**:1229-1236.

Berchowitz, L. E., K. E. Francis, A. L. Bey, and G. P. Copenhaver. 2007. The role of AtMUS81 in interference-insensitive crossovers in A. thaliana. PLoS Genet. **3**:e132.

Biswas, S., and J. M. Akey. 2006. Genomic insights into positive selection. Trends Genet. **22**:437-446.

Blanc, G., and K. H. Wolfe. 2004. Functional divergence of duplicated genes formed by polyploidy during Arabidopsis evolution. Plant Cell **16**:1679-1691.

Borde, V. 2007. The multiple roles of the Mre11 complex for meiotic recombination. Chromosome Res. **15**:551-563.

Buis, J., Y. Wu, Y. Deng, J. Leddon, G. Westfield, M. Eckersdorff, J. M. Sekiguchi, S. Chang, and D. O. Ferguson. 2008. Mre11 nuclease activity has essential

roles in DNA repair and genomic stability distinct from ATM activation. Cell **135**:85-96.

Chen, C., W. Zhang, L. Timofejeva, Y. Gerardin, and H. Ma. 2005. The Arabidopsis ROCK-N-ROLLERS gene encodes a homolog of the yeast ATP-dependent DNA helicase MER3 and is required for normal meiotic crossover formation. Plant J. **43**:321-334.

Copenhaver, G. P. 2005. Plant genetics: when not to interfere. Curr. Biol. **15**:R290-291.

Couteau, F., F. Belzile, C. Horlow, O. Grandjean, D. Vezon, and M.-P. Doutriaux. 1999. Random chromosome segregation without meiotic arrest in both male and female meiocytes of a *dmc1* mutant of Arabidopsis. Plant Cell **11**:1623-1634.

Dion, E., L. Li, M. Jean, and F. Belzile. 2007. An Arabidopsis MLH1 mutant exhibits reproductive defects and reveals a dual role for this gene in mitotic recombination. Plant J. **51**:431-440.

Esch, E., J. M. Szymaniak, H. Yates, W. P. Pawlowski, and E. S. Buckler. 2007. Using crossover breakpoints in recombinant inbred lines to identify quantitative trait loci controlling the global recombination frequency. Genetics **177**:1851-1858.

Eyre-Walker, A., R. L. Gaut, H. Hilton, D. L. Feldman, and B. S. Gaut. 1998. Investigation of the bottleneck leading to the domestication of maize. Proc. Nat. Acad. Sci. USA **95**:4441-4446.

Franklin, A. E., J. McElver, I. Sunjevaric, R. Rothstein, B. Bowen, and W. Z. Cande. 1999. Three-dimensional microscopy of the Rad51 recombination protein during meiotic prophase. Plant Cell **11**:809-824.

Fu, Y. X., and W. H. Li. 1993. Statistical tests of neutrality of mutations. Genetics **133**:693-709.

Gaut, B. S., and J. F. Doebley. 1997. DNA sequence evidence for the segmental allotetraploid origin of maize. Proc. Natl. Acad. Sci. USA **94**:6809-6814.

Gaut, B. S., S. I. Wright, C. Rizzon, J. Dvorak, and L. K. Anderson. 2007. Recombination: an underappreciated factor in the evolution of plant genomes. Nat. Rev. Genet. **8**:77-84.

Gore, M. A., J. M. Chia, R. J. Elshire, Q. Sun, E. S. Ersoz, B. L. Hurwitz, J. A. Peiffer, M. D. McMullen, G. S. Grills, J. Ross-Ibarra, D. H. Ware, and E. S. Buckler. 2009. A first-generation haplotype map of maize. Science **326**:1115-1117.

Gornall, R. J. 1983. Recombination systems and plant domestication. Biol. J. Linn. Soc. **20**:375-383.

Grelon, M., D. Vezon, G. Gendrot, and G. Pelletier. 2001. AtSPO11-1 is necessary for efficient meiotic recombination in plants. EMBO J. **20**:589-600.

Hartung, F., and H. Puchta. 2000. Molecular characterisation of two paralogous SPO11 homologues in Arabidopsis thaliana. Nucleic Acids Res. **28**:1548-1554.

Hartung, F., S. Suer, and H. Puchta. 2007. Two closely related RecQ helicases have antagonistic roles in homologous recombination and DNA repair in Arabidopsis thaliana. Proc. Natl. Acad. Sci. USA **104**:18836-18841.

Higgins, J. D., S. J. Armstrong, F. C. Franklin, and G. H. Jones. 2004. The Arabidopsis MutS homolog AtMSH4 functions at an early step in recombination: evidence for two classes of recombination in Arabidopsis. Genes Dev. **18**:2557-2570.

Hudson, R. R. 2002. Generating samples under a Wright-Fisher neutral model. Bioinformatics **18**:337-338.

Hudson, R. R. 1990. Gene genealogies and the coalescent process. . Oxf. Surv. Evol. Biol. **7**:1-44.

Hudson, R. R., M. Kreitman, and M. Aguade. 1987. A test of neutral molecular evolution based on nucleotide data. Genetics **116**:153-159.

Huelsenbeck, J. P., and F. Ronquist. 2001. MRBAYES: Bayesian inference of phylogenetic trees. Bioinformatics **17**:754-755.

Hughes, A. L., and M. Nei. 1988. Pattern of nucleotide substitution at major histocompatibility complex class I loci reveals overdominant selection. Nature **335**:167-170.

Huson, D. H., and D. Bryant. 2006. Application of phylogenetic networks in evolutionary studies. Mol. Biol. Evol. **23**:254-267.

Innan, H., and Y. Kim. 2004. Pattern of polymorphism after strong artificial selection in a domestication event. Proc. Natl. Acad. Sci. USA **101**:10667-10672.

Jones, G. H., and F. C. Franklin. 2006. Meiotic crossing-over: obligation and interference. Cell **126**:246-248.

Keeney, S., C. N. Giroux, and N. Kleckner. 1997. Meiosis-specific DNA double-strand breaks are catalyzed by Spo11, a member of a widely conserved protein family. Cell **88**:375-384.

Kosakovsky Pond, S. L., D. Posada, M. B. Gravenor, C. H. Woelk, and S. D. Frost. 2006. GARD: a genetic algorithm for recombination detection. Bioinformatics **22**:3096-3098.

Larkin, M. A., G. Blackshields, N. P. Brown, R. Chenna, P. A. McGettigan, H. McWilliam, F. Valentin, I. M. Wallace, A. Wilm, R. Lopez, J. D. Thompson, T. J. Gibson, and D. G. Higgins. 2007. Clustal W and Clustal X version 2.0. Bioinformatics **23**:2947-2948.

Li, J., L. C. Harper, I. Golubovskaya, C. R. Wang, D. Weber, R. B. Meeley, J. McElver, B. Bowen, W. Z. Cande, and P. S. Schnable. 2007. Functional analysis of maize RAD51 in meiosis and double-strand break repair. Genetics **176**:1469-1482.

Li, W., C. Chen, U. Markmann-Mulisch, L. Timofejeva, E. Schmelzer, H. Ma, and B. Reiss. 2004. The Arabidopsis *AtRAD51* gene is dispensable for vegetative development but required for meiosis. Proc. Natl. Acad. Sci. USA **101**:10596-10601.

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

Liu, K., M. Goodman, S. Muse, J. S. Smith, E. Buckler, and J. Doebley. 2003. Genetic structure and diversity among maize inbred lines as inferred from DNA microsatellites. Genetics **165**:2117-2128.

Louis, E. J., and R. H. Borts. 2003. Meiotic recombination: too much of a good thing? Curr. Biol. **13**:R953-955.

Maddison, W. P., and D. R. Maddison. 2009. Mesquite: a modular system for evolutionary analysis. http://mesquiteproject.org.

Malik, S. B., M. A. Ramesh, A. M. Hulstrand, and J. M. Logsdon, Jr. 2007. Protist homologs of the meiotic Spo11 gene and topoisomerase VI reveal an evolutionary history of gene duplication and lineage-specific loss. Mol. Biol. Evol. **24**:2827-2841.

Markmann-Mulisch, U., M. Z. Hadi, K. Koepchen, J. C. Alonso, V. E. Russo, J. Schell, and B. Reiss. 2002. The organization of Physcomitrella patens RAD51 genes is unique among eukaryotic organisms. Proc. Natl. Acad. Sci. USA **99**:2959-2964.

Matsuoka, Y., Y. Vigouroux, M. M. Goodman, G. J. Sanchez, E. Buckler, and J. Doebley. 2002. A single domestication for maize shown by multilocus microsatellite genotyping. Proc. Natl. Acad. Sci. USA **99**:6080-6084.

McDonald, J. H., and M. Kreitman. 1991. Adaptive protein evolution at the Adh locus in Drosophila. Nature **351**:652-654.

McMullen, M. D., S. Kresovich, H. S. Villeda, P. Bradbury, H. Li, Q. Sun, S. Flint-Garcia, J. Thornsberry, C. Acharya, C. Bottoms, P. Brown, C. Browne, M. Eller, K. Guill, C. Harjes, D. Kroon, N. Lepak, S. E. Mitchell, B. Peterson, G.

Pressoir, S. Romero, M. Oropeza Rosas, S. Salvo, H. Yates, M. Hanson, E. Jones, S. Smith, J. C. Glaubitz, M. Goodman, D. Ware, J. B. Holland, and E. S. Buckler. 2009. Genetic properties of the maize nested association mapping population. Science **325**:737-740.

Mercier, R., S. Jolivet, D. Vezon, E. Huppe, L. Chelysheva, M. Giovanni, F. Nogue, M. P. Doutriaux, C. Horlow, M. Grelon, and C. Mezard. 2005. Two meiotic crossover classes cohabit in Arabidopsis: one is dependent on MER3,whereas the other one is not. Curr. Biol. **15**:692-701.

Mimida, N., H. Kitamoto, K. Osakabe, M. Nakashima, Y. Ito, W. D. Heyer, S. Toki, and H. Ichikawa. 2007. Two alternatively spliced transcripts generated from OsMUS81, a rice homolog of yeast MUS81, are up-regulated by DNA-damaging treatments. Plant Cell Physiol. **48**:648-654.

Muse, S. V., and B. S. Gaut. 1994. A likelihood approach for comparing synonymous and nonsynonymous nucleotide substitution rates, with application to the chloroplast genome. Mol. Biol. Evol. **11**:715-724.

Namsaraev, E. A., and P. Berg. 1998. Interaction of Rad51 with ATP and Mg2+ induces a conformational change in Rad51. Biochemistry **37**:11932-11939.

Neale, M. J., and S. Keeney. 2006. Clarifying the mechanics of DNA strand exchange in meiotic recombination. Nature **442**:153-158.

Nei, M., and W. H. Li. 1979. Mathematical model for studying genetic variation in terms of restriction endonucleases. Proc. Natl. Acad. Sci. USA **76**:5269-5273.

Nichols, M. D., K. DeAngelis, J. L. Keck, and J. M. Berger. 1999. Structure and function of an archaeal topoisomerase VI subunit with homology to the meiotic recombination factor Spo11. EMBO J. **18**:6177-6188.

Nielsen, R. 2005. Molecular signatures of natural selection. Annu. Rev. Genet. **39**:197-218.

Nielsen, R., and Z. Yang. 1998. Likelihood models for detecting positively selected amino acid sites and applications to the HIV-1 envelope gene. Genetics **148**:929-936.

Nylander, J. A. A. 2004. MrModeltest. Evolutionary Biology Centre, Uppsala University., Uppsala, Sweden.

Otto, S. P., and Y. Michalakis. 1998. The evolution of recombination in changing environments. Trends Ecol. Evol. **13**:145-151.

Page, R. D. M. 1996. TREEVIEW: An application to display phylogenetic trees on personal computers. Computer Applications Biosci. **12**:357-358.

Pawlowski, W. P., I. N. Golubovskaya, L. Timofejeva, R. B. Meeley, W. F. Sheridan, and W. Z. Cande. 2004. Coordination of meiotic recombination, pairing, and synapsis by PHS1. Science **303**:89-92.

Pawlowski, W. P., C. J. Wang, I. N. Golubovskaya, J. M. Szymaniak, L. Shi, O. Hamant, T. Zhu, L. Harper, W. F. Sheridan, and W. Z. Cande. 2009. Maize AMEIOTIC1 is essential for multiple early meiotic processes and likely required for the initiation of meiosis. Proc. Natl. Acad. Sci. USA **106**:3603-3608.

Piperno, D. R., A. J. Ranere, I. Holst, J. Iriarte, and R. Dickau. 2009. Starch grain and phytolith evidence for early ninth millennium B.P. maize from the Central Balsas River Valley, Mexico. Proc. Natl. Acad. Sci. USA **106**:5019-5024.

Posada, D., and K. A. Crandall. 2001. Intraspecific gene genealogies: trees grafting into networks. Trends Ecol. Evol. **16**:37-45.

Puizina, J., J. Siroky, P. Mokros, D. Schweizer, and K. Riha. 2004. Mre11 deficiency in Arabidopsis is associated with chromosomal instability in somatic cells and Spo11-dependent genome fragmentation during meiosis. Plant Cell **16**:1968-1978.

Rockmill, B., J. C. Fung, S. S. Branda, and G. S. Roeder. 2003. The Sgs1 helicase regulates chromosome synapsis and meiotic crossing over. Curr. Biol. **13**:1954-1962.

Ross-Ibarra, J. 2004. The evolution of recombination under domestication: A test of two hypotheses. Am. Nat. **163**:105-112.

Saleem, M., B. C. Lamb, and E. Nevo. 2001. Inherited differences in crossing over and gene conversion frequencies between wild strains of Sordaria fimicola from "Evolution Canyon". Genetics **159**:1573-1593.

Sanchez-Moran, E., K. Osman, J. D. Higgins, M. Pradillo, N. Cunado, G. H. Jones, and F. C. Franklin. 2008. ASY1 coordinates early events in the plant meiotic recombination pathway. Cytogenet. Genome Res. **120**:302-312.

Schnable, P. S.D. WareR. S. FultonJ. C. SteinF. WeiS. PasternakC. LiangJ. ZhangL. FultonT. A. GravesP. MinxA. D. ReilyL. CourtneyS. S. KruchowskiC. TomlinsonC. StrongK. DelehauntyC. FronickB. CourtneyS. M. RockE.

BelterF. DuK. KimR. M. AbbottM. CottonA. LevyP. MarchettoK. OchoaS. M. JacksonB. GillamW. ChenL. YanJ. HigginbothamM. CardenasJ. WaligorskiE. ApplebaumL. PhelpsJ. FalconeK. KanchiT. ThaneA. ScimoneN. ThaneJ. HenkeT. WangJ. RuppertN. ShahK. RotterJ. HodgesE. IngenthronM. CordesS. KohlbergJ. SgroB. DelgadoK. MeadA. ChinwallaS. LeonardK. CrouseK. ColluraD. KudrnaJ. CurrieR. HeA. AngelovaS. RajasekarT. MuellerR. LomeliG. ScaraA. KoK. DelaneyM. WissotskiG. LopezD. CamposM. BraidottiE. AshleyW. GolserH. KimS. LeeJ. LinZ. DujmicW. KimJ. TalagA. ZuccoloC. FanA. SebastianM. KramerL. SpiegelL. NascimentoT. ZutavernB. MillerC. AmbroiseS. MullerW. SpoonerA. NarechaniaL. RenS. WeiS. KumariB. FagaM. J. LevyL. McMahanP. Van BurenM. W. VaughnK. YingC. T. YehS. J. EmrichY. JiaA. KalyanaramanA. P. HsiaW. B. BarbazukR. S. BaucomT. P. BrutnellN. C. CarpitaC. ChaparroJ. M. ChiaJ. M. DeragonJ. C. EstillY. FuJ. A. JeddelohY. HanH. LeeP. LiD. R. LischS. LiuZ. LiuD. H. NagelM. C. McCannP. SanMiguelA. M. MyersD. NettletonJ. NguyenB. W. PenningL. PonnalaK. L. SchneiderD. C. SchwartzA. SharmaC. SoderlundN. M. SpringerQ. SunH. WangM. WatermanR. WestermanT. K. WolfgruberL. YangY. YuL. ZhangS. ZhouQ. ZhuJ. L. BennetzenR. K. DaweJ. JiangN. JiangG. G. PrestingS. R. WesslerS. AluruR. A. MartienssenS. W. CliftonW. R. McCombieR. A. Wing, and R. K. Wilson. 2009. The B73 maize genome: complexity, diversity, and dynamics. Science **326**:1112-1115.

Schommer, C., A. Beven, T. Lawrenson, P. Shaw, and R. Sablowski. 2003. AHP2 is required for bivalent formation and for segregation of homologous chromosomes in Arabidopsis meiosis. Plant J. **36**:1-11.

Simmons, M. P., and H. Ochoterena. 2000. Gaps as characters in sequence-based phylogenetic analyses. Syst. Biol. **49**:369-381.

Soria-Carrasco, V., G. Talavera, J. Igea, and J. Castresana. 2007. The K tree score: quantification of differences in the relative branch length and topology of phylogenetic trees. Bioinformatics **23**:2954-2956.

Stacey, N. J., T. Kuromori, Y. Azumi, G. Roberts, C. Breuer, T. Wada, A. Maxwell, K. Roberts, and K. Sugimoto-Shirasu. 2006. Arabidopsis SPO11-2 functions with SPO11-1 in meiotic recombination. Plant J. **48**:206-216.

Sterken, R., R. Kiekens, E. Coppens, I. Vercauteren, M. Zabeau, D. Inze, J. Flowers, and M. Vuylsteke. 2009. A population genomics study of the Arabidopsis core cell cycle genes shows the signature of natural selection. Plant Cell **21**:2987-2998.

Swofford, D. L. 2003. PAUP. Phylogenetic analysis using parsimony (and other methods). Sinauer Associates, Sunderland, MA.

Tajima, F. 1993. Simple methods for testing the molecular evolutionary clock hypothesis. Genetics **135**:599-607.

Tajima, F. 1989. Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. Genetics **123**:585-595.

Tamura, K., J. Dudley, M. Nei, and S. Kumar. 2007. MEGA4: Molecular Evolutionary Genetics Analysis (MEGA) software version 4.0. Mol. Biol. Evol. **24**:1596-1599.

Tenaillon, M. I., M. C. Sawkins, A. D. Long, R. L. Gaut, J. F. Doebley, and B. S. Gaut. 2001. Patterns of DNA sequence polymorphism along chromosome 1 of maize (*Zea mays ssp. mays* L.). Proc. Natl. Acad. Sci. USA **98**:9161-9166.

Tenaillon, M. I., J. U'Ren, O. Tenaillon, and B. S. Gaut. 2004. Selection versus demography: A multilocus investigation of the domestication process in maize. Mol. Biol. Evol. **21**:1214-1225.

Tian, F., N. M. Stevens, and E. S. t. Buckler. 2009. Tracking footprints of maize domestication and evidence for a massive selective sweep on chromosome 10. Proc. Natl. Acad. Sci. USA **106**:9979-9986.

Vermaak, D., S. Henikoff, and H. S. Malik. 2005. Positive selection drives the evolution of *rhino*, a member of the heterochromatin protein 1 family in *Drosophila*. PLoS Genet. **1**:96-108.

Villeneuve, A. M., and K. J. Hillers. 2001. Whence meiosis? Cell **106**:647-750.

Wang, R. L., A. Stec, J. Hey, L. Lukens, and J. Doebley. 1999. The limits of selection during maize domestication. Nature **398**:236-239.

Waterworth, W. M., C. Altun, S. J. Armstrong, N. Roberts, P. J. Dean, K. Young, C. F. Weil, C. M. Bray, and C. E. West. 2007. NBS1 is involved in DNA repair and plays a synergistic role with ATM in mediating meiotic homologous recombination in plants. Plant J. **52**:41-52.

Watterson, G. A. 1975. On the number of segregating sites in genetical models without recombination. Theor. Popul. Biol. **7**:256-276.

Wright, S. I., I. V. Bi, S. G. Schroeder, M. Yamasaki, J. F. Doebley, M. D. McMullen, and B. S. Gaut. 2005. The effects of artificial selection on the maize genome. Science **308**:1310-1314.

Yamasaki, M., M. I. Tenaillon, I. V. Bi, S. G. Schroeder, H. Sanchez-Villeda, J. F. Doebley, B. S. Gaut, and M. D. McMullen. 2005. A large-scale screen for artificial selection in maize identifies candidate agronomic loci for domestication and crop improvement. Plant Cell **17**:2859-2872.

Yang, Z. 2007. PAML 4: phylogenetic analysis by maximum likelihood. Mol. Biol. Evol. **24**:1586-1591.

Yang, Z., W. S. Wong, and R. Nielsen. 2005. Bayes empirical bayes inference of amino acid sites under positive selection. Mol. Biol. Evol. **22**:1107-1118.

Zhai, W., R. Nielsen, and M. Slatkin. 2009. An investigation of the statistical power of neutrality tests based on comparative and population genetic data. Mol. Biol. Evol. **26**:273-283.

CHAPTER 3

ANALYSIS OF VARIATION IN CROSSOVER FREQUENCY IN MAIZE

3.1 Introduction

Meiotic recombination is one of the most important elements in life cycles of all sexually reproducing eukaryotes. Meiosis consists of two nuclear divisions: meiosis I and, meiosis II, which result in halving the chromosome number in each of the daughter cells. Accurate segregation of chromosomes is ensured by formation of crossovers between homologous chromosomes during prophase I. Crossovers are the sites of physical exchanges of segments between homologous chromosomes and are formed from a subset of double strand breaks. Double strand breaks are produced by topoisomerase-like protein Spo11, as a first step in the recombination pathway during meiosis (*1*). Not all DSBs are repaired as crossovers (*2*). In wild-type maize, nearly 500 DSBs are formed (*3*) and out of these about 20 are repaired as crossovers. DSBs that are not repaired as crossovers become noncrossovers or 'gene conversion' events. Bivalents that fail to form crossovers fail to segregate properly resulting in aneuploidy. Thus, crossover formation has to be under a very stringent control so that a bivalent formed during prophase I is held together is held together until metaphase I to ensure proper chromosome segregation (*4-6*).

It has been well documented that the number and distribution of crossovers during meiosis is are not random and are subject to stringent control (*7*). In order to ensure proper segregation during anaphase I, each bivalent requires at least one obligate crossover (*8*). Most species have two crossovers per bivalent (one per

chromosome arm), while more than two crossovers are rare. Cytological studies in humans and mouse spermatocytes revealed that most bivalents form only one or two crossovers but differences in numbers exist between sexes and chromosomes (*9-13*). A study on crossover numbers in human oocytes, found nearly three crossovers per bivalent (*14*). C. elegans mostly exhibits one crossover per chromosome (*15*). Studies in genus *Caledia* showed that the number of crossovers/nucleus is quite consistent within species (*16*).

Crossover formation is controlled in such a way that at least one crossover forms per bivalent but multiple crossovers are spaced far from each other. Spacing of multiple crossovers is affected by the phenomenon of crossover interference. Interference was first observed in *Drosophila* (*17*). Crossover interference reduces the probability of a second crossover forming in vicinity of a crossover so that distance between crossovers is larger than the expected distance if the two crossovers were to form independent of each other (*8, 17*). It has been speculated that this phenomenon is regulated by mechanical stress experienced due to expansion of chromatin. The hypothesis proposes that crossovers form so that this stress could be relieved (*6*). Another crossover is not formed in the vicinity because the mechanical stress is not enough to generate another break (*6*). Another model known as a 'counting model' has also been proposed to explain interference (*18*). According to this model, there needs to be a set number of non-crossovers before another neighboring crossover can form (*18*). Data from many species such as budding yeast, mouse, tomato, Arabidopsis and maize show that a small fraction of crossovers are independent of crossover interference (*18-24*). Accordingly, COs are categorized into class I, or

interfering crossovers, and class II, or non-interfering crossovers. In most species, including plants, the majority of crossovers are class I. These crossovers are formed by a group of meiosis-specific proteins called ZMM (*21, 25*). Non- interfering crossovers, are formed by a different set of proteins, including Mus81 and Mms4 (*26, 27*). In yeast and tomato, nearly 30% of all crossovers are class II, while in Arabidopsis, 15% of all crossovers belong to this class (*7*).

Crossover formation is also regulated by a phenomenon of "crossover homeostasis". It was shown in budding yeast that crossover numbers vary little with varying DSB numbers, so that even when there are fewer DSBs, the same number of crossovers per bivalent is formed at the expense of non-crossovers (*28*).

Crossovers can be visualized on a genome-wide and chromosome level through cytogenetic analysis by observing chiasmata. Chiasmata are the sites where homologous chromosomes are physically connected to each other during during the diplotene and diakinesis stages of prophase I. They are best visible at diakinesis, Chiasmata mark the sites of crossovers, which has been shown through BrdU labeling experiments (*29-31*). Thus, they have been long regarded as proxies for the number of crossovers. For the present study, I analyzed crossover numbers in inbreds and hybrids of maize. Maize is an excellent system for this type of analysis for several reasons:

A. In maize, meiosis is highly synchronized within anthers and flowers. Meiotic stages can be easily identified and because of the synchronization, one can

harvest a lot of material at a given stage, making maize as an excellent material to perform cytology.

B. Maize chromosomes are large and thus are easily observed under microscope and are excellent for cytological studies.

C. Maize harbors an exceptionally large amount of within-species genetic variation (*32*).

Besides chiasmata, there is another type of cytological structures known as recombination nodules that have been found to be associated with crossovers. Recombination nodules (RNs) are formed in synaptonemal complex during meiotic prophase I. Early RNs represent the sites of DSB repair, while late RNs represent sites of crossing-over. In several species it was shown that late RNs exhibit similar patterns of localization as crossovers (*33*). A 1:1 correspondence between the late RNs number and the chiasmata number has been found in lily and tomato (*34, 35*).

Genome-wide recombination rates are likely to be affected by numerous cis- and trans-acting factors. It has been documented that deletion on one chromosome arm affects the number of chiasmata on the other arm. A recent study of chiasma distribution in a wheat–rye addition line containing rye chromosome 5 in the hexaploid wheat background showed that a loss of about 70% of the distal portion of the long arm arm of wheat chromosome 5 resulted in more COs on the short arm (*36*).

Presence of supernumerary B chromosomes boosts genome-wide recombination. These chromosomes are mainly heterochromatic and are not essential for life of the organism (*37*). B chromosomes have been found in over 200 species of

plants and animals (*37*). The first evidence that B chromosomes can raise chiasma frequencies in A (normal) chromosomes came from studies in grasshoppers (*38*). Subsequently, there were many other studies in several species, including maize, rye, and lily, which further confirmed this hypothesis (*37, 39, 40*). It was also shown that B chromosomes cause redistribution of chiasma from distal to more proximal positions on chromosomes (*40*). In contrast to these studies, in wheat and wheat-aegilops hybrids, it has been found that B chromosomes cause a significant reduction in chiasma frequency (*41, 42*).

An increase in recombination frequency associated with the presence of B chromosomes in maize has also been shown through genetic means. A study of crossover frequency between four markers on chromosome 3 showed increasing recombination rates with increasing numbers of B chromosomes (*43*).

Several studies in plants have suggested that heterozygocity affects meiotic crossover rates. It has been documented in pearl millet that hybrids had significantly higher numbers of chiasma than their inbred parents (*44*). It was also reported in the same study that newly formed hybrids had significantly higher chiasma numbers than established hybrids (*44*).

A study in maize reported that the number of COs is affected by allelic variation at the locus between the parents (*45*). In this study, recombination was analyzed at the *bronze* locus between a variety of heteroalleles, which differed from each other in the level of plolymorpism ranging from point mutations to large indels (*45*). It was shown that polymorphic heteroalleles resulted in more crossovers than more similar parental alleles (*45*).

It has also been documented that recombination frequency is higher in polyploids. In recent studies, it was shown that Arabidopsis tetraploids as well as allotriploids and tetraploids in Brassica have significantly higher numbers of chiasmata than diploids (*46, 47*).

Although the mechanistic aspects of meiotic recombination in plants especially in Arabidopsis, maize, and rice, have received a lot of attention in past few years (*7, 48*), very little is known about natural variation in crossover frequencies within a species (*49, 50 2008*). In this study, I analyzed crossover frequencies in thirteen diverse inbreds of maize with the aim of identifying lines that show extreme variation in chiasma numbers. I found that maize inbreds differed significantly from each other with respect to chiasma numbers.

3.2 Materials and Methods

3.2.1 Plant material

Thirteen maize inbreds, B73, B97, CML69, CML103, CML228, CML277, CML322, Hp301, Ki11, Mo17, Mo18w, M162w, and Oh43 were selected from a collection of NAM population parents. NAM population parent inbreds represent a diverse collection of maize germplasm and thus offers the best resource to study genetic variation in this species (*51-53*). Plants were grown in 1:1 Promix :calcined clay soil mix in growth chamber facility in Weill Hall, Cornell University with 31 C day temp and 22 C night temp with 12 hr light. The light intensity was maintained at 800 µmol/m2s. Meiocytes were harvested when plants were about 6-7 weeks old. This time was variable due to different inbred backgrounds.

3.2.2 Flower fixation

To determine the meiosis stage, anthers were stained using the acetocarmine staining and viewed under a light microscope. Entire tassels were fixed in Carnoy's solution containing 3:1 ethanol:glacial acetic acid for 24 hours and then stored in 70% ethanol at -20 C.

3.2.3 Chromosome spreading

Chromosomes at the diakinesis stage of meiotic prophase I were visualized using the well-established chromosome-spreading technique (*54*) with some minor modifications. Fixed meiocytes were washed in citric acid buffer for 10 minutes and incubated in an enzyme mixture consisting of 2% Onazuka R10 cellulase, 1% mecerozyme, 1% cytohelocase, 4 mM citric acid, and 6 mM sodium citrate, pH4.8 at room temperature for 60 minutes. Enzyme-digested anthers were transferred to a slide, macerated with a needle in a small drop of 95:5 (v/v) acetic acid : methanol solution. The solution containing meiocytes was dropped from about 30 cm height on a glass slide slides for effective spreading. After subsequent air-drying, chromosomes were cross-linked to the slides using UV light (1200 J). Finally, slides were stained with 500 μL of DAPI (4,6-diamidino-2-phenylindole) solution (1 μg mL^{-1}) and washed with 1X PBS for 10 minutes, three times. The slides were then mounted with DABCO and sealed with nail polish.

3.2.4 Fluorescence microscopy

A DeltaVision imaging workstation (Applied Precision, Issaquah, WA) was used to generate 3-dimensional stacks of images that were taken every 0.15 μm across the whole cell nucleus. The 3-D image stacks were deconvolved and analyzed using the SoftWoRx software (Applied Precision, Issaquah, WA). In most cases, I was able to clearly identify the number of chiasmata for a given bivalent. In some ambiguous cases where it was not clear, 3D reconstructions of the imaged nuclei were generated. The 3-D reconstructions allowed viewing chromosomes at all possible angles and made it easier to evaluate the different shapes of bivalents at diakinesis.

3.3 Results and Discussion

3.3.1 Chiasmata frequency analysis in maize inbreds

To survey natural variation in genome-wide recombination rates in maize, I examined chiasmata during the diakinesis stage of meiosis I in a set of thirteen maize inbred lines. Chiasmata were counted in chromosome spreads. To enhance our ability to correctly identify and detect all chiasmata, I used 3D image reconstruction, which allowed me to clearly see the morphology of bivalents (Figure 3.1). Using this method, it was easy to clearly see chiasmata by analyzing bivalent configurations. There are two very distinct configurations of bivalents during diakinesis: rods and rings (Figure 3.1). Rods have only one chiasma per bivalent while rings have two chiasmata, one at each end (Figure 3.1).

There are two main challenges in determining the accurate number of chiasmata per chromosome, which could lead to over or under estimation of the total

number: (1) chromosome twists can be mistaken for chiasmata and (2) nearby chiasmata cannot be distinguished from each other. The 3D reconstruction method allowed me to differentiate between chromosome twists and *bona fide* chiasma because bivalents could be viewed from all different angles (Figure 3.1c, 3.1d). Chiasmata present near each other make it harder to distinguish multiple from one chiasma (Figure 3.1e). In such cases, to avoid false negatives, the size of the chiasma was taken into account. Typically, a chiasma is 1-2 um in length so if two chiasmata were next to each other, they can be distinguished based on size (Figure 3.1e). These features made it easier to count multiple chiasmata on a bivalent.

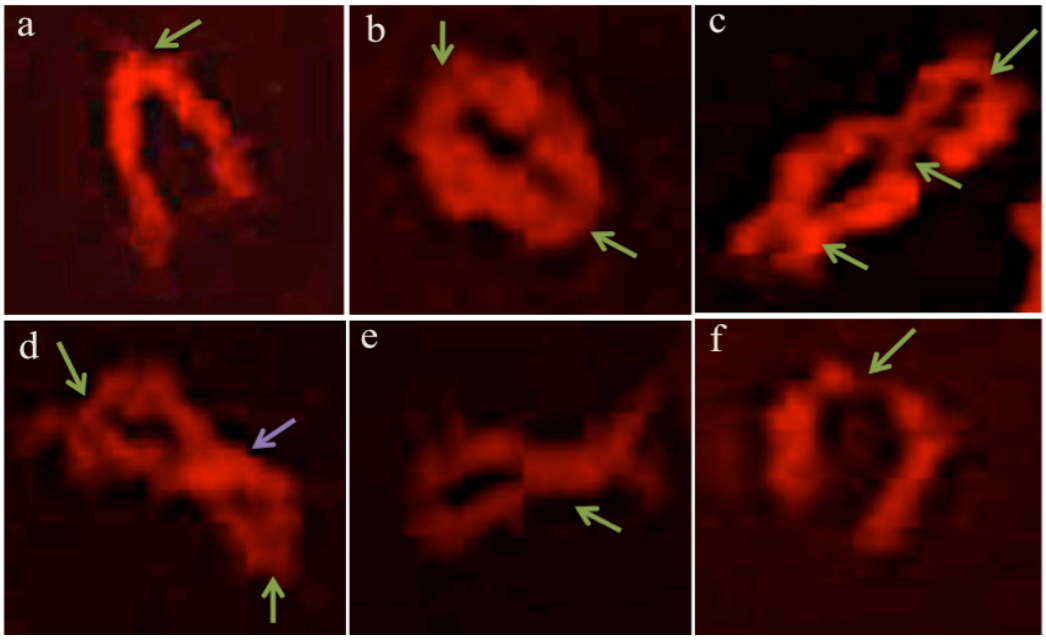
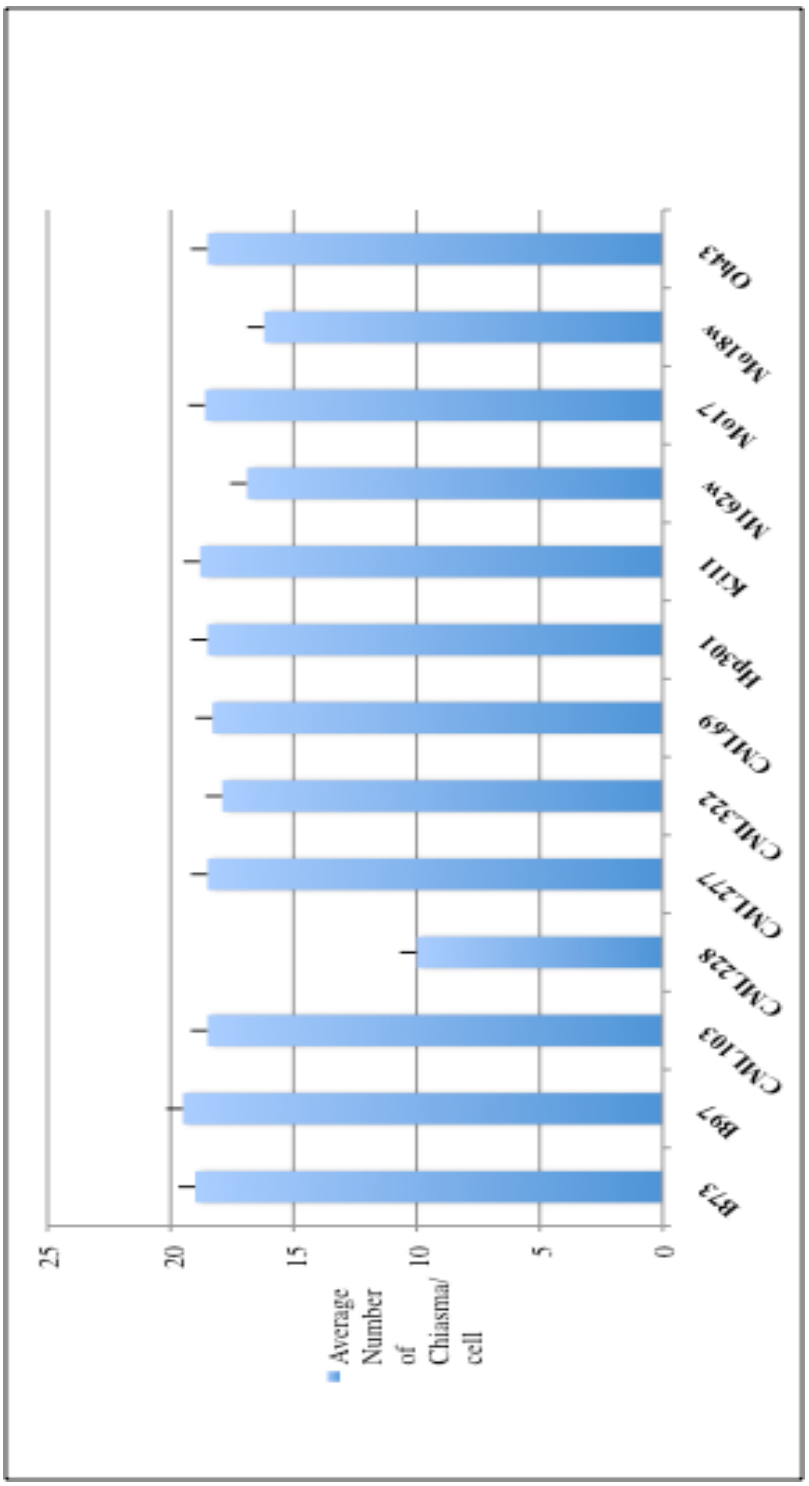


Figure 3.1: Combining chromosome spreading with 3D reconstruction of images allows us to precisely determine the number of chiasma, chiasmata are pointed in the images by arrows: Top panel (a-c) shows typical chromosome morphology for bivalents that one, two or three chiasmata; the bottom panel (d-f) shows ambiguous bivalent morphologies that appear as three, one and no chiasma; a) rod bivalent that has one chiasma at one end, b) a typical ring bivalent, with one chiasma at each end, c) a bivalent showing three chiasmata: two terminal and one interstitial, d) A bivalent appears showing two terminal chiasmata and one twist (pointed by purple arrow) appearing as chiasma e) a bivalent which has two chiasma closely spaced to each other f) a bivalent appear as two univalents.

Figure 3.2: Average number of chiasmata per cell in thirteen maize inbreds



Among the thirteen maize inbred lines that I analyzed, the chiasmata numbers varied from 10.2 to 19.5. I observed the highest average number of chiasmata per cell in the B97 inbred (19.5/cell) and the lowest in CML228 (10.2/cell). As CML228 was clearly an 'outlier' in this analysis (Table 3.1, Figure 3.2), to make sure that the low chaismata counts

Table 3.1: Variation in average number of chiasma per cell in maize inbreds:

S. No.	Inbred	Number of cells	Average Number of Chiasma/cell	Standard Error
1	B73	25	19.0	0.2
2	B97	40	19.5	0.113
3	CML103	46	18.5	0.142
4	CML228	44	10.2	0.241
5	CML277	45	18.5	0.174
6	CML322	42	17.9	0.256
7	CML69	40	18.3	0.136
8	Hp301	25	18.5	0.244
9	Ki11	45	18.8	0.22
10	M162w	50	16.9	0.201
11	Mo17	46	18.6	0.131
12	Mo18w	65	16.2	0.334
13	Oh43	45	18.5	0.17

With the exception of CML228, the differences in chiasmata numbers among the inbreds were rather small. To determine whether these differences were statistically significant, I conducted a one-way analysis of variance. The results of this analysis showed that in the vast majority of comparisons, the differences between inbreds were statistically significant (Table 3.2a). These data also showed that within-inbred variation was very small in most cases. Since CML228 exhibited a higher degree of

variation than the other inbreds, I performed another analysis that specifically excluded CML228. The results from this analysis are listed in Table 3.2b. The differences among inbreds in the later analysis were also found to be significant.

Table 3.2a: Analysis of variance test for variation in average number of chiasma per cell in maize inbreds (p=0.001)

Sources of Variation	Sum of Squares	d.f.	Mean Sum of squares	F-value
Between groups	4070.0	12	339.2	217.2
Error	821.3	526	1.561	
Total	4891	538		

Table 3.2b: Analysis of variance test for variation in average number of chiasma per cell in maize inbreds except CML228 (p=0.001)

Sources of Variation	Sum of Squares	d.f.	Mean Sum of squares	F-value
Between groups	459.8	11	41.8	27.57
Error	683.9	451	1.516	
Total	1144.0	462		

Since CML228 exhibited a much lower number of chiasmata per bivalent then the other inbreds, I analyzed the overall meiosis progression in CML228 plants. CML228 meiocytes seemed to progress through meiosis normally. However, I found that synchronization of meiosis progression among flowers on a tassel was disturbed in this inbred. In maize, the oldest meiotic stages are found in anthers located in the center of the tassel, followed by younger stages on the top and bottom of the tassel. In CML228, I observed that meiotic stages were not synchronized in this way. Instead,

the stages were intertwined and presented a complex arrangement in which flowers with meiocytes at earlier stages of meiosis such as zygotene and pachytene were intermingled with flowers tat much later stages of meiosis like tetrads.

Surveys of chismata numbers have been previously undertaken in Arabidopsis and mouse. In Arabidopsis, eight diverse accessions were compared shown to differ significantly from each other with respect to chiasma numbers ranging from, 7.9 to 9.36 chiasma per cell (*49*). Chiasma numbers in these accessions were estimated through analyzing bivalent configuration (rods vs. rings) at metaphase I.

Arabidopsis chromosomes are smaller than maize chromosomes and and the cytology is challenging. Furthermore, chromosomes are more condensed at metaphase than at diakinesis, making it harder to visualize the actual chiasma number. It is possible that chiasma numbers observed in the Arabidopsis study were under or over estimated due to these factors. The study in mouse, which compared crossover numbers in eight mouse strains revealed that the average crossover number per cell varied from 22.6 to 23.9 (*9*). Although, the variation in the number of crossovers between strains was not statistically significant, the study reported a significant correlation between chromosome length and crossover (performed on one mouse strain) (*9*).

In addition to recording the number of chiasmata, I analyzed my data set to see what percentage of bivalents had one, two, or three chiasma. Inbreds differed from each other significantly in this aspect. CMl228 had 99% bivalents with a single chiasma/cell while B97 had only about 9% bivalents with one chiasma per cell; the highest and lowest values among all inbreds analyzed. Most inbreds had between 10

to 20% bivalents with one chiasma per cell with the exception of M162w (31% bivalents with one chiasma) and Mo18w (41% bivalents with one chiasma). This was expected as M162w and Mo18w also had lower average number of chiasma per cell (16.8 and 16) compared to other inbreds.

Most inbreds analyzed in this study had between 80 to 88% bivalents with two chiasmata per cell. The exceptions were M162w, Mo18w, and CML228, which showed relatively low overall chiasma numbers and exhibited high fractions of bivalents with only a single chiasma. Another exception was CML103. In this inbred, the average chiasma number per cell was higher than in M162w, Mo18w, or CML228 (18.5 chiasmata/meiocyte) but a higher percentage of bivalents per cell exhibited three chiasmata (7.5 %) compared to the other inbreds.

There are several factors that can affect the number of crossovers within species. These factors are genetic (*49, 55, 56*), environmental (*57*), developmental (*57*), and genomic (*45, 58, 59*). Environmental factors, such as temperature, have been shown to affect crossover frequency in many different species (*60*). However, different species exhibit different effects of temperature on CO frequencies, ranging from positive correlation to negative correlation (*60*).

Developmental effects on CO frequencies have been reported in Arabidopsis, where it was shown that CO frequencies varied depending on the position of the flower (*57*). Primary flowers, which are derived from the apical meristem, had lower CO frequency than secondary flowers (flowers on braches of primary flowers) and tertiary (flowers on braches of secondary flowers) flowers, which are derived from lateral meristems indicating that developmental position affects recombination

frequency (*57*). No correlation between plant age and CO frequency was observed in this study (*57*).

Genetics controllers of CO frequencies can be trans-acting (genes that have a regulatory effect on crossover frequency) or cis-acting (structural elements of the chromosomes). Recombination frequency variation due to differences in genetic background has been documented in several plant species; Arabidopsis, maize and Petunia hybrids (*49, 55, 56*).

The study in Arabidopsis and the present study in maize revealed that although the variation of chiasma numbers among lines was statistically significant, the overall extent of the variation was rather limited. The requirement for an obligatory chiasma in each bivalent defines the minimum number of CO events per cell. These data suggest that, at least in plants, there is also a mechanism that limits the maximum number of crossovers. A high number of CO events could, for example, introduce genomic instability. Potential consequences of genomic instability could outweigh the benefits of additional recombination and, thus, avoiding high recombination frequencies may be desirable.

3.3.2 Chiasmata frequency analysis in hybrids

In addition to examining chiasma frequencies in homozygous inbreds, I have also examined chiasma frequencies in three maize hybrids generated from NAM parents. I chose B73 x CML228, B73 x Mo17, and B73 x Mo18w for this analysis. I chose B73 x Mo17 because B73 and Mo17 belong to different heterotic groups and, the B73 x Mo17 inbred exhibits strong heterosis (*61*).

B73 x CML228 was chosen because in a recent study on recombination variation in recombinant inbred lines of the NAM population, it was shown that the B73x CML228 cross exhibited the highest recombination frequency out of 25 progenies analyzed in this study. The B73 x Mo18w progeny exhibited lowest recombination frequency, and I also selected this hybrid for further analysis (*53*).

I found that B73 x Mo17 hybrid had the average chiasmata number per cell that was the highest among the three hybrids (19.5 per meiocyte). B73 x Mo18w exhibited the lowest number (16.2 chiasmata per cell) (Figure 3.3). The B73 x CML228 hybrid exhibited 17.8 chiasmata per cell on average. The B73 x CML228 and B73 x Mo17 hybrids exhibited higher crossover frequency than the average of their parental means while the B73 x Mo18w hybrid exhibited crossover frequency, which was similar to Mo18w but lower than the parental mean (Figure 3.3).

Previously, recombination frequencies have been examined using genetic mapping in 25 recombinant inbred line families (S5 generation) (*53*). This study reported significant variation in hybrid families ranging from –104.3 cM for B73 × Mo18W to +269.4 cM for B73 × CML228. Our results are not in agreement with these data and this could be due to several reasons: First, recombinant inbred lines used in the McMullen et al, study have gone through several rounds of meiosis; while in our study, the crossover numbers were estimated based on a single meiosis. Second, in the McMullen study, both male and female meiotic events were accounted for, while for our study, only male meiosis was taken into account.

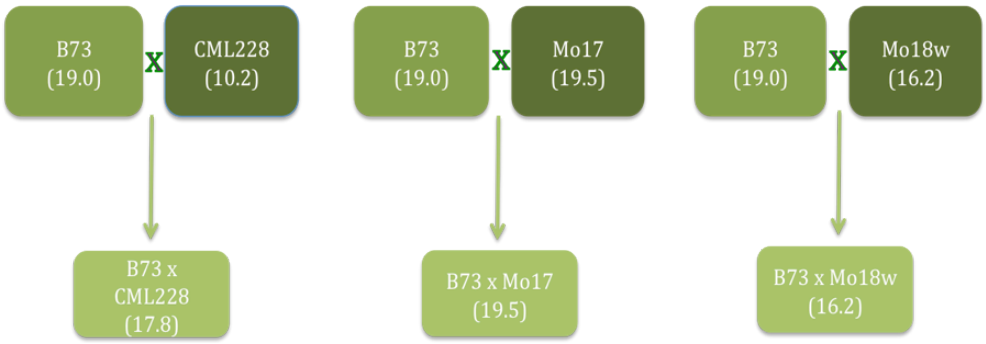


Figure 3.3: Average number of chiasmata per cell in hybrids and parents

Male vs. female recombination frequency differences were first observed in *Drosophila* and silk worm, where females were found to have more recombination frequency than males (*62, 63*). Similar observation of higher recombination in females than males was made in humans (*64, 65*) and mice [(*66*)]. In Arabidopsis, it has been shown that recombination frequency in males is generally higher than females (*66*). Similarly, male recombination frequency has been reported to be higher in maize (*67, 68*). The sex-related differences in recombination frequency are not uniform across entire genome. For example, several genetic intervals in maize are longer in female (*68*). In humans, similarly some regions were same in genetic size although overall female genetic maps are longer than male (*65*).

3.4 Conclusions

In this study on chiasma variation in thirteen inbreds of maize, I documented that maize inbreds significantly differ from each other with respect to average chiasma numbers per cell. This is the first documentation of significant within species variation in crossover numbers in a complex genome. The variation observed between different inbreds suggests the existence of genetic factors that differ from each other in these inbred lines, leading to differences in recombination frequency. These could be genes or other structural or regulatory elements on chromosomes This study will form the basis of further studies to understand the genetic basis of within-species variations in recombination frequencies.

REFERENCES

1. S. Keeney, Mechanism and control of meiotic recombination initiation. *Current topics in developmental biology* **52**, 1 (2001).
2. P. B. Moens *et al.*, The time course and chromosomal localization of recombination-related proteins at meiosis in the mouse are compatible with models that can resolve the early DNA–DNA interactions without reciprocal recombination. *J. Cell Sci.* **115**, 1611 (2002).
3. W. P. Pawlowski *et al.*, Coordination of meiotic recombination, pairing, and synapsis by PHS1. *Science* **303**, 89 (Jan 2, 2004).
4. D. K. Bishop, D. Zickler, Early decision; meiotic crossover interference prior to stable strand exchange and synapsis. *Cell* **117**, 9 (Apr 2, 2004).
5. K. J. Hillers, Crossover interference. *Current biology : CB* **14**, R1036 (Dec 29, 2004).
6. G. V. Borner, N. Kleckner, N. Hunter, Crossover/noncrossover differentiation, synaptonemal complex formation, and regulatory surveillance at the leptotene/zygotene transition of meiosis. *Cell* **117**, 29 (Apr 2, 2004).
7. C. Mezard, J. Vignard, J. Drouaud, R. Mercier, The road to crossovers: plants have their say. *Trends in genetics : TIG* **23**, 91 (Feb, 2007).
8. G. H. Jones, The control of chiasma distribution. *Symposia of the Society for Experimental Biology* **38**, 293 (1984).

9. L. K. Anderson, A. Reeves, L. M. Webb, T. Ashley, Distribution of crossing over on mouse synaptonemal complexes using immunofluorescent localization of MLH1 protein. *Genetics* **151**, 1569 (Apr, 1999).

10. A. L. Barlow, M. A. Hulten, Crossing over analysis at pachytene in man. *European journal of human genetics : EJHG* **6**, 350 (Jul-Aug, 1998).

11. L. Froenicke, L. K. Anderson, J. Wienberg, T. Ashley, Male mouse recombination maps for each autosome identified by chromosome painting. *American journal of human genetics* **71**, 1353 (Dec, 2002).

12. D. A. Laurie, R. W. Palmer, M. A. Hulten, Chiasma derived genetic lengths and recombination fractions: a 46, XY, t(9; 10) (p22; q24) reciprocal translocation. *Annals of human genetics* **49**, 135 (May, 1985).

13. N. M. Lawrie, C. Tease, M. A. Hulten, Chiasma frequency, distribution and interference maps of mouse autosomes. *Chromosoma* **104**, 308 (Dec, 1995).

14. C. Tease, M. A. Hulten, Inter-sex variation in synaptonemal complex lengths largely determine the different recombination rates in male and female germ cells. *Cytogenetic and genome research* **107**, 208 (2004).

15. W. B. Wood, The nematode Caenorhabditis elegans. *3. Wood,W.B.(1988).ThenematodeCaenorhabditiselegans(Cold Spring Harbor, NY: Cold Spring Harbor Laboratory Press*, (1988).

16. D. D. Shaw, G. R. Knowles, Comparative chiasma analysis using a computerised optical digitiser. *Chromosoma* **59**, 103 (Dec 16, 1976).

17. H. J. Muller, The mechanism of crossing-over. *Am. Nat.* **50**, 193 (1916).

18. F. W. Stahl *et al.*, Does crossover interference count in Saccharomyces cerevisiae? *Genetics* **168**, 35 (Sep, 2004).

19. N. M. Hollingsworth, S. J. Brill, The Mus81 solution to resolution: generating meiotic crossovers without Holliday junctions. *Genes & development* **18**, 117 (Jan 15, 2004).

20. H. Guillon, F. Baudat, C. Grey, R. M. Liskay, B. de Massy, Crossover and noncrossover pathways in mouse meiosis. *Molecular cell* **20**, 563 (Nov 23, 2005).

21. F. G. Lhuissier, H. H. Offenberg, P. E. Wittich, N. O. Vischer, C. Heyting, The mismatch repair protein MLH1 marks a subset of strongly interfering crossovers in tomato. *The Plant cell* **19**, 862 (Mar, 2007).

22. J. D. Higgins, S. J. Armstrong, F. C. Franklin, G. H. Jones, The Arabidopsis MutS homolog AtMSH4 functions at an early step in recombination: evidence for two classes of recombination in Arabidopsis. *Genes & development* **18**, 2557 (Oct 15, 2004).

23. R. Mercier *et al.*, Two meiotic crossover classes cohabit in Arabidopsis: one is dependent on MER3,whereas the other one is not. *Current biology : CB* **15**, 692 (Apr 26, 2005).

24. M. Falque, L. K. Anderson, S. M. Stack, F. Gauthier, O. C. Martin, Two types of meiotic crossovers coexist in maize. *The Plant cell* **21**, 3915 (Dec, 2009).

25. A. Lynn, R. Soucek, G. V. Borner, ZMM proteins during meiosis: crossover artists at work. *Chromosome research : an international journal on the*

molecular, supramolecular and evolutionary aspects of chromosome biology **15**, 591 (2007).

26. T. de los Santos *et al.*, The Mus81/Mms4 endonuclease acts independently of double-Holliday junction resolution to promote a distinct subset of crossovers during meiosis in budding yeast. *Genetics* **164**, 81 (May, 2003).

27. M. C. Whitby, Making crossovers during meiosis. *Biochemical Society transactions* **33**, 1451 (Dec, 2005).

28. E. Martini, R. L. Diaz, N. Hunter, S. Keeney, Crossover homeostasis in yeast meiosis. *Cell* **126**, 285 (Jul 28, 2006).

29. C. Tease, Cytological detection of crossing-over in BUdR substituted meiotic chromosomes using the fluorescent plus Giemsa technique. *Nature* **272**, 823 (Apr 27, 1978).

30. J. W. Allen, BrdU-dye characterization of late replication and meiotic recombination in Armenian hamster germ cells. *Chromosoma* **74**, 189 (1979).

31. A. Latos-Bielenska, W. Vogel, Frequency and distribution of chiasmata in Syrian hamster spermatocytes studied by the BrdU antibody technique. *Chromosoma* **99**, 267 (Aug, 1990).

32. E. S. Buckler, B. S. Gaut, M. D. McMullen, Molecular and functional diversity of maize. *Current opinion in plant biology* **9**, 172 (Apr, 2006).

33. S. M. Albini, G. H. Jones, Synaptonemal complex-associated centromeres and recombination nodules in plant meiocytes prepared by an improved surface-spreading technique. *Experimental cell research* **155**, 588 (Dec, 1984).

34. S. M. Stack, L. K. Anderson, Two-dimensional spreads of synaptonemal complexes from solanaceous plants. II. Synapsis in Lycopersicon esculentum (tomato). *Am. J. Bot.* **73**, 264 (1986a).

35. S. M. Stack, L. K. Anderson, Two-dimensional spreads of synaptonemal complexes from solanaceous plants. III. Recombination nodules and crossing over in Lycopersicon esculentum (tomato). *Chromosoma* **94**, 253 (1986b).

36. T. Naranjo, N. T. Valenzuela, E. Perera, Chiasma frequency is region specific and chromosome conformation dependent in a rye chromosome added to wheat. *Cytogenetic and genome research* **129**, 133 (Jul, 2010).

37. R. N. Jones, H. Rees, B Chromosomes. *Academic Press, London*, (1982).

38. J. F. Barker, Climatological Distribution of a Grasshopper Supernumerary Chromosome. *Evolution* **20**, 665 (1966).

39. G. Bell, A. Burt, B-chromosomes: germ-line parasites which induced changes in host recombination. *Parasitology* **100 Suppl**, S19 (1990).

40. E. J. Ward, The effect of accessory chromatin on chiasma distribution in maize. *Canadian journal of genetics and cytology. Journal canadien de genetique et de cytologie* **18**, 479 (Sep, 1976).

41. G. Simchen, Y. Zarchi, J. Hillel, Supernumerary chromosomes in the second outbreeding species of the wheat group. *Chromosoma* **33**, 63 (1971).

42. A. Vardi, G. A. Dover, The effect of B chromosomes on meiotic and pre-meiotic spindles and chromosome pairing in Triticum-Aegilops Hybrids. *Chromosoma* **38**, 367 (1972).

43. U. W. Ayonoadu, H. Rees, The regulation of mitosis by B-chromosomes in rye. *Experimental cell research* **52**, 284 (Sep, 1968).

44. H. K. Srivastava, H. S. Balyan, Increased chiasma frequency in some hybrids of pearl millet. *J. Hered* **68**, 338 (1977).

45. H. K. Dooner, Extensive interallelic polymorphisms drive meiotic recombination into a crossover pathway. *The Plant cell* **14**, 1173 (May, 2002).

46. M. Leflon *et al.*, Crossovers get a boost in Brassica allotriploid and allotetraploid hybrids. *The Plant cell* **22**, 2253 (Jul, 2010).

47. A. Pecinka, W. Fang, M. Rehmsmeier, A. A. Levy, O. Mittelsten Scheid, Polyploidization increases meiotic recombination frequency in Arabidopsis. *BMC biology* **9**, 24 (2011).

48. R. Mercier, M. Grelon, Meiosis in plants: ten years of gene discovery. *Cytogenetic and genome research* **120**, 281 (2008).

49. E. Sanchez-Moran, S. J. Armstrong, J. L. Santos, F. C. Franklin, G. H. Jones, Variation in chiasma frequency among eight accessions of Arabidopsis thaliana. *Genetics* **162**, 1415 (Nov, 2002).

50. L. K. Anderson *et al.*, High-resolution crossover maps for each bivalent of Zea mays using recombination nodules. *Genetics* **165**, 849 (Oct, 2003).

51. J. Yu, J. B. Holland, M. D. McMullen, E. S. Buckler, Genetic design and statistical power of nested association mapping in maize. *Genetics* **178**, 539 (Jan, 2008).

52. M. D. McMullen *et al.*, Genetic properties of the maize nested association mapping population. *Science* **325**, 737 (Aug 7, 2009).

53. M. A. Gore *et al.*, A first-generation haplotype map of maize. *Science* **326**, 1115 (Nov 20, 2009).

54. Jones G, H.-H. Jeds, Classical and Molecular Cytogenetics of Arabidopsis thaliana. *Oxford University Press, Oxford*, (1996).

55. Robbins T.P., Gerats A.G.M., Fiske H., J. R.A., Suppression of recombination in wide hybrids of petunia hybrids as revealed by genetic mapping of marker transgenes. *Theor. Appl. Genet.* **90**, 957 (1995).

56. E. Esch, J. M. Szymaniak, H. Yates, W. P. Pawlowski, E. S. Buckler, Using crossover breakpoints in recombinant inbred lines to identify quantitative trait loci controlling the global recombination frequency. *Genetics* **177**, 1851 (Nov, 2007).

57. K. E. Francis *et al.*, Pollen tetrad-based visual assay for meiotic recombination in Arabidopsis. *Proceedings of the National Academy of Sciences of the United States of America* **104**, 3913 (Mar 6, 2007).

58. J. Dvorak, M. C. Luo, Z. L. Yang, Restriction fragment length polymorphism and divergence in the genomic regions of high and low recombination in self-fertilizing and cross-fertilizing aegilops species. *Genetics* **148**, 423 (Jan, 1998).

59. S. Barth, A. E. Melchinger, B. Devezi-Savula, T. Lubberstedt, Influence of genetic background and heterozygosity on meiotic recombination in Arabidopsis thaliana. *Genome / National Research Council Canada = Genome / Conseil national de recherches Canada* **44**, 971 (Dec, 2001).

60. J. Y. Wilson, Temperature effect on chiasma frequency in the bluebell endymion nonscriptus. *Chromosoma* **10**, 337 (1959).

61. J. A. Birchler, D. L. Auger, N. C. Riddle, In search of the molecular basis of heterosis. *The Plant cell* **15**, 2236 (Oct, 2003).

62. T. H. Morgan, Complete linkage in the second chromosome of the male of Drosophila. . *Science* **36**, 719 (1912).

63. J. Haldane, Sex ratio and unisex sterility in hybrid animals. . *J Genet* **12**, 101 (1922).

64. K. W. Broman, J. C. Murray, V. C. Sheffield, R. L. White, J. L. Weber, Comprehensive human genetic maps: individual and sex-specific variation in recombination. *American journal of human genetics* **63**, 861 (Sep, 1998).

65. H. Donis-Keller *et al.*, A genetic linkage map of the human genome. *Cell* **51**, 319 (Oct 23, 1987).

66. I. Y. Vizir, A. B. Korol, Sex differences in recom-bination frequency in Arabidopsis. . *Heredity* **65**, 379 (1990).

67. M. M. RHOADES, Differentratesofcrossingoverinthe male andfemalegametesofmaize. . *J. Am. SOC.Agron.* **33:**, 603 (1941).

68. D. S. Robertson, Different frequency in the recovery of crossover products from male and female gametesof plants hypoploid for B-A translocations in maize. *Genetics* **107**, 117 (1984).