ABSTRACT

Title of Dissertation:	CHARACTERIZATION OF GENETIC RECOMBINATION AND ITS INFLUENCING FACTORS IN CATTLE
	Botong Shen, Doctor of Philosophy, 2017
Dissertation directed by:	Assistant Professor Li Ma, Department of Animal and Avian Sciences

Meiotic recombination is a fundamental biological process in which parental genetic materials are exchanged during egg or sperm development. Recombination is necessary for proper chromosomal disjunction during meiosis. Aberrations in this process have been confirmed as the cause of aneuploidy, leading to a potentially deleterious outcome. Along with mutation, recombination is a major force to promote genetic diversity and drive the evolution of genomes. Despite the importance of recombination, the frequency and location of recombination vary wildly within and between individuals, populations, and species. In this thesis, I characterized patterns of recombination in the cattle genome and conducted a comprehensive study of the effect of genetics, sex and age on recombination and its evolution using a uniquely large cattle database hosted at the USDA, where over a million animals with full pedigree information have been genotyped and new data are being generated at an increasing speed. First, we characterized five PRDM9 alleles and generated allele-

specific recombination maps using data derived from over 239,000 meioses in Holstein. We found one allele of PRDM9 to be very different from others in both protein composition and recombination landscape. By comparing recombination maps from sperm and pedigree data, we validated the quality of pedigree-based results. Second, we extended our analysis in recombination patterns to four major U.S. dairy cattle breeds, Holstein, Jersey, Ayrshire, and Brown Swiss. We identified over 8.9 million crossover events and constructed eight genome-wide recombination maps for the two sexes in four cattle breeds. We confirmed a longer male genetic map in bovine and found breed-specific recombination hotspots. Our GWAS analyses confirmed seven loci associated with genome-wide recombination rate and the association of the PRDM9 gene with hotspot usage in two sexes and multiple cattle breeds. Third, we explored the plastic nature of recombination in cattle by examining the effect of maternal age and temperature using data derived from 36,999 three-generation families in Holstein for which temperature data were available. We presented a quadratic relationship between recombination frequency and maternal age and a positive correlation between temperature and recombination rate. By analyzing large genomic datasets with pedigree information in cattle, these studies advanced our understanding of meiotic recombination in a domestic livestock species.

CHARACTERIZATION OF GENETIC RECOMBINATION AND ITS INFLUENCING FACTORS IN CATTLE

by

Botong Shen

Dissertation submitted to the Faculty of the Graduate School of the
University of Maryland, College Park, in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
2017

Advisory Committee:
Assistant Professor Li Ma, Chair
Professor Jiuzhou Song
Associate Professor Steve Mount
Dr. George Liu
Dr. Derek M. Bickhart

Acknowledgements

First and foremost, I wish to express my thanks to my advisor, Dr. Li Ma, for your guidance, inspiration, patience, and understanding in both research and personal life. I particularly want to thank you for giving me the opportunity to join your laboratory and work on these exciting projects.

To Dr. Jiuzhou Song, thank you for your suggestions on my studies and personal life.

To Dr. George Liu, I am incredibly grateful for your help for our collaborations project.

I would also like to acknowledge Dr. Sridhar Hannenhalli, Dr. Derek Bickhart and Dr. Steve Mount for their comments, advice, and support.

I want to thank my labmates, Jicai Jiang, Padhi Abinash, Zhiying Wang and friends from Dr. Song's and Dr. Liu's labs. We shared an office and great times together.

I thank the Animal and Avian Sciences department for a great and supportive environment.

Finally, I'd like to thank my family members for their love, support and encouragement.

Table of Contents

List of Abbreviations

AY:	Ayrshire
bp:	Base pairs
BS:	Brown Swiss
CCDC43:	Coiled-coil domain containing protein43
CI:	Confidence interval
cM:	centiMorgan
CO:	Crossover
CPLX1:	Complexin 1
DMC1:	DNA meiotic recombinase 1
DNA:	Deoxyribonucleic acid
DSB:	Double-strands breaks
GAK:	Cyclin G associated kinase
GWAS:	Genome-wide association study
H3K4me3:	Histone H3 lysine 4 trimethylation
HFM1:	ATP dependent DNA helicase homolog 1
HO:	Holstein
HS:	Hotspot
HWE:	Hardy-Weinberg equilibrium
JE:	Jersey
Kb:	Kilobase
LD:	Linkage disequilibrium
MAF:	Minor allele frequency
Mb:	Megabase
MLH:	MutL homolog
MSH:	MutS homolog
NCBI:	National Center for Biotechnology Information
NEK9:	NIMA related kinase 9
NUB1:	Negative regulator of ubiquitin like proteins 1
PCA:	Principal component analysis
PCGF9:	Polycomb group ring finger 3

PCR:	Polymerase chain reaction
PDZK1:	PDZ domain containing protein 1
PRDM9:	PR/SET domain 9
RAD51:	RAD51 recombinase
REC8:	Meiotic recombination protein 8
RNA:	Ribonucleic acid
RNAseq:	RNA sequencing
RNF212:	Ring finger protein 212
SC:	Synaptonemal complex
SYCE:	Synaptonemal complex central element Protein
SYCP:	Synaptonemal complex protein
SNP:	Single-nucleotide polymorphism
TEX12:	Testis expressed 12
TSS:	Transcription starting site
UCSC:	University of California, Santa Cruz
UGCG:	UDP-Glucose ceramide glucosyltransferase
Znf:	Zinc finger

Chapter 1: Review of Literature

Introduction

Meiotic recombination is a major force that drives the evolution of genome structure by creating a mosaic of parental genetic materials and passing it to the next generation. It also facilitates the proper pairing and segregation of homologous chromosomes during meiotic prophase I. Aberrant meiotic recombination has been confirmed as the cause of aneuploidy that leads to a potentially deleterious outcome (Hassold and Hunt 2001, Lipkin, Moens et al. 2002). The frequency and placement of recombination have been found to vary wildly between and among genders, ages, individuals, populations and species.

The meiotic recombination process starts with a double strand break (DSB). The DSBs can be repaired either as a crossover or as a non-crossover (Gerton and Hawley 2005, de Massy 2013). In this chapter, I will first review the biological mechanisms involving meiotic recombination, including the undergoing chromosomal changes during meiosis prophase I, the proteins involved, and the timing differences between males and females.

I will then introduce the current methods used to detect recombination events. This includes molecular assays that directly observe the recombination during certain meiotic stages and also includes indirect approaches that could infer recombination events from genotype data of unrelated individuals in a population or related individuals in pedigree.

I will also describe the variation in recombination frequency and recombination target sites among different species and the factors that could affect these recombination features. This includes PRDM9 protein and its function in regulating recombination distribution along the genome. Sex is another major contributing factor in recombination frequency as male and female genetic maps differ in both genome-wide and local fine scales.

I will then move on to the genome-wide association studies (GWAS) of the frequency and location of recombination which aimed to identify genetic variants associated with those recombination features. GWAS studies of recombination in humans, mice, and cattle have identified some candidate genes that are conserved across species. In the end, I will review the effect of environmental stresses on recombination and provide some background knowledge about the major dairy cattle breeds in U.S.

As the genetic mechanisms of meiotic recombination are conserved across mammalian species, the research in this dissertation aims to commit a comprehensive study of the effect of genetics, sex and age on recombination and its evolution using a uniquely large cattle database, which would help advance our current understanding of the recombination process.

Overview of Meiotic Recombination

Prophase I of Meiosis

Prophase I of the first meiotic division is a crucial step in the reproductive cycle of sexually reproducing eukaryotes as it is the phase when meiotic

recombination occurs. Prophase I is followed by three other stages of meiosis, metaphase I, anaphase I, and telophase I. Based on the status of the homologous chromosomes, prophase I can be further divided into five different stages, leptonema, zygonema, pachynema and diplonema and diakinesis (Figure 1.1). A pre-meiotic S phase takes place before prophase I where the DNA is replicated and the chromosomes become sister chromatids (Bell and Dutta 2002). Entering prophase I, the first stage if prophase I is the leptonema (derived from Greek meaning “thin threads”), where chromosomes, i.e. each consisting of two sister chromatids, become individualized to form visible strands within the nucleus. The lateral elements of the synaptonemal complex also assemble to bind homologous chromosome (Yang and Wang 2009). This is a very short stage where the structure of chromosomes starts to condense.

During zygonema (derived from Greek meaning “paired threads”), the synapsis of homologous chromosomes takes place, facilitated by the assembly of central elements of the synaptonemal complex. The homologous chromosomes align together by the transverse filament in a zipper like fashion, resulting in pairs of chromosomes equal in length and in position of the centromere (Yang and Wang 2009). The pairing process is highly precise, and the paired chromosomes are also called bivalent or tetrad chromosomes. Double-strand breaks take place at this stage along the synapsis of homologous chromosomes (Gruhn, Rubio et al. 2013).

The pachynema (derived from Greek meaning “thick threads”), is the stage where recombination happens and non-sister chromatids of homologous chromosome may exchange segments of genome. However, sex chromosomes only exchange over

a small region because the shape and structure differences. At this stage, the synaptonemal complex is fully assembled, mediating the exchange of genetic materials, which is done by the repair mechanism that also repairs DSBs. The most accepted mechanism is the Szostak model of recombination (Sugawara and Mikamo 1983). When the DSB occurs, a gap on the DNA is opened and one of the strands grows at the 3'-end and invades a homologous duplex, displacing its old strand, and forming a D loop. Then the strand elongation continues by repair replication until it anneals to the complementary sequences. Repair replication starting from the other 3'-end repairs the gap and the branch migration generates two Holliday junctions. The junctions can be resolved by cutting the inner or the outer strands. As a result, there are two possible outcomes, crossover or non-crossover gene conversion (Baudat, Imai et al. 2013).

In the diplonema (derived from Greek meaning "two threads") stage, the SC degrades and so the homologous chromosomes separate from each other a little. Some transcription of the DNA happens as the structure of chromosomes uncoils a bit; however, the homologous chromosomes remain tightly bound at chiasmata, where the crossover happens. The chiasmata remain on the chromosome pairs until they are severed at the transition of anaphase I.

The last stage of prophase I is diakinesis (derived from Greek meaning "moving through"). At this stage, the structure of chromosomes continues to condense, resulting in the four parts of the visible tetrads. The sites of the chiasmata overlap together, making them easily recognizable. The nucleoli disappear, the nuclear membrane disintegrates, and the meiotic spindle begins to form.

The prophase I is followed by metaphase I, anaphase I, and telophase I. During these stages, the homologous chromosomes are pulled to opposite poles of the cell by spindle. As a result, each daughter cell now has half the number of chromosomes but each chromosome consists of a pair of chromatids. The chromosomes uncoil back into chromatin while the sister chromatids remain attached until Meiosis II.

Proteins involved in recombination.

Meiotic recombination happens after the synapsis of homologous chromosomes, which is assisted by the formation of synaptonemal complex (SC) (Zickler 2006). The synaptonemal complex is zipper-like structures between homologous chromosomes and consists of three structural elements (Figure 1.2). Lateral elements of SC that attach to the homologous chromosome are formed by SYCP2 and SYCP3. The transverse filament formed by SYCP1 links the lateral elements to the central element and holds the two homologous chromosomes together. The central elements are in parallel of the chromosome axis and consist of SYCE1, SYCE2, SYCE3, and TEX12 (Westergaard and von Wettstein 1972). The lateral elements begin to form during leptonema and complete their pairing during the zygonema stage. The SC usually becomes disassembled after pachynema stage. A programmed DNA double-strand break (DSB) follows the formation of SC, and the location of DSB is affected by the action of PRDM9 protein, which places H3 lysine 4 trimethylation marks (H3K4me3) on the genome (Parvanov, Petkov et al. 2010). DSBs are catalyzed by the meiotic topoisomerase-like protein SPO11. SPO11 is broadly conserved across eukaryotes (Malik, Ramesh et al. 2007), and two isoforms

of SPO11 were found, SPO11α and SPO11β. Previous studies suggested that SPO11β rather than SPO11α is responsible for generating most meiotic DSBs (de Massy 2013). The open 3' single-strand DNA is then recognized by RAD51 and DMC1, which facilitate the in strand invasion and homologous pairing of the open 3' single-strand DNA. The next stage mediated by two DNA binding protein MLH1 and MSH4 are crucial to the outcome of the DSB repair as the recombination intermediates are processed either to Holliday junctions (crossover intermediates) or single-end strand invasions (non-crossover intermediates) (Grelon 2016). In the final diplonema substage, the homologues are physically held together by the crossovers, a structure called chiasmata, during the end of prophase when cohesions are removed. Furthermore, chiasmata are necessary to stabilize the homologs on the metaphase I plate and to promote normal segregation at anaphase I (Figure 1.3).

The Timing of Meiotic Recombination

Previous studies have showed major differences in recombination frequency and placement between males and females in mammal (Lynn, Ashley et al. 2004). Due to the physiological differences between male and female gamete generation cycles, the timing of meiotic recombination varies wildly. In humans, the female starts the meiotic process through a diplotene stage of prophase I in fetal ovaries. After birth, the oocytes undergoes an arrest period before prophase I, during which the synapsis and the chiasmata remain intact. Meiosis resumes after puberty and only a part of the oocytes would ovulate and the meiosis process for these oocytes could vary in decades (Crow 2000).

For males, the spermatogonium stem cells undergo meiosis continuously through the reproduction lifespan of the individual. However, as spermatogonium stem cells also undergo mitotic proliferation with approximately 64 days per spermatogenesis cycle, mutations are accumulated in older fathers due to replication error. This contributed to the paternal age effect on germlines (Arbeithuber, Betancourt et al. 2015). However, recent studies regarding the paternal-age-related psychiatric disorders in offspring suggested age-related mutations are unlikely to explain much of the increased risk of psychiatric disorders in children of older fathers, as the genetic risk factors shared by old fathers turned to be a credible explanation (Gratten, Wray et al. 2016).

Methods for Detection of Recombination

A recombination map shows the frequency of recombination events proportional to the physical distances along each chromosomes. The unit of recombination is Morgan that represents the number of recombination events between two physical locations. Recombination rate is usually measured in centimorgan (cM) units, which is 0.01 Morgan. One centimorgan means that there is a 1% chance that two physical locations will be separated by a recombination event in one meiosis.

Recombination rates can be measured both directly and indirectly. Indirect methods build recombination maps using the genetic data from pedigree or unrelated individuals in a population. The direct methods measure the recombination frequency by examining the chiasma or recombination related proteins during some specific stages of meiosis.

Pedigree Method

The principle of using pedigree to detect recombination is to track the transmission of alleles from one generation to the next generation. Although it is the first method used to infer recombination events, it is still widely used to date. By tracing the allele transmission, we can identify recombination events between two genetic markers. The main advantages of pedigree based study is that a genetic map can be constructed and recombination events are able to be assigned to each individuals. And the location as well as the genotype of the genetic markers are crucial as higher density marker would result in a finer scale genetic map. The main challenge of the pedigree method is the cost of collecting sample and genotyping for each individual, and usually a large sample size is preferred for construction of recombination map in a higher resolution.

Linkage Disequilibrium Method

Linkage disequilibrium (LD) based methods were developed to infer recombination rate in a population, in which the genotyped individuals can be unrelated. This method is based on the observation that adjacent SNP markers tend to form clusters and associate together as a block. This phenomena is defined as high levels of linkage disequilibrium, which is a measurement of the association between loci. When an allele from one locus is inherited completely independent of another allele from another location, they are considered to be in linkage equilibrium. The LD between markers are broken down by recombination events in the intervening space between them, with higher recombination rates resulting in more rapid decay of LD. This method is particularly suited for natural populations where only a few samples

are available and pedigree data is lacking. A major drawback of the linkage disequilibrium method is that only sex-average results will be generated.

Single-Sperm Genome-Wide Genotyping

This method separates single sperm from one individual followed by whole-genome amplification and high-throughput genotyping or sequencing. Then a genome-wide individual phasing/haplotyping can be carried out and an individual-level recombination map can be constructed. This method can detect crossover events, non-crossover events, mutations, and even aneuploidy. Using this method, Wang et al. reported an average of 22.8 recombination events, 5 to 15 gene conversion events, and 25 to 36 de novo mutations in human sperm cells (Wang, Fan et al. 2012). Another study of 99 sperm from a single Asian individual reported aneuploidy in 4% of the cells and 26 recombination events per single-sperm cell in humans (Lu, Zong et al. 2012).

Molecular Assay

Recombination events can be mapped by detection of DMC1 associated single-stranded DNA. DMC1 covers the single-stranded DNA resulting from double-strand breaks generated by SPO11. Using targeted immunoprecipitation against DMC1 in chromatin which is extracted from adult testes, the DNA fragments can be obtained and sequenced to map crossover events as well as non-crossovers on the genome.

In addition to DMC1, SPO11 protein can also be used to study meiotic recombination. By detecting the SPO11 oligonucleotides which are covalently

attached to SPO11, a SPO11 mediated DSB events can be mapped. However, this method has only been used in budding yeast (Pan, Sasaki et al. 2011).

PRDM9 and Regulation of Recombination Distribution

PRDM9 Alleles and Recombination Hotspots

PRDM9 (PR domain containing 9) is a histone methyltransferase consisting of three major domains. A N-terminal KRAB (Kruppel-associated box) domain is suggested to be involved in protein-protein interactions. A central PR/SET domain is served as the trimethyl transferase which would work on the lysine 4 on histone 3, leaving H3K4me3 marks on the genome. A C-terminal domain contains varying number of tandem C2H2 zinc finger (ZnF) arrays and is assumed to have a DNA-binding function (Hayashi, Yoshida et al. 2005).

Zinc finger (ZnF) arrays have great variability regarding the number and composition even between closely related species. A number of PRDM9 alleles could also be found within a species. In Holstein and Jersey cattle, the two most common breeds of dairy cattle in the U.S., we have sequenced and reported a total of ten different ZnF arrays and five different PRDM9 alleles, most of which carry six zinc finger repeats. We found one allele of PRDM9 to be very different from others regarding the protein composition and recombination landscape, which suggested a causative role of this allele on the association between PRDM9 and recombination in the cattle population we identified previously (Ma, O'Connell et al. 2015).

Recombination events are not evenly distributed along the genome. Sometimes more recombination events take place in small genomic regions known as

“hotspots.” Hotspots are typically 1-2 kilobase regions where a significant increase, usually on orders of magnitude, in recombination rate was observed compared to the background recombination rate.

In humans, nine PRDM9 alleles were found with allele A and B being the major alleles (Baudat, Buard et al. 2010). Allele A and B have 13 zinc finger repeats and only differ at one amino acid. Allele A and B are mostly present in European heritage while allele C were later found to be predominate in African heritage (Berg, Neumann et al. 2011). European ancestry with allele A and B have significantly different recombination hotspots than African ancestry with allele C, where allele A and B recognize and bind to a 13-mer motif CCNCCNTNNCCNC, whereas allele C binds to a 17-mer motif CCNCNNTNNNCNTNNCC.

Another study constructed allele specific DSB maps of allele A, B and C in humans (Pratto, Brick et al. 2014). Consistent with previous study, allele A and B shared 88% of DSB hotspots, while allele A and C only shared 43% of the DSB hotspots and the remaining 56% of allele C DSB hotspots were found to be allele C specific. Two distinct consensus motifs were identified to be enriched at the centers of allele A and allele C defined hotspots, in accordance to the previously reported motifs. By studying individual DSB maps with PRDM9 heterozygosity composition, they claimed that the heterozygosity of PRDM9 also modulates hotspot strength.

PRDM9 is a major factor determining the recombination hotspots in many species; however, it appears to be only a piece of the puzzle when we compare recombination landscape across taxa. Evidence from Drosophila and other organisms suggests that there are other mechanism for meiosis when PRDM9 is absent.

Recombination without PRDM9

Organisms that possess PRDM9 gene show a rapid turnover of the recombination hotspots. The differences in the zinc finger repeats and the DNA binding motifs often result in the divergence of recombination rates between species at a fine scale. In mice, mutations of the PRDM9 gene have previously been associated with infertility (Hayashi, Yoshida et al. 2005). In addition, many species can still undergo meiotic recombination in spite of the absence of functional PRDM9, these includes dogs, reptiles, birds, plants, yeast, nematodes, and Drosophila (Birtle and Ponting 2006, Oliver, Goodstadt et al. 2009, Ponting 2011, Heil and Noor 2012). These phenomena raise great questions such as the regulation of recombination distribution in the genomes or the presence of recombination hotspots and turnovers without PRDM9 regulation.

Dog carries inert versions of PRDM9 genes with multiple disruptive mutations (Axelsson, Webster et al. 2012, Auton, Li et al. 2013). The inactivation of PRDM9 was not the result of domestication as the close related wolves and coyotes were also found to carry disrupted PRDM9 (Munoz-Fuentes, Di Rienzo et al. 2011). Despite the absence of functional PRDM9, dogs are able to complete the meiosis process and produce fertile offspring, and the hotspots in dogs are enriched around promoter regions (Auton, Li et al. 2013). This raises a question that whether PRDM9 is required during meiosis.

In the absence of PRDM9, the recombination hotspots persist within genomes. In a recent study in yeast, by comparing genome-wide recombination initiation maps from widely divergent Saccharomyces species, the recombination hotspots were

found to frequently overlap with promoters regions in multiple species. In addition, the hotspot positions are also conserved (Lam and Keeney 2015). Another study in birds constructed fine-scale genetic maps for two bird species: the zebra finch, *Taeniopygia guttata*, and the long-tailed finch, *Poephila acuticauda* (Singhal, Leffler et al. 2015). Recombination hotspots were found in both species and were enriched near promoter regions. Most hotspots were shared between the two species despite of tens of millions of years after divergence. Thus, recombination in species lacking a functional PRDM9 gene still shows similar patterns of hotspot localization and evolution.

Variation in Recombination Features among Species

Recombination Features in Cattle

Recombination in cattle is of great interest to the dairy and beef industries, and much effort of my thesis has been focused here. Previously, two cattle studies were reported on male recombination using the bovine 50K SNP chip. Sandor et al. characterized cattle male meiotic recombination using 10,192 bulls from the Netherlands and 3783 bulls from New Zealand with 19,487 SNPs in common between the two groups (Sandor, Li et al. 2012). Weng et al. reported male recombination features and related genetic loci in beef cattle with a moderate sample size,2,778 Angus and 1,485 Limousin sire-offspring pairs (Weng, Saatchi et al. 2014). Sandor et al. (2012) reported an association between recombination hotspot usage and PRDM9 in bulls, but localized the gene to chromosome X. However, PRDM9 has four paralogues in the bovine genome and previous studies have found

signals of positive selection associated with the copy on chromosome 1 (Oliver, Goodstadt et al. 2009). In a large-scale study of cattle sex-specific recombination, we have recently shown that the PRDM9 paralogue on chromosome 1 is associated with recombination hotspot usage in the U.S. Holstein population (Ma, O'Connell et al. 2015). A recent study using three cattle populations from France, New Zealand, and Netherlands confirmed a longer genetic map in males (23.3 Morgan) than in female (21.4 Morgan), their GWAS results showed RNF212B, a paralog for RNF212, might also be involved in recombination (Kadri, Harland et al. 2016).

Recombination Features in Other Species

Recombination rates vary between individuals, populations, and species. The past few years have seen remarkable progress in the development of fine scale maps and in revealing novel modifiers of recombination rate. However, some comparisons of recombination maps show seemingly contradictory results, particularly in relation to conservation and divergence of recombination rates. It is important to point out that studies in variation in recombination rate among other species can benefit the study of recombination in cattle in an evolutionary point of view. Recombination studies have been done in a variety of species, especially in *Arabidopsis*, fruit fly, and *C.elegans*. A list of previous studies in vertebrate is shown in Table 1.1. It is worth noticing that many recombination studies so far used LD-based methods to estimated recombination rates, without distinguishing the difference of recombination pattern between males and females. In addition, the results from some studies using the pedigree-based methods still lack conclusive results, due to the low resolution of the genotype data, small sample size, and lacking of complete pedigree information.

Recombination rates vary in closely related species. Humans diverged from chimpanzee around six million years ago, though human and chimpanzee share about 96% of their genome. Previous study on chimpanzee constructed a LD-based recombination map, and the recombination patterns tend to be conserved to the sex-average recombination pattern in humans. However, the recombination rates are not conserved at fine scales between human and chimpanzee (Ptak, Hinds et al. 2005). The difference is at the recombination hotspots of the chimpanzee, as the binding motif predicted based on the sequence of Chimpanzee PRDM9 is drastically different than the predicted binding sequence of human PRDM9. In fact, across all orthologous zinc finger protein, PRDM9 shows the most divergence among them. In addition, PRDM9 shows rapid evolution with variation in the number of zinc finger as well as the constitution of the zinc finger repeats (Oliver, Goodstadt et al. 2009, Myers, Bowden et al. 2010). It was suggested and recently proved by us in cattle that these changes in PRDM9 would have generated recombination rate differences.

Sexual Dimorphism in Recombination

In cattle, we've characterized the sex difference of recombination features in both large and fine scales. In Holstein, the male map is 10% longer than the female map, and the sex difference is mostly pronounced in the subtelomeric regions. We identified 1,792 male and 1,885 female putative recombination hotspots, with 720 hotspots shared between sexes (Ma, O'Connell et al. 2015). Kadri et al. conformed that average number of COs was found to be larger in males (23.3) than in females (21.4) in a combination of cattle population from France, New Zealand and Netherlands (Kadri, Harland et al. 2016).

Studies in sheep also showed that males have a higher recombination rate than females (Johnston, Berenos et al. 2016). However, in most studied species, including humans, it is suggested that mammalian females have a longer recombination map than males (Table 1.1).

One possible explanation for the sex differences in recombination frequency is the different chromatin configurations across meiotic prophase. The length of the SC has been measured, and female SC is substantially longer than male SC. When packaging DNA into a shorter length of SC in males, a larger DNA loop would form and thus a smaller portion of DNA would be susceptible to recombination (Gruhn, Rubio et al. 2013).

Genetic Basis of Recombination

GWAS Studies in Cattle

A study by Sandor et al. using over 10,000 meiosis in bovine sperm cells first identified genetic variants in REC8 and RNF212 to influence the genome-wide recombination rate, while genetic variants in PRDM9 influence genome-wide hotspot usage (Sandor, Li et al. 2012). Our follow-up GWAS study used over 3,200 sires and 53,000 dams of Holstein cattle and identified a total of thirteen loci to have significance effect on recombination rates, including genetic variants in CPLX1, REC8, REC14, NEK9, MSH4 and PRDM9. The GWAS results for recombination hotspot usage showed that a single associated locus in PRDM9 in both males and females (Ma, O'Connell et al. 2015). Another study by Kadri et al. using over 14,000 animals identified 5 likely causative coding variants in HFM1, MSH4, RNF212,

MLH3 and MSH5 and 1 noncoding variants in RNF212B (Kadri, Harland et al. 2016).

GWAS Studies in Other Species

In Soay sheep, a genome-wide association study of global recombination rate showed the strongest associations at locus RNF212, along with 5 other candidate loci at CLPX1, GAK, PCGF3, REC8 and RNF212B (Johnston, Berenos et al. 2016).

In humans and mice, considerable variation in recombination rate among individuals has been discovered from pedigree-based studies (Shifman, Bell et al. 2006, Kong, Thorleifsson et al. 2010). In humans, several genes have been identified to be associated with individual-level variation in recombination rate, including CPLX1, SMH4, RNF212, CCDC43, RAD21L, SMEK1 and PRDM9 (Baudat, Buard et al. 2010, Kong, Thorleifsson et al. 2010, Myers, Bowden et al. 2010, Parvanov, Petkov et al. 2010). Additionally, locations of recombination crossovers are mainly regulated by the PRDM9 protein during the initiation of meiotic recombination (Baudat, Buard et al. 2010, Kong, Thorleifsson et al. 2010, Myers, Bowden et al. 2010, Parvanov, Petkov et al. 2010). Recombination hotspots have been identified in humans and mice, and PRDM9 has been found to be associated with the percentage of crossovers in hotspots that is termed as ‘hotspot usage’ (Parvanov, Petkov et al. 2010). A GWAS study by Chowdhury et al. identified two loci significantly associated with female recombination rate (KIAA1462, PDZK1), and the other two with male rate (UGCG, NUB1) (Fledel-Alon, Leffler et al. 2011).

Recombination in Response to Intrinsic and Extrinsic Factors

Maternal Age Effect on Recombination

Several studies have been done in the past decade regarding the maternal age effect on recombination but yielded conflicting results. In humans, a study in 2004 using 14,140 female meioses found a positive correlation between maternal recombination counts of an offspring and maternal age, with an estimate of 0.082 recombination events more per year (s.e. = 0.012; $P < 1 \times 10^{-8}$) (Kong, Barnard et al. 2004). Coop et al. analyzed pedigree data from 728 meiosis and found that mothers over the age of 35 have 3.1 more recombination events than mothers under 35 years of age (Coop, Wen et al. 2008). However, a study on 195 maternal meiosis in French-Canadian pedigree showed a negative relation between recombination and maternal age with an effect of -0.4 per year (Hussin, Roy-Gagnon et al. 2011). Another study on Asian populations also detected a weak but significant negative effect of maternal age on recombination rate (Bleazard, Ju et al. 2013). A multi-cohort analysis on over 6,000 meioses reported significant increase in the crossover count with age (Martin et al, 2015}.

In addition to the cited research on human subjects, a study in mice reported females exhibited a significant decrease in chiasma frequency with increasing age (Speed 1977) but no other studies have been conducted in mammals.

Temperature Effect on Recombination

The temperature effect on meiotic recombination rates is complicated and mostly done in small organisms which the environmental factors can be regulated.

Positive correlations between temperature and recombination were found in *Arabidopsis thaliana*, *Caenorhabditis elegans* and the *Locusta migratoria* (Church and Wimber 1969, Rose and Baillie 1979, Francis, Lam et al. 2007). A negative correlation was reported in *Allium ursinum and Melanoplus femurrubrum* (Church and Wimber 1969, Loidl 1989) and some found positive correlation in *Drosophila melanogaster* and *Chorthippus parallelus* (Stern 1926). A recent study in Drosophila found that increased exposure to heat shock conditions is associated with a non-linear increase in meiotic recombination rates (Jackson, Nielsen et al. 2015). These suggest that recombination could respond to environmental stress and confer increased adaptive potential to their offspring.

Major Dairy Cattle Breeds in U.S.

Holstein: Holstein originated in Europe, and the major historical development of this breed occurred in Netherlands. The original stock were the black animals and white animals of the Batavians and Friesians, migrant European tribes who settled in the Rhine Delta region about 2,000 years ago. Holstein was first introduced to America in 1852 (Kaupe, Winter et al. 2004).

Jersey: The Jersey breed originated on the Island of Jersey, a small British island in the English Channel off the coast of France. Jersey is one of the oldest dairy breeds and was reported by authorities as being purebred for nearly six centuries. Jersey cattle were brought to the United States in the 1850's (de Roos, Hayes et al. 2008).

Brown Swiss: Brown Swiss originated in Switzerland, it became prominent among dairy breeds about a 100 years ago. The first Brown Swiss breed in the United States was declared in 1906 (Worede, Forabosco et al. 2013).

Ayrshire: The Ayrshire breed originated in the County of Ayr in Scotland, prior to 1800. A majority of the breeding in the Ayrshire was from Dutch or Flemish cattle that were also used in the formation of the Holstein breed (Schulman, Viitala et al. 2004).

A phylogeny study of 134 domesticated bovid breeds reveals that Brown Swiss is closely related to Jersey, as the Finnish Ayrshire is closely related to Holstein (Decker, McKay et al. 2014).

References

Arbeithuber, B., A. J. Betancourt, T. Ebner and I. Tiemann-Boege (2015). "Crossovers are associated with mutation and biased gene conversion at recombination hotspots." Proc Natl Acad Sci U S A 112(7): 2109-2114.

Archibald, A. L., C. S. Haley, J. F. Brown, S. Couperwhite, H. A. McQueen, D. Nicholson, W. Coppieters, A. Van de Weghe, A. Stratil, A. K. Wintero and et al. (1995). "The PiGMaP consortium linkage map of the pig (Sus scrofa)." Mamm Genome 6(3): 157-175.

Auton, A., Y. R. Li, J. Kidd, K. Oliveira, J. Nadel, J. K. Holloway, J. J. Hayward, P. E. Cohen, J. M. Greally and J. Wang (2013). "Genetic recombination is targeted towards gene promoter regions in dogs." PLoS Genet 9(12): e1003984.

Axelsson, E., M. T. Webster, A. Ratnakumar, C. P. Ponting, K. Lindblad-Toh and L. Consortium (2012). "Death of PRDM9 coincides with stabilization of the recombination landscape in the dog genome." Genome research 22(1): 51-63.

Baudat, F., J. Buard, C. Grey, A. Fledel-Alon, C. Ober, M. Przeworski, G. Coop and B. de Massy (2010). "PRDM9 is a major determinant of meiotic recombination hotspots in humans and mice." Science 327(5967): 836-840.

Baudat, F., Y. Imai and B. de Massy (2013). "Meiotic recombination in mammals: localization and regulation." Nat Rev Genet 14(11): 794-806.

Bell, S. P. and A. Dutta (2002). "DNA replication in eukaryotic cells." Annu Rev Biochem 71: 333-374.

Berg, I. L., R. Neumann, S. Sarbajna, L. Odenthal-Hesse, N. J. Butler and A. J. Jeffreys (2011). "Variants of the protein PRDM9 differentially regulate a set of human meiotic recombination hotspots highly active in African populations." Proc Natl Acad Sci U S A 108(30): 12378-12383.

Birtle, Z. and C. P. Ponting (2006). "Meisetz and the birth of the KRAB motif." Bioinformatics 22(23): 2841-2845.

Bleazard, T., Y. S. Ju, J. Sung and J. S. Seo (2013). "Fine-scale mapping of meiotic recombination in Asians." BMC Genet 14: 19.

Campbell, C. L., C. Bherer, B. E. Morrow, A. R. Boyko and A. Auton (2016). "A Pedigree-Based Map of Recombination in the Domestic Dog Genome." G3 (Bethesda).

Church, K. and D. E. Wimber (1969). "Meiosis in the grasshopper: chiasma frequency after elevated temperature and x-rays." Can J Genet Cytol 11(1): 209-216.

Cox, A., C. L. Ackert-Bicknell, B. L. Dumont, Y. Ding, J. T. Bell, G. A. Brockmann, J. E. Wergedal, C. Bult, B. Paigen, J. Flint, S. W. Tsaih, G. A. Churchill and K. W. Broman (2009). "A new standard genetic map for the laboratory mouse." Genetics 182(4): 1335-1344.

Coop, G., X. Wen, C. Ober, J. K. Pritchard and M. Przeworski (2008). "High-resolution mapping of crossovers reveals extensive variation in fine-scale recombination patterns among humans." Science 319(5868): 1395-1398.

Crawford, A. M., K. G. Dodds, A. J. Ede, C. A. Pierson, G. W. Montgomery, H. G. Garmonsway, A. E. Beattie, K. Davies, J. F. Maddox, S. W. Kappes and et al. (1995). "An autosomal genetic linkage map of the sheep genome." Genetics 140(2): 703-724.

Crow, J. F. (2000). "The origins, patterns and implications of human spontaneous mutation." Nat Rev Genet 1(1): 40-47.

de Massy, B. (2013). "Initiation of meiotic recombination: how and where? Conservation and specificities among eukaryotes." Annu Rev Genet 47: 563-599.

de Roos, A. P., B. J. Hayes, R. J. Spelman and M. E. Goddard (2008). "Linkage disequilibrium and persistence of phase in Holstein-Friesian, Jersey and Angus cattle." Genetics 179(3): 1503-1512.

Decker, J. E., S. D. McKay, M. M. Rolf, J. Kim, A. Molina Alcala, T. S. Sonstegard, O. Hanotte, A. Gotherstrom, C. M. Seabury, L. Praharani, M. E. Babar, L. Correia de Almeida Regitano, M. A. Yildiz, M. P. Heaton, W. S. Liu, C. Z. Lei, J. M. Reecy, M. Saif-Ur-Rehman, R. D. Schnabel and J. F. Taylor (2014). "Worldwide patterns of ancestry, divergence, and admixture in domesticated cattle." PLoS Genet 10(3): e1004254.

Fledel-Alon, A., E. M. Leffler, Y. Guan, M. Stephens, G. Coop and M. Przeworski (2011). "Variation in human recombination rates and its genetic determinants." PLoS One 6(6): e20321.

Francis, K. E., S. Y. Lam, B. D. Harrison, A. L. Bey, L. E. Berchowitz and G. P. Copenhaver (2007). "Pollen tetrad-based visual assay for meiotic recombination in Arabidopsis." Proc Natl Acad Sci U S A 104(10): 3913-3918.

Gerton, J. L. and R. S. Hawley (2005). "Homologous chromosome interactions in meiosis: diversity amidst conservation." Nat Rev Genet 6(6): 477-487.

Gratten, J., N. R. Wray, W. J. Peyrot, J. J. McGrath, P. M. Visscher and M. E. Goddard (2016). "Risk of psychiatric illness from advanced paternal age is not predominantly from de novo mutations." Nat Genet 48(7): 718-724.

Grelon, M. (2016). "Meiotic recombination mechanisms." C R Biol 339(7-8): 247-251.

Groenen, M. A., P. Wahlberg, M. Foglio, H. H. Cheng, H. J. Megens, R. P. Crooijmans, F. Besnier, M. Lathrop, W. M. Muir, G. K. Wong, I. Gut and L. Andersson (2009). "A high-density SNP-based linkage map of the chicken genome reveals sequence features correlated with recombination rate." Genome Res 19(3): 510-519.

Gruhn, J. R., C. Rubio, K. W. Broman, P. A. Hunt and T. Hassold (2013). "Cytological studies of human meiosis: sex-specific differences in recombination originate at, or prior to, establishment of double-strand breaks." PLoS One 8(12): e85075.

Hassold, T. and P. Hunt (2001). "To err (meiotically) is human: the genesis of human aneuploidy." Nat Rev Genet 2(4): 280-291.

Hayashi, K., K. Yoshida and Y. Matsui (2005). "A histone H3 methyltransferase controls epigenetic events required for meiotic prophase." Nature 438(7066): 374-378.

Heil, C. S. and M. A. Noor (2012). "Zinc finger binding motifs do not explain recombination rate variation within or between species of Drosophila." PLoS One 7(9): e45055.

Hussin, J., M. H. Roy-Gagnon, R. Gendron, G. Andelfinger and P. Awadalla (2011). "Age-dependent recombination rates in human pedigrees." PLoS Genet 7(9): e1002251.

Jackson, S., D. M. Nielsen and N. D. Singh (2015). "Increased exposure to acute thermal stress is associated with a non-linear increase in recombination frequency and an independent linear decrease in fitness in Drosophila." Bmc Evolutionary Biology 15.

Johnston, S. E., C. Berenos, J. Slate and J. M. Pemberton (2016). "Conserved Genetic Architecture Underlying Individual Recombination Rate Variation in a Wild Population of Soay Sheep (Ovis aries)." Genetics 203(1): 583-598.

Kadri, N. K., C. Harland, P. Faux, N. Cambisano, L. Karim, W. Coppieters, S. Fritz, E. Mullaart, D. Baurain, D. Boichard, R. Spelman, C. Charlier, M. Georges and T. Druet (2016). "Coding and noncoding variants in HFM1, MLH3, MSH4, MSH5, RNF212, and RNF212B affect recombination rate in cattle." Genome Res 26(10): 1323-1332.

Kappes, S. M., J. W. Keele, R. T. Stone, R. A. McGraw, T. S. Sonstegard, T. P. Smith, N. L. Lopez-Corrales and C. W. Beattie (1997). "A second-generation linkage map of the bovine genome." Genome Res 7(3): 235-249.

Kaupe, B., A. Winter, R. Fries and G. Erhardt (2004). "DGAT1 polymorphism in Bos indicus and Bos taurus cattle breeds." J Dairy Res 71(2): 182-187.

Kong, A., J. Barnard, D. F. Gudbjartsson, G. Thorleifsson, G. Jonsdottir, S. Sigurdardottir, B. Richardsson, J. Jonsdottir, T. Thorgeirsson, M. L. Frigge, N. E. Lamb, S. Sherman, J. R. Gulcher and K. Stefansson (2004). "Recombination rate and reproductive success in humans." Nat Genet 36(11): 1203-1206.

Kong, A., G. Thorleifsson, D. F. Gudbjartsson, G. Masson, A. Sigurdsson, A. Jonasdottir, G. B. Walters, A. Jonasdottir, A. Gylfason and K. T. Kristinsson (2010). "Fine-scale recombination rate differences between sexes, populations and individuals." Nature 467(7319): 1099-1103.

Lam, I. and S. Keeney (2015). "Nonparadoxical evolutionary stability of the recombination initiation landscape in yeast." Science 350(6263): 932-937.

Lipkin, S. M., P. B. Moens, V. Wang, M. Lenzi, D. Shanmugarajah, A. Gilgeous, J. Thomas, J. Cheng, J. W. Touchman, E. D. Green, P. Schwartzberg, F. S. Collins and P. E. Cohen (2002). "Meiotic arrest and aneuploidy in MLH3-deficient mice." Nat Genet 31(4): 385-390.

Loidl, J. (1989). "Effects of Elevated-Temperature on Meiotic Chromosome Synapsis in Allium-Ursinum." Chromosoma 97(6): 449-458.

Lu, S., C. Zong, W. Fan, M. Yang, J. Li, A. R. Chapman, P. Zhu, X. Hu, L. Xu and L. Yan (2012). "Probing meiotic recombination and aneuploidy of single sperm cells by whole-genome sequencing." Science 338(6114): 1627-1630.

Lynn, A., T. Ashley and T. Hassold (2004). "Variation in human meiotic recombination." Annu Rev Genomics Hum Genet 5: 317-349.

Ma, L., J. R. O'Connell, P. M. VanRaden, B. Shen, A. Padhi, C. Sun, D. M. Bickhart, J. B. Cole, D. J. Null, G. E. Liu, Y. Da and G. R. Wiggans (2015). "Cattle Sex-Specific Recombination and Genetic Control from a Large Pedigree Analysis." PLoS Genet 11(11): e1005387.

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(12): 2827-2841.

Munoz-Fuentes, V., A. Di Rienzo and C. Vila (2011). "Prdm9, a major determinant of meiotic recombination hotspots, is not functional in dogs and their wild relatives, wolves and coyotes." PLoS One 6(11): e25498.

Marklund, L., M. Johansson Moller, B. Hoyheim, W. Davies, M. Fredholm, R. K. Juneja, P. Mariani, W. Coppieters, H. Ellegren and L. Andersson (1996). "A comprehensive linkage map of the pig based on a wild pig-Large White intercross." Anim Genet 27(4): 255-269.

Morton, N. E. (1991). "Parameters of the human genome." Proc Natl Acad Sci U S A 88(17): 7474-7476.

Myers, S., R. Bowden, A. Tumian, R. E. Bontrop, C. Freeman, T. S. MacFie, G. McVean and P. Donnelly (2010). "Drive against hotspot motifs in primates implicates the PRDM9 gene in meiotic recombination." Science 327(5967): 876-879.

Oliver, P. L., L. Goodstadt, J. J. Bayes, Z. Birtle, K. C. Roach, N. Phadnis, S. A. Beatson, G. Lunter, H. S. Malik and C. P. Ponting (2009). "Accelerated evolution of the Prdm9 speciation gene across diverse metazoan taxa." PLoS Genet 5(12): e1000753.

Pan, J., M. Sasaki, R. Kniewel, H. Murakami, H. G. Blitzblau, S. E. Tischfield, X. Zhu, M. J. Neale, M. Jasin, N. D. Socci, A. Hochwagen and S. Keeney (2011). "A hierarchical combination of factors shapes the genome-wide topography of yeast meiotic recombination initiation." Cell 144(5): 719-731.

Parvanov, E. D., P. M. Petkov and K. Paigen (2010). "Prdm9 controls activation of mammalian recombination hotspots." Science 327(5967): 835-835.
Ponting, C. P. (2011). "What are the genomic drivers of the rapid evolution of PRDM9?" Trends in Genetics 27(5): 165-171.

Pratto, F., K. Brick, P. Khil, F. Smagulova, G. V. Petukhova and R. D. Camerini-Otero (2014). "Recombination initiation maps of individual human genomes." Science 346(6211): 1256442.

Ptak, S. E., D. A. Hinds, K. Koehler, B. Nickel, N. Patil, D. G. Ballinger, M. Przeworski, K. A. Frazer and S. Paabo (2005). "Fine-scale recombination patterns differ between chimpanzees and humans." Nat Genet 37(4): 429-434.

Rose, A. M. and D. L. Baillie (1979). "The Effect of Temperature and Parental Age on Recombination and Nondisjunction in CAENORHABDITIS ELEGANS." Genetics 92(2): 409-418.

Sandor, C., W. Li, W. Coppieters, T. Druet, C. Charlier and M. Georges (2012). "Genetic variants in REC8, RNF212, and PRDM9 influence male recombination in cattle." PLoS Genet 8(7): e1002854.

Schulman, N. F., S. M. Viitala, D. J. de Koning, J. Virta, A. Maki-Tanila and J. H. Vilkki (2004). "Quantitative trait Loci for health traits in Finnish Ayrshire cattle." J Dairy Sci 87(2): 443-449.

Shifman, S., J. T. Bell, R. R. Copley, M. S. Taylor, R. W. Williams, R. Mott and J. Flint (2006). "A high-resolution single nucleotide polymorphism genetic map of the mouse genome." PLoS biology 4(12): e395.

Singer, A., H. Perlman, Y. Yan, C. Walker, G. Corley-Smith, B. Brandhorst and J. Postlethwait (2002). "Sex-specific recombination rates in zebrafish (Danio rerio)." Genetics 160(2): 649-657.

Singhal, S., E. M. Leffler, K. Sannareddy, I. Turner, O. Venn, D. M. Hooper, A. I. Strand, Q. Li, B. Raney, C. N. Balakrishnan, S. C. Griffith, G. McVean and M. Przeworski (2015). "Stable recombination hotspots in birds." Science 350(6263): 928-932.

Speed, R. M. (1977). "The effects of ageing on the meiotic chromosomes of male and female mice." Chromosoma 64(3): 241-254.

Stern, C. (1926). "An Effect of Temperature and Age on Crossing-Over in the First Chromosome of Drosophila Melanogaster." Proc Natl Acad Sci U S A 12(8): 530-532.

Sugawara, S. and K. Mikamo (1983). "Absence of Correlation between Univalent Formation and Meiotic Nondisjunction in Aged Female Chinese Hamsters." Cytogenetics and Cell Genetics 35(1): 34-40.

Tortereau, F., B. Servin, L. Frantz, H. J. Megens, D. Milan, G. Rohrer, R. Wiedmann, J. Beever, A. L. Archibald, L. B. Schook and M. A. Groenen (2012). "A high density recombination map of the pig reveals a correlation between sex-specific recombination and GC content." BMC Genomics 13: 586.

Wang, J., H. C. Fan, B. Behr and S. R. Quake (2012). "Genome-wide single-cell analysis of recombination activity and de novo mutation rates in human sperm." Cell 150(2): 402-412.

Weng, Z. Q., M. Saatchi, R. D. Schnabel, J. F. Taylor and D. J. Garrick (2014). "Recombination locations and rates in beef cattle assessed from parent-offspring pairs." Genet Sel Evol 46: 34.

Westergaard, M. and D. von Wettstein (1972). "The synaptinemal complex." Annu Rev Genet 6: 71-110.

Wong, A. K., A. L. Ruhe, B. L. Dumont, K. R. Robertson, G. Guerrero, S. M. Shull, J. S. Ziegle, L. V. Millon, K. W. Broman, B. A. Payseur and M. W. Neff (2010). "A comprehensive linkage map of the dog genome." Genetics 184(2): 595-605.

Worede, G. M., F. Forabosco, B. Zumbach, V. Palucci and H. Jorjani (2013). "Evaluation of genetic variation in the international Brown Swiss population." Animal 7(7): 1060-1066.

Yang, F. and P. J. Wang (2009). "The Mammalian synaptonemal complex: a scaffold and beyond." Genome Dyn 5: 69-80.

Zickler, D. (2006). "From early homologue recognition to synaptonemal complex formation." Chromosoma 115(3): 158-174.

Tables

Species	Year	Authors	Female(cm)	Male(cm)	Female/Male ratio
Cattle	1997	Kappes et.al	2879	2808	1.03
	2015	Ma et.al	2320	2550	0.91
	2016	Druet et.al	2140	2330	0.92
Human	1991	Morton	4782	2809	1.70
	2002	Kong et.al	4281	2591	1.65
	2008	Coop et.al	4133	2599	1.53
	2010	Kong et.al	4071	2287	1.78
	2013	Bieazard et.al	3934	2503	1.57
	2014	Kong et.al	3868	1925	2.01
	2015	Campbell et.al	4355	2707	1.61
Mouse	2009	Cox et.al	1495	1375	1.09
Sheep	1995	Crawford	1907	2143	0.89
	2016	Pemberton et.al	2860	3748	0.77
Pig	1995	Archibald et.al	2150	1650	1.30
	1996	Marklund et.al	2565	1830	1.40
	2012	Groenen et.al	2244	1782	1.26
	2012	Groenen et.al	2545	1747	1.46
Dog	2010	Wong et.al	2276	1909	1.19
	2016	Campbell et.al	2162	1816	1.19
Chicken	2009	Groenen et.al	3097.7	2913.7	1.08
Zebrafish	2002	Singer et.al	2582.7	942.5	2.74

Table 1.1 Autosomal map length estimates in several vertebrate species.

Figures

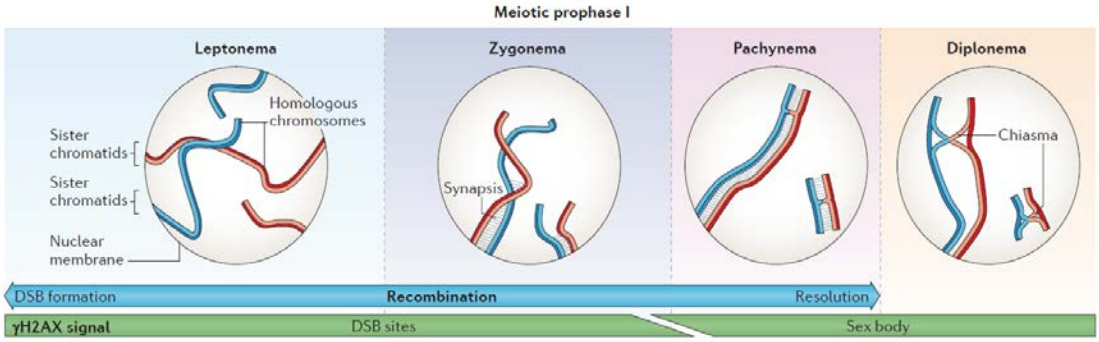


Figure 1.1 Chromosome organization and cytology during meiotic prophase I. Chromosome organization during meiotic prophase I (consisting of leptonema, zygonema, pachynema and diplonema) is illustrated with two pairs of homologous chromosomes, each split into two sister chromatids. Adopted from Baudat (2013).

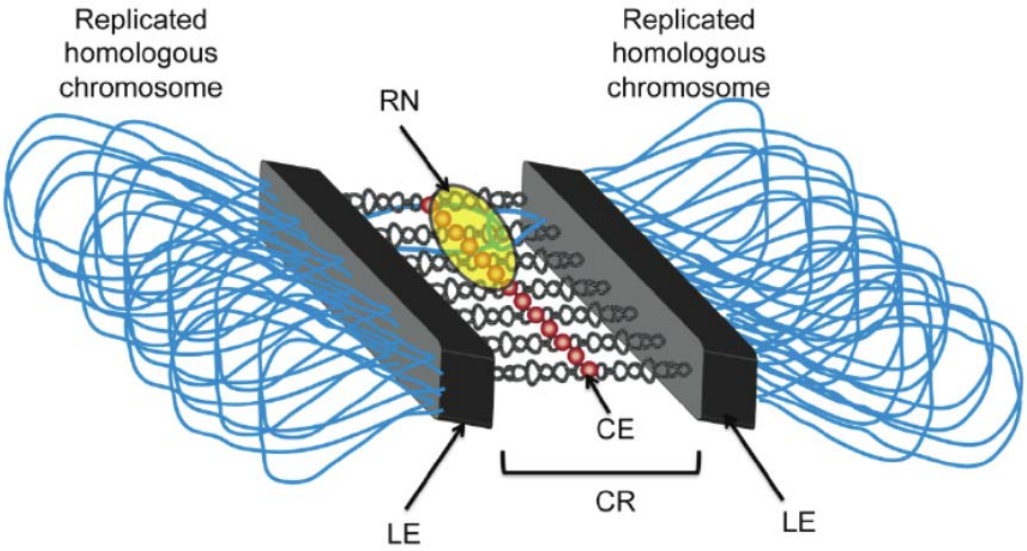


Figure 1.2 Schematic representation of synaptonemal complex. The replicated homologous chromosomes are anchored to the lateral elements (LEs) of the synaptonemal complex (SC) while the genetic exchange between these homologous (referred as crossing over) takes place at the late recombination nodule (RN), that is tethered to the central region (CR). Adopted from Hernández (2012).

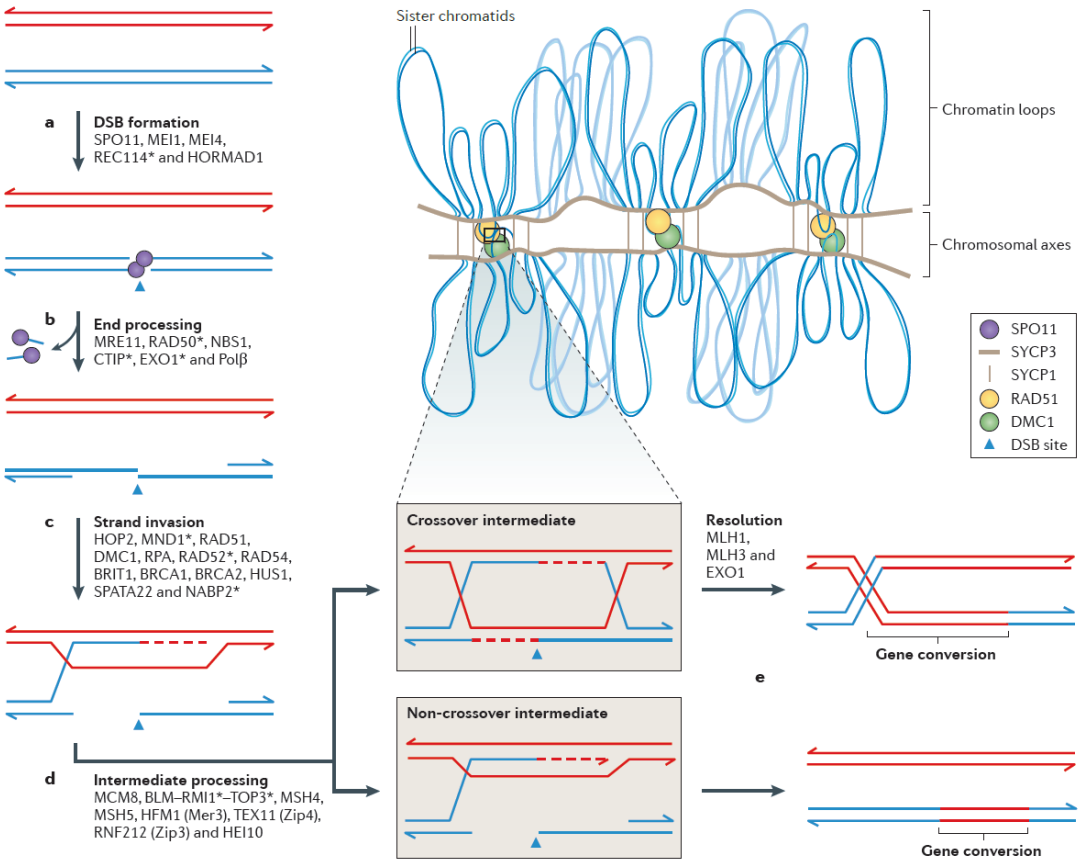


Figure 1.3 Proteins involved in mammalian meiotic recombination. Adopted from Baudat (2013).

Chapter 2: Construction of PRDM9 Allele-Specific Recombination Maps in Cattle Using Large-Scale Pedigree Analysis and Genome-Wide Single Sperm Genomics

Abstract

PRDM9 contributes to hybrid sterility and species evolution. However, its role is to be confirmed in cattle, a major domesticated livestock species. We previously found an association near PRDM9 with cattle recombination features, but the causative variants are still unknown. Using millions of genotyped cattle with pedigree information, we characterized five PRDM9 alleles and generated allele-specific recombination maps. By examining allele-specific recombination patterns, we observed the impact of PRDM9 on global distribution of recombination, especially in the two ends of chromosomes. We also showed strong associations between recombination hotspot regions and functional mutations within PRDM9 zinc finger domain. More importantly, we found one allele of PRDM9 to be very different from others in both protein composition and recombination landscape, indicating the causative role of this allele on the association between PRDM9 and cattle recombination. When comparing recombination maps from sperm and pedigree data, we observed similar genome-wide recombination patterns, validating the quality of pedigree-based results. Collectively, these evidence supported PRDM9 alleles as causal variants for the reported association with cattle recombination. Our study comprehensively surveyed the bovine PRDM9 alleles, generated allele-specific recombination maps, and expanded our understanding of the role of PRDM9 on genome distribution of recombination.

Introduction

Meiotic recombination promotes population diversity by reshuffling parental genetic variants into the next generation and providing novel combinations of genes for selection and evolution (Barton and Charlesworth 1998, Stumpf and McVean 2003, Kauppi, Jeffreys et al. 2004, Coop and Przeworski 2006, Paigen and Petkov 2010). Meiotic recombination is also important in determining proper chromosomal segregation (Coop and Przeworski 2006). Recombination hotspots are usually clustered into narrow genomic regions that are specified by the PR domain-containing 9 (PRDM9) gene in human and mouse (Baudat, Buard et al. 2010, Berg, Rita et al. 2010, Myers, Bowden et al. 2010, Parvanov, Petkov et al. 2010). The high polymorphism level (number and type) in the tandem-repeat zinc finger (ZnF) regions of PRDM9 has drawn wide interest and attention (Kono, Tamura et al. 2014, Schwartz, Roach et al. 2014, Ahlawat, Sharma et al. 2016). In mammals, the number of PRDM9 ZnF varies from 6 to 19 with highly diverse ZnF components between and within species, likely evolving under strong positive selection (Coop and Przeworski 2006, Ponting 2011, Buard, Rivals et al. 2014, Kono, Tamura et al. 2014, Lesecque, Glémin et al. 2014). Some species, like canids, carry inert versions of PRDM9 genes with multiple disruptive mutations (Axelsson, Webster et al. 2012, Auton, Li et al. 2013). Although the polymorphism level of PRDM9 is dramatically higher than other genes in many mammalian species, the diversity of PRDM9 has only recently been documented in cattle (Bos taurus) (Ahlawat, Sharma et al. 2016, Padhi, Shen et al. 2017), which has been domesticated since the Neolithic period (Loftus, MacHugh et al. 1994) and whose effective population size continued to shrink from tens of

thousands prior to domestication to hundreds at present (Boitard, Rodriguez et al. 2016).

Considerable variation in recombination rate between individuals has been documented in mammals and other species (Shifman, Bell et al. 2006, Kong, Thorleifsson et al. 2010, Nachman and Payseur 2012, Balcova, Faltusova et al. 2016, Hunter, Huang et al. 2016). A recent study reported different locations of double-strand breaks between PRDM9 alleles in humans, indicating a critical role of PRDM9 in individual-level variations of recombination (Pratto, Brick et al. 2014). PRDM9 has also been shown to drive evolutionary erosion of hotspots in Mus musculus through haplotype-specific initiation of meiotic recombination (Baker, Kajita et al. 2015). In a further study of Mus musculus PRDM9, Smagulova et al. found hotspot erosion governed the preferential usage of PRDM9 alleles and increased sequence diversity at hotspots that become active in the hybrids (Baker, Kajita et al. 2015). Because crossovers were disfavored at such hotspots, it was assumed that sequence divergence generated by hotspot turnover may create an impediment for recombination in hybrids, potentially leading to reduced fertility and, eventually, speciation. Through these mechanisms, PRDM9 has been considered as an important player in speciation (Payseur 2016). Moreover, re-engineering the ZnFs of PRDM9 with human alleles reversed hybrid sterility in mouse (Davies, Hatton et al. 2016). Despite the important role of PRDM9 on recombination, a genome-wide evaluation of the recombination patterns of different PRDM9 alleles has been lacking in mammals, mainly due to the limited pedigree data in mammalian species.

Large-scale pedigree data is crucial for reconstructing fine-scale recombination maps and for studying patterns of recombination. Genomic evaluations in the cattle industry have accumulated tons of genotype data with pedigree information. The USDA-ARS Animal Genomics and Improvement Laboratory (AGIL) maintains a large database that includes millions of genotyped cattle, a unique resource for studying meiotic recombination with unprecedented power. In a large-scale study of cattle pedigree, recently, we have reported strikingly different recombination patterns between males and females, and identified several loci associated with recombination rate and hotspot usage in both sexes, including the PRDM9 gene on chromosome 1 (Ma, O'Connell et al. 2015). Additionally, in contrast to human and mouse studies that reported a male to female ratio of 0.6 to 0.9 in genome-wide recombination rate (Jensen-Seaman, Furey et al. 2004, Paigen, Szatkiewicz et al. 2008, Kong, Thorleifsson et al. 2010), a much higher ratio of 1.1 to 1.2 was found in cattle (Ma, O'Connell et al. 2015, Kadri, Harland et al. 2016), suggesting marked divergence in the sex-specific recombination rate in these placental mammals. However, it remains unclear what the causative mutations are for the cattle PRDM9 association and how PRDM9 alleles impact recombination features.

While pedigree-based studies have been widely applied, there are two other methods for measuring recombination based on either linkage disequilibrium (LD) patterns or sperm typing. Sperm typing and single-sperm genomics evaluate recombination at either a regional scale or genome-wide level (Hubert, MacDonald et al. 1994, Wang, Fan et al. 2012). Although global recombination maps obtained by

these methods are more or less consistent with one another (Clark, Wang et al. 2010), measurable local differences that attributed to different genetic features cannot be ruled out. Using single sperm genomics approach, Wang et al. reported an average of 22.8 recombination events, 5 to 15 gene conversion events, and 25 to 36 de novo mutations in human sperm cells (Wang, Fan et al. 2012). Using the same method, another study reported aneuploidy in 4% of the cells and 26 recombination events per single-sperm cell in humans (Lu, Zong et al. 2012). These studies affirmed the robustness of single sperm genomics as alternative to pedigree-based approaches in recombination research.

Using large-scale pedigree analysis and genome-wide single sperm genomics, the objective of this study was to 1) Characterize PRDM9 ZnF sequence variations in two cattle breeds; 2) Correlate different PRDM9 alleles with recombination features and generate PRDM9 allele-specific recombination maps in two sexes of cattle; and 3) Assess the reliability of pedigree and single sperm-typing based approaches and compare recombination patterns between sperms and live-born offspring.

Materials and methods

Validation of Bovine Prdm9 Gene Structure

The Ensembl database has two coding sequences for the Zinc finger (ZnF) repeat of the bovine PRDM9 gene, located respectively at Chr1: 45,034,069-45,034,571 (ZnF1) and Chr1: 45,078,067-45,078,685 (ZnF2) (Hubbard, Barker et al. 2002). Using evidence from multiple sources, we confirmed ZnF1 as the correct ZnF coding sequence for PRDM9. First, we separately amplified the two ZnF repeat

regions using genomic DNA of 8 bulls from two cattle breeds (Holstein and Hereford). Due to the high diversity in the ZnF regions of PRDM9 in other species, we expected to observe polymorphisms in the ZnF region of bovine PRDM9. However, we found nucleotide variation in only the ZnF1 region; no polymorphisms were observed in the ZnF2 region. To further validate the two ZnF sequences for PRDM9, we attempted to amplify them from the cDNA of bovine testis tissue using two specifically designed primers. Only ZnF1 was successfully amplified (Supplementary Figure 2.1). The reverse primer for ZnF1 is located in a non-coding region, suggesting that ZnF1 and ZnF2 are disconnected in the mRNA level. These results were used to confirm the gene structure of bovine PRDM9, and we used the ZnF1 region of PRDM9 in all following analyses.

Amplification and Sequencing of Cattle PRDM9 Znf Repeats

Cattle testis tissues were collected as per the ethical guidelines of USDA-ARS animal use and care protocol. The testis was snap-frozen in liquid nitrogen immediately after excision and kept at -80 ºC until further use. Total RNA was isolated and reverse-transcribed to cDNA. Frozen semen and somatic tissue of animal HOUSA000072190767 (Supplemental Table 2.2), together with its parent somatic tissues were donated by Selected Sires, Inc (Plain City, OH , USA). Other Holstein and Jersey cattle DNA samples were obtained from the Cooperative Dairy DNA Repository (CDDR) at USDA-ARS. The PCR and sequencing primers were designed using Primer-BLAST (Ye, Coulouris et al. 2012), and the PRDM9 gene in the cattle reference genome (UMD 3.1.1) was used as template (Zimin, Delcher et al. 2009). We confirmed the specificity of primers by comparing with the newest version of the

cattle genome (Btau_5.0.1, GCA_000003205.6). The primer pairs for PCR amplification using cDNA as the template were designed by crossing two or more exons to avoid potential DNA contamination. The primer pairs for PRDM9 ZnF repeat amplification were designed with 579 bp (5' end) and 157 bp (3' end) of unique, non-repetitive flanking sequences around the ZnF repeat. All the primers used in the present study are listed in Supplementary Table 2.1.

The PCR amplification was performed with 50 μL reaction volume according to Taq DNA polymerase manufacturer's protocol (Taq PCR Master Mix Kit, Qiagen, Hilden, Germany), and the genomic DNA was amplified on a bioRad MyIQ thermocycler. The PCR cycle for PRDM9 ZnF repeat amplification was as follows: initial denaturation at 95 °C for 5 min; followed by 40 cycles of 95 °C for 45 s, annealing at 63°C for 40 s; primer extension at 72 °C for 1 min 20 s; and final extension at 72 °C for 10 min. All the amplified products were run in 1.5% agarose gel, the bands were cut, and DNA was purified with the MinElute Gel Extraction Kit provided by the Qiagen (Hilden, Germany). Purified PCR products were ligated to the pGEM-T Easy vector using the pGEM-T Vector System I (Promega, Madison, WI, USA) and transformed into DH5α (subcloning efficiency) competent cells (Invitrogen / Life Technologies, Carlsbad, CA, USA). Examples of Agarose gel electrophoresis of PCR products are shown in Supplementary Figure 2.2. At least 10 single colonies for each PCR product were randomly picked. Plasmid was extracted using Plasmid Miniprep System (Promega, Madison, WI, USA) and sequenced using the T7 and SP6 primers at two companies, GENEWIZ (South Plainfield, NJ, USA) and Macrogene (Rockville, MD, USA).

To account for the short tandem repetitive nature of ZnF repeats, only forward and reverse sequences with sufficient overlap (at least 168 nt sequences covering two ZnF repeats) were used to prepare the contig and the consensus sequence for each clone was obtained. We used at least three clones with identical sequence to support a valid consensus sequence. To get both alleles with the same length from an individual, we generated at least 6 assembled sequences for each PCR product and used more clones for those individuals whose two alleles showed an unbalanced ratio. The detailed clone number and allele information for all animals were listed in Supplementary Table 2.2 and Supplementary Table 2.3.

Estimation of Recombination Rate in Cattle Pedigree

We used similar approaches as describe before (Fazakerley, Naghiloo et al. 2015). In brief, we extracted three-generation families from Holstein and Jersey pedigree. Within a family, we require that the offspring, at least one parent, and at least one grandparent were genotyped by SNP arrays. In a three-generation family, we phased the two haplotypes of an animal (second and third generations) based on the parental genotypes, and crossover locations were identified by comparing either a paternal or maternal haplotype of an offspring (third generation) to its corresponding parent's two haplotypes (second generation). Based on the location of a crossover, a recombination event was assigned to an interval flanked by two SNPs that are informative (phased heterozygote in the second generation). To construct recombination maps of SNPs, we estimated recombination rate between consecutive SNPs based on the identified crossover events by assigning a recombination event as equal probabilities to all consecutive SNP intervals between the two informative

SNPs. When constructing recombination maps, we only included high-quality data where all members of a three-generation family were genotyped by at least 50K SNP chips. For quality control purposes, we also removed animals (<1%) that have more than 45 crossover events genome-wide, based on the distribution of crossover events in all data. The sex chromosomes were excluded from all analyses due to the poor quality of the genome assembly.

Global and Local Comparisons of Recombination Maps

To show the global distribution of recombination rates along the chromosomes, we adopted a smooth spline model of recombination rates against relative physical locations on chromosomes using the smooth spline function implemented in R 3.2.4 (R 2014, Fazakerley, Naghiloo et al. 2015). We divided the recombination data into subgroups based on the PRDM9 genotype and sex, and generated global recombination maps for each of the PRDM9 alleles in both males and females.

To identify PRDM9 dependent hotspot regions, we compared recombination rate locally in a SNP interval between three PRDM9 genotypes of allele 5 (allele 5 homozygote, allele 5 heterozygote and non-allele 5 genotypes). Given a SNP interval and a pair of groups, we generated a 2×2 table of recombination data (number of meioses w/o recombination in two genotype groups) and applied a Chi-square test to determine whether recombination rate is different between the two groups. Since the frequency of allele 5 in Holsteins is about 9%, there are unbalanced numbers of animals for the three genotype groups for allele 5 (allele 5 homozygote, heterozygote, and non-allele 5 homozygote). Male and female data also have different sample sizes.

To account for the issue of unbalanced sample sizes, we used evidence from multiple sources to select the most significant allele-specific hotspots for allele 5 based on the following rules: 1) Same direction of the difference in recombination rate for males and females; 2) P-value for the comparison between allele 5 homozygote and non-allele 5 genotypes is less than 0.05; 3) P-value in the comparison of allele 5 heterozygote versus allele 5 homozygote or non-allele 5 genotypes is less than 0.05; and 4) At least one P-value from (2) or (3) is less than 2 × 10-6. For non-allele 5, we excluded the comparison between allele 5 homozygote and non-allele 5 genotypes due to the very limited sample sizes and low statistical power.

Whole Genome Amplification, Genotyping and QC of Single Sperm DNA

Sperm cells were obtained from semen samples stored in liquid nitrogen. Frozen semen was thawed in 37 ºC water for 30-45 seconds. Sperm cells were diluted by PBS+1% BSA and washed twice. Then the sperm cells were further diluted to a proper resolution using PBS+1% BSA on a petri-dish, and active single sperm cells were picked up manually by pipetting into a reaction tube under micromanipulator. Whole-genome amplification was performed according to the manufacturer's protocol using a Single Cell Whole Genome Amplification Kit (Yikon Genomics, Shanghai, China) based on the MALBAC (Multiple Annealing and Looping Based Amplification Cycles) technology. In brief, a single sperm cell was initially lysed and pre-amplified by primers supplied in the kit with 8 cycles with multiple annealing steps. Fragments with variable length at random starting positions were generated by polymerase extension for multiple cycles by exponential amplification step. PCR products were purified and sent to Neogen/GeneSeek (Lansing, MI, USA) for

genotyping by the Illumina® BovineHD Genotyping BeadChip assay. PCR amplification quality was confirmed by showing the percentage of SNP successfully genotyped per Mb evenly distributed along the whole genome (Supplementary Figure 2.3). To improve the genotyping accuracy for single sperm cells, we applied a stringent cutoff on the raw genotyping quality score to call genotypes, eliminated heterozygous SNP calls, and removed SNPs that had conflict with the sire genotype. In total, we performed whole genome amplification for 97 single sperm cells.

Phasing and Inferring Crossovers in Single Sperms

Genomic DNA of Animal 102 and the parents were extracted from ear tissues and genotyped together with single sperm DNA at GeneSeek using Illumina® BovineHD Genotyping BeadChip assay. Heterozygous SNPs (hetSNPs) of the bull were phased to two haploids (paternal and maternal) based on the genotypes of two parents. In total, over 76.33% of the 193,066 hetSNPs were phased. Before inferring crossovers in sperms, we applied additional quality control procedures, including general call score ≥ 0.7 for a SNP and homozygote rate ≥ 0.8 for a sample (Supplementary Figure 2.4 and Supplementary Table 2.4). After QC, single sperm samples had phased hetSNP numbers ranging from 11,762 to 87,648. To improve accuracy, we applied the Viterbi algorithm in a hidden Markov model in R to identify crossovers in the sperm haplotype as a transition between paternal and maternal status (R 2014). After filtering samples with abnormal numbers of crossovers (>45), we obtained 1,526 autosomal crossover events from 56 high quality single sperms. Using pedigree data of the same bull, we identified 12,089 crossovers from 556 offspring that were genotyped with 7K, 10K or 50K SNP chips.

Results

Znf Variants of PRDM9 in Two Cattle Breeds

Due to the highly repetitive nature of PRDM9 ZnFs, next-generation sequencing is unable to accurately measure the target sequence because one read cannot capture the full length of the ZnF region. To fully characterize genetic variation of cattle PRDM9, we PCR-amplified, cloned, sequenced, and phased the ZnFs of PRDM9 for 25 and 17 influential bulls, respectively, for Holstein and Jersey, the two most common breeds of dairy cattle in the US. Based on amino acid composition, a total of ten different ZnF arrays (color coded in Figure 2.1) and five different alleles (combination of multiple ZnF arrays) were recovered from the 42 bulls that are representative of the dairy population because these chosen bulls typically had thousands of daughters and sons (Supplementary Table 2.2). The five alleles of PRDM9 were named by the order of allele frequency in the Holstein sample. In contrast to primate and mouse PRDM9 that has 6 to 19 ZnF repeats (Kono, Tamura et al. 2014, Schwartz, Roach et al. 2014), the cattle PRDM9 appears to have fewer ZnF repeats, ranging from 6 to 7 for each allele (Figure 2.1a). With exception of allele 3 that possesses seven ZnF repeats, all the remaining alleles consist of six repeats. Despite the small effective population size of domestic cattle (Boitard, Rodriguez et al. 2016), the amino acid residues at positions –1, 3 and 6 of the zinc finger alpha helix, which were predicted to be in contact with DNA motifs (Oliver, Goodstadt et al. 2009, Schwartz, Roach et al. 2014), are highly polymorphic. Note that the ten different ZnF arrays may have the same composition at these three positions as ZnF arrarys were defined by the full length amino acid sequence. At the

nucleotide level, all of the changes at these three sites are non-synonymous. The amino acid alignments of the cattle ZnF repeats with the corresponding repeats of other species, including human, rhesus monkey, rat and mouse, are shown in Supplementary Figure 2.5. While the ZnFs of cattle PRDM9 show consistent patterns with other species, several amino acid residues are unique to cattle. For example, cattle have Serine and Glycine at positions –4 and –9, respectively, but the corresponding positions in all other species are Glycine and Arginine (Supplementary Figure 2.5).

Although we identified five alleles in the Holstein cattle (Supplementary Table 2.2), only four alleles (alleles 1-4) were observed in Jerseys (Supplementary Table 2.3). As the two breeds are closely related, we observed minor differences in allele frequency between Holstein and Jersey (Figure 2.1a). For instance, allele 1 is most frequent in Holstein (38%), but allele 2 appears to be dominant in Jersey (41.2%). The most striking difference between the two breeds is the uniqueness of allele 5 in Holstein (8%). Interestingly, allele 5 is the most distinct allele compared to other alleles with no shared ZnF arrays between allele 5 and non-allele 5 (Figure 2.1b). The amino acid alignments between allele 5 and the rest of the alleles revealed unique amino acid substitutions at the DNA-contact sites (i.e., –1, 3 and 6) and nearby positions (–2 and 2). Consistently, allele 5 and non-allele 5 (referred to as the rest of the alleles) were predicted to preferentially recognize distinct DNA motifs (Supplementary Figure 2.6). In a previous GWAS study of recombination in cattle, we identified a SNP (rs110661033, Chr1: 45113934) downstream of the PRDM9 gene that was associated with genome-wide hotspot usage and recombination rate

(Fazakerley, Naghiloo et al. 2015). Intriguingly, the minor allele "A" of SNP rs110661033 was perfectly linked with PRDM9 allele 5 in the current study ($R^2 = 1$; Supplementary Table 2.2). All the Holstein bulls that carry one copy of allele 5 are heterozygous at SNP rs110661033 (G/A). We further confirmed this linkage between PRDM9 (non-allele 5/allele 5) and SNP rs110661033 (G/A) by sequencing PRDM9 in an additional five Holstein bulls selected by the SNP genotype, in which two animals are heterozygous at rs110661033 and carry one copy of allele 5 at PRDM9, and three bulls are homozygous (A/A) at rs110661033 and carry two copies of allele 5 (Supplementary Table 2.2). Altogether, these results demonstrate that PRDM9 ZnF alleles are likely the causative mutations associated with the recombination features in cattle, with PRDM9 allele 5/non-allele 5 combinations fully explaining the association between rs110661033 and recombination rate and hotspot usage that we previously reported (Fazakerley, Naghiloo et al. 2015).

PRDM9 Allele-Specific Recombination Maps in Holstein

Superior bulls have been extensively used in dairy cattle breeding through artificial insemination. In the national dairy database maintained at AGIL, a bull typically has hundreds to thousands of daughters genotyped. The 25 Holstein bulls selected for sequencing of PRDM9 were chosen because they have large numbers of genotyped offspring to study recombination (Supplementary Table 2.2). To have an overview of the recombination patterns of PRDM9 alleles, we divided the 25 Holstein bulls into six groups based on their PRDM9 genotypes: allele 1 homozygote, allele 2 homozygote, allele 1/2 heterozygote, allele 3 carrier (homozygote or heterozygote), allele 4 carrier (homozygote or heterozygote), and allele 5 carrier (heterozygote).

After splitting, each group has 3 to 6 bulls and more than 2,300 offspring genotyped by 50K SNP chips, allowing us to generate PRDM9-specific recombination maps for individual groups (Supplementary Figure 2.7). The six recombination maps exhibited similar global patterns of recombination with two peak regions, one near the 10% from the beginning of a chromosome and the other to the end. However, allele 5 carrier group showed a larger recombination rate than other alleles at both peak regions. Note that the difference between allele 5 carrier group and other groups is small due to the limitation of sample size and mixed grouping of heterozygote and homozygote individuals. Still, this observation is consistent with the observed differences in protein sequences between the five alleles, where allele 5 has the most distinct ZnFs compared to other alleles.

Using an imputation-based approach, we next attempted to generate allele-specific recombination maps for each of the five PRDM9 alleles. Based on the LD patterns between PRDM9 alleles and nearby SNPs or haplotypes in the 25 Holstein bulls, we identified markers tagging each of the five PRDM9 alleles (Supplementary Table 2.5). Using this tagging information, we extracted all available recombination data and constructed recombination maps for each of the five PRDM9 alleles in both males and females (Figure 2.2). In total, we extracted 1,369,139 three-generation families in Holstein, with each family including one offspring, at least one parent (maternal or paternal), and at least one grandparent. A total of 239,116 three-generation families were genotyped by chips with at least 50K SNPs. We inferred over 3.7 million paternal and over 2 million maternal crossover events from these three-generation families. As mentioned previously, SNP rs110661033 was in perfect

LD with allele 5 with a maximum R2 value of 1. Using this SNP (MAF = 9.4%) as a tagging marker, we extracted all animals (bulls and cows) carrying allele 5 and the corresponding three-generation families, with the number of families for allele 5 homozygote, allele 5 heterozygote and non-allele 5 homozygote genotypes being 637, 28,759 and 120,990 in males and 719, 15,548 and 72,513 in females, respectively. Similarly, using tagging haplotypes or SNPs, we extracted animals and three-generation families for each of the other four alleles (Supplementary Table 2.5).

With an expanded data set using imputation, we assembled enough data to generate recombination maps for each of the five PRDM9 alleles in two sexes (Figure 2.2). Across the five alleles, the average number of crossovers for males and females are 25.10 and 22.74 respectively, which is consistent with previously reported higher recombination rate in bulls than in cows (Fazakerley, Naghiloo et al. 2015). While males and females exhibited different recombination maps across the genome, most notably near the end of chromosomes, PRDM9 alleles were associated with differences in global recombination maps within each sex (Figure 2.2). One striking observation was that among the five PRDM9 alleles, allele 5 showed the largest recombination maps in both sexes (Figure 2.2a and Figure 2.2b). Bulls with allele 5 homozygote, allele 5 heterozygote and non-allele 5 homozygote genotypes had a genetic map length of 27.0, 25.7 and 24.9 Morgans, respectively; and for cows, the map length dropped to 23.6, 23.4 and 22.6 Morgans for the three genotypes, respectively. These results also confirmed the previous association of SNP rs110661033 with genome-wide recombination rates with the minor allele (linked with allele 5) increasing recombination rate (Fazakerley, Naghiloo et al. 2015). For

both males and females, the largest difference in recombination rate between PRDM9 alleles fell into the two recombination peak regions, with animals carrying two copies of allele 5 having highest recombination rate and animals carrying other alleles showing lower recombination rates (Figure 2.2c and Figure 2.2d). Interestingly, animals carrying one copy of allele 5 (heterozygote) have recombination rates in between the two homozygotes, but are closer to allele 5 homozygous animals, especially in the recombination peak regions. Since allele 5 clearly stands out from the rest PRDM9 alleles in protein composition and recombination patterns, we generated three recombination maps for the three genotypes of allele 5 and non-allele 5 in each of the two sexes (Supp_Data_1).

PRDM9 Allele-Specific Recombination Maps in Jersey

The 17 sequenced Jerseys had 4 PRDM9 alleles (alleles 1-4) and 7 genotype combinations (Supplementary Table 2.3). Similarly, we evaluated the LD patterns between PRDM9 alleles and nearby SNPs or haplotypes in the Jersey samples, and identified allele-tagging markers for Jersey (Supplementary Table 2.6). As the genotyped pedigree is smaller in Jersey than in Holstein, we focused on allele 2 that had enough data to study allele-specific recombination maps. In total, 11 of 17 animals with allele 2 carry at least one minor allele of SNP Hapmap26498-BTA-33060. An association test further confirmed the correlation between this SNP and genome-wide hotspot usage in Jersey (P-value = 5.4 × 10-3). Note that this SNP was not linked with allele 2 in Holsteins, possibly due to different LD patterns between cattle breeds. Using this tagging marker of allele 2 (MAF = 41%), we assembled 26,945, 71,726 and 52,870 three-generation families for allele 2 homozygote, allele 2

heterozygote and non-allele 2 homozygote genotypes, respectively. Among these families, 3,959, 5,660 and 2,106 were genotyped by 50K SNP chips in males, and 1,016, 1,214 and 262 genotyped by 50K chips in females. The male map lengths are 23.3, 23.6 and 24.1 Morgans for allele 2 homozygote, allele 2 heterozygote and non-allele 2 genotypes, respectively; while female maps are 22.0, 22.0 and 22.6 Morgans in length, respectively (Supplementary Figure 2.8). Consistent with the patterns found in Holstein, main differences between PRDM9 alleles in Jersey were also found in the two recombination peak regions: in the centromere peak region, allele 2 homozygote map had the lowest recombination rate in both males and females; and in the telomere region, allele 2 carriers had the lowest recombination rate in females and was close to the lowest in males. Note that recombination rate in Jersey is in general slightly smaller than that in Holstein, likely due to lower polymorphism levels of markers in Jersey as the SNP chips were originally designed for Holstein.

PRDM9 Allele-Specific Hotspot Regions

Recombination rates were calculated between consecutive SNPs for animals that carry different PRDM9 alleles, showing allele-specific distributions of recombination across the cattle genome (Figure 2.3). We here used the term “hotspot region” instead of “hotspot” because our SNP intervals were much larger (average 44 Kb) than typical recombination hotspot regions in human and mouse studies. We focused on the most distinct PRDM9 alleles in this analysis, allele 5 and non-allele 5. To find allele-specific hotspot regions for allele 5 and non-allele 5, we applied a Chi-square test to identify SNP intervals with significantly different recombination rate between the three genotypes of allele 5 and non-allele 5 in both sexes of Holstein

(Figure 2.3). Based on the genome-wide significance level of 8.3×10-7 after Bonferroni correction, in males, we identified 7 SNP intervals with different recombination rate between allele 5 homozygote and non-allele 5 genotypes, 369 intervals between allele 5 heterozygote and non-allele 5 genotypes, and only 1 interval between allele 5 homozygote and allele 5 heterozygote, respectively. In females, the numbers of intervals with different recombination rate between the three genotypes were 2, 36 and 0, respectively. The different number of identified intervals for these comparisons reflected different sample sizes and statistical power, because the recombination data involved more bulls than cows, and had more animals carrying non-allele 5 than allele 5. To account for the issue of unbalanced sample sizes, we adopted multiple evidence to select the most significant allele-specific hotspot regions of allele 5: 1) Same direction of the difference in recombination rate for males and females, 2) Small P-value for the comparison between allele 5 homozygote and non-allele 5 genotype, and 3) Small P-value in the comparison of allele 5 heterozygote versus allele 5 homozygote or non-allele 5 genotype. For non-allele 5, we dropped the comparison between allele 5 homozygote and non-allele 5 genotype due to the limited sample sizes and low statistical power.

Among the top 5 hotspot regions of allele 5, one hotspot region, located at chr13: 14923596-15017558, is shared between males and females (Table 2.1). In males, the recombination rate increased 13 fold from 0.0006 to 0.009 between animals with non-allele 5 and two copies of allele 5. In females, animals carrying two copies of allele 5 showed a five-fold increase. To further investigate the recombination patterns of allele-specific hotspot regions, we zoomed into the local

recombination maps of the three PRDM9 genotypes of allele 5 (Supplementary Figure 2.9). Generally, for the allele 5 favored hotspot regions, we found a striking peak of the recombination rates of allele 5 compared to other genotypes. As for non-allele 5 hotspot regions, we also observed higher recombination rates for animals carrying alleles other than 5. In both cases, recombination rates of allele 5 heterozygote would mimic either allele 5 or non-allele 5 since the heterozygote animals had both alleles and binding motifs.

Dominant Effect of Allele 5 on Recombination in Two Sexes

Using both sharing of recombination hotspot region and correlation in recombination rate, we found animals carrying one copy of allele 5 were more similar to animals with two copies of allele 5 than to animals carrying non-allele 5. To evaluate the sharing of recombination hotspot region, we tentatively defined hotspot regions as the SNP intervals with recombination rate 2.5 standard deviations greater than the genome-wide average. For Holstein males, a total of 1361, 1353, and 1365 hotspot regions were identified for allele 5 homozygote, allele 5 heterozygote and non-allele 5 animals, respectively. Consistent with observations in human studies (Pratto, Brick et al. 2014), each PRDM9 genotype had its unique hotspot regions (Supplementary Figure 2.10). Allele 5 homozygote, allele 5 heterozygote and non-allele 5 genotypes each had 887, 737 and 1,032 unique hotspot regions, with only 97 hotspot regions shared by all three genotypes. Excluding the 97 hotspot regions that were shared by all three genotypes, allele 5 and non-allele 5 had only 47 hotspot regions in common. Consistent with the patterns observed in the global recombination maps, allele 5 heterozygote shared more hotspot regions with allele 5 homozygote

than with non-allele 5 genotype (330 vs 189), indicating an observed dominant effect of allele 5. A similar pattern was found in the Holstein cows (Supplementary Figure 2.10). Allele 5 homozygote, allele 5 heterozygote and non-allele 5 genotypes each had 801, 583 and 855 allele-specific hotspot regions, with 101 hotspot regions shared across all three genotypes. Excluding the common hotspots, allele 5 homozygote and allele 5 heterozygote genotypes shared 314 hotspot regions, but allele 5 heterozygote and non-allele 5 genotypes had only 260 hotspot regions in common. Using a correlation analysis of recombination rates across three PRDM9 genotypes and two sexes, we found higher correlations between allele 5 homozygote and heterozygote than between allele 5 heterozygote and non-allele 5 in both males and females (Supplementary Figure 2.11), confirming the observed dominant effect of allele 5 to other PRDM9 alleles.

Enrichment of PRDM9 Binding Motifs in Recombination Hotspot Regions

Using a computational approach designed for ZnF proteins (Persikov and Singh 2013), we predicted the binding motifs of the PRDM9 alleles (Supplementary Figure 2.6). Since the predicted motifs are almost the same for alleles 1-4, we evaluated the enrichment of PRDM9 binding motifs in recombination hotspot regions by focusing on allele 1 and allele 5. Based on the position weight matrices of the predicted motifs, we extracted the degenerated 17-bp motifs ‘ANNANNANNANNANGGC’ and ‘CGNNANNAGCANNANNA’ for allele 1 and allele 5, respectively. Here, we used allele 1 to represent non-allele 5 alleles because allele 1 was the most frequent and there was little difference in binding motif between allele 1 and other non-allele 5 alleles. In the whole bovine reference genome (UMD

3.1.1), allele 5 motif is 1.17 times more prevalent than allele 1 motif. Compared to this genome-wide baseline ratio of 1.17, non-allele 5 hotspot regions had a decreased ratio of 1.14, and allele 5 hotspot regions had an increased ratio of 1.20 in Holstein males (Supplementary Table 2.7). When measured in various subsets of hotspot regions, this ratio showed a consistent trend, 1.14 in non-allele 5 and allele 5 heterozygous shared hotspot regions, 1.19 in allele 5 homozygous and allele 5 heterozygous shared hotspot regions, 1.14 in non-allele 5 specific hotspot regions, and 1.21 in allele 5 specific hotspot regions. Similar enrichment trends were observed in females (Supplementary Table 2.7). Collectively, although we used computationally predicted motifs and our hotspot regions were larger than typical recombination hotspot regions, we did observe an expected trend that allele 5 motif was enriched in allele 5 hotspot regions and allele 1 motif enriched in non-allele 5 hotspot regions.

Comparison of Recombination Maps from Single Sperm-Typing and Pedigree Data

The frequent usage of artificial insemination in cattle provided an opportunity to compare recombination patterns between sperm cells and live-born offspring. To check the consistency of the recombination patterns inferred from pedigree data, we characterized the recombination patterns of a single Holstein bull using both pedigree and single sperm-typing. The comparison involved 56 high-quality single sperms genotyped on BovineHD (770K) SNP chip and 556 live-born offspring genotyped with various chips of more than 50K SNPs (see Materials and Methods). Using the same approach of pedigree analysis, we calculated recombination rates in single sperm data between adjacent SNPs on a 50K SNP chip, so the two recombination

maps could be compared with the same number of SNPs. Although sperm data have more SNPs than pedigree data to begin with, the sperm-based recombination map supported the differences between different PRDM9 alleles identified from pedigree data (Supplementary Figure 2.12). Overall, recombination rates from the single sperm data (allele1/allele2) showed the highest correlation with the pedigree-based rates from the same bull, and this correlation continuously decreased with recombination maps from animals carrying zero, one and two copies of allele 5 (Supplementary Figure 2.12a). As expected, the sperm recombination rates are more similar to male recombination rates than that of females in all groups. In addition, we observed the same trend using the sharing of top 5% recombination intervals between sperm and pedigree data (Supplementary Figure 2.12b).

By comparing recombination maps from sperm cells and the pedigree of the same bull, we found the same number but slightly different preferred locations of recombination between sperm cells and live-born offspring. At the chromosome level, the number of crossovers from pedigree and sperm data were correlated with $r = 0.77$ across 29 autosomes (Supplementary Figure 2.13). Since the sperm data had a higher density of SNPs, we manually decreased the SNP number of sperm data to levels comparable to pedigree data. When using similar numbers of SNPs in both samples (10K in sperms and 7K~50K in pedigree), we saw no difference in the total number of recombination events: 21.91 ± 0.5 SE (±3.8 SD) in sperms and 21.65 ± 0.12 SE (±3.8 SD) in pedigree data (Supplementary Figure 2.14). To compare the patterns of recombination in pedigree and sperm data, we generated global recombination maps by collapsing all 29 autosomes into a standardized chromosome (Figure 2.4). Overall,

the pedigree and sperm recombination maps showed a very similar pattern except near the two ends of chromosomes (standardized locations 10%-20% and >80% on a chromosome). At the beginning of a chromosome (10%-20%), the pedigree-based map exhibited a higher recombination rate than the sperm-based map. To the end of a chromosome (>80%), the recombination rate of sperms reached a peak near the 85% of a chromosome, whereas the pedigree-based recombination rate continuously increased to the end of chromosomes. To avoid potential biases from SNP coverage in sperm cells, we checked the distribution of SNP numbers along the genome and found no clear deficiency of SNPs to the two ends of chromosomes (Supplementary Figure 2.15). However, a Chi-square test found no significant differences (P-value > 0.05) at both locations, possibly due to the small number of sperms tested.

Discussion

The purpose of this study was to characterize PRDM9 allele-specific localizations of meiotic recombination in widely commercialized cattle breeds (Loftus, MacHugh et al. 1994, Boitard, Rodriguez et al. 2016). Meiotic recombination is known to enhance genetic and phenotypic variations in sexually reproducing organisms, and PRDM9 regulates the location of double-stranded breaks and thus recombination events in most of the placental mammals (Baudat, Buard et al. 2010, Parvanov, Petkov et al. 2010). Recent studies in humans showed that the PRDM9 alleles have dramatic influences on the localization and turnover of the recombination hotspot via increased sequence diversity (Pratto, Brick et al. 2014, Smagulova, Brick et al. 2016); however, it is unclear whether these patterns can also be observed in cattle, which has unique demographic histories and more importantly, persistently

succumbed to intense selective pressures in the past 10,000 years after domestication. In addition, multiple recent GWAS studies have identified candidate genes associated with recombination features in cattle, including PRDM9 (Sandor, Li et al. 2012, Ma, O'Connell et al. 2015, Kadri, Harland et al. 2016). The present study not only identified the PRDM9 allele-specific variations including the amino acid substitutions at three functionally important sites (i.e., -1, 3, and 6) of the ZnFs, but also provided strong evidence of allele-specific localizations of recombination events with most pronounced differences at the two ends of chromosomes in cattle. Importantly, these evidence suggested a specific allele (allele 5) as the causal variant for the PRDM9 association with cattle recombination.

While four out of five PRDM9 alleles were shared between the two cattle breeds and showed similar global recombination patterns, one allele (allele 5) that was unique to Holstein exhibited the highest recombination rates at two recombination peak regions. Interestingly, this pattern is correlated with the observed amino acid substitutions at the functional sites of the PRDM9 ZnFs. In addition, previous studies have reported strong associations between the allele-specific localization and turnover of recombination hotspot regions and the polymorphisms of PRDM9 in humans (Berg, Rita et al. 2010, Pratto, Brick et al. 2014, Smagulova, Brick et al. 2016). Although the diversity levels in PRDM9 alleles and number of ZnF repeats were relatively lower in cattle than primates and rodents (Kono, Tamura et al. 2014, Schwartz, Roach et al. 2014), alleles with unique amino acid substitutions appeared to have dramatic differences in the binding-specificity and the distribution of recombination events. For instance, as observed in this study, animals that carry

allele 5 have unique recombination hotspot regions that are distinguishable from hotspot regions modulated by non-allele 5, consistent with studies in humans (Pratto, Brick et al. 2014, Smagulova, Brick et al. 2016). As the PRDM9 ZnFs were predicted to bind sequence motifs that are enriched in recombination hotspot (Grey, Barthès et al. 2011, Ségurel, Leffler et al. 2011), one might speculate uneven distributions of binding affinities of the PRDM9 alleles across the cattle genome. Computationally, we predicted two17-bp motifs 'CGNNANNAGNANNANNA' and 'ANNANNANNANNANGGC' of the most common allele (i.e., allele 1) and allele 5, respectively. We also reported consistent enrichment patterns of these binding motifs in corresponding recombination hotspot regions. Collectively, these variations in PRDM9 ZnFs are associated with the location and intensity of recombination in cattle. However, due to the limited resolution of SNP densities, the strength of enrichment is relatively low and the predicted binding motifs need to be further validated in future experiments.

Given the fact that sampling and genotype errors may potentially bias the pedigree-based results, we further confirmed our findings using a single-sperm genomics approach (Lu, Zong et al. 2012, Wang, Fan et al. 2012). Although the genome-wide recombination rates from these two approaches were consistent, we found some differences, especially at the two locations of recombination peaks, between the pedigree and sperm-based recombination maps. These findings were in agreement with the previous studies in humans, which showed that although the recombination maps from the pedigree and sperm-typing methods are largely consistent, considerable differences were observed at a higher resolution. Since the

sperms used in the present study were active and viable, the differences in fitness before fertilization are small between the sperm samples and the sperms that ended up in live-born offspring. However, different fitness between sperms and live-born offspring may still lie in the selection process between sperm-egg fertilization and embryo development till birth. Although it is intuitively unclear as to what factors drive such differences, based on our results and previous reports (Ye, Coulouris et al. 2012), we postulate the differences in sperm fitness during and after fertilization to be one of the plausible explanations.

Conclusions

Taken together, in this study we characterized the PRDM9 sequence diversity in multiple cattle breeds and generated PRDM9 allele-specific global and local recombination maps in individual bulls. The large cattle pedigree provided us the power to show differences in genome-wide recombination maps of PRDM9 alleles. For the first time, we showed the impact of PRDM9 on the global distribution of recombination on the genome, particularly in the telomere and centromere regions. Using genome-wide single sperm genotyping, we validated the quality of pedigree-based recombination maps. Collectively, these results will provide new insights into the regulatory functions of PRDM9 on meiotic recombination, which further contribute to our understanding of genome evolution in mammals.

Acknowledgements

This work was supported in part by AFRI grants No. 2011-67015-30183, 2013-67015-20951 and No. 2016-67015-24886 from the USDA NIFA and grant No.

US-4997-17 from the US-Israel Binational Agricultural Research and Development Fund. We thank two anonymous reviewers for the constructive comments that have improved the manuscript. We thank Reuben Anderson and Alexandre Dimitriv for technical assistance. We thank Select Sires, Inc. (Charles Settler and Adam Oswalt) for sample donations. The Council on Dairy Cattle Breeding (Bowie, MD) and the Cooperative Dairy DNA Repository (Columbia, MO) are also acknowledged for providing data used in this study. JBC and GEL were also supported by appropriated projects 1265-31000-096-00, "Improving Genetic Predictions in Dairy Animals Using Phenotypic and Genomic Information", and 8042-31000-104-00, “Enhancing Genetic Merit of Ruminants Through Genome Selection and Analysis”, of the Agricultural Research Service of the United States Department of Agriculture, respectively. Mention of trade names or commercial products in this article is solely for the purpose of providing specific information and does not imply recommendation or endorsement by the US Department of Agriculture. The USDA is an equal opportunity provider and employer. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

References

Ahlawat, S., P. Sharma, R. Sharma, R. Arora and S. De (2016). "Zinc finger domain of the PRDM9 gene on chromosome 1 exhibits high diversity in ruminants but its paralog PRDM7 contains multiple disruptive mutations." PloS one 11(5): e0156159.

Ahlawat, S., P. Sharma, R. Sharma, R. Arora, N. Verma, B. Brahma, P. Mishra and S. De (2016). "Evidence of positive selection and concerted evolution in the rapidly evolving PRDM9 zinc finger domain in goats and sheep." Animal Genetics 47(6): 740-751.

Auton, A., Y. R. Li, J. Kidd, K. Oliveira, J. Nadel, J. K. Holloway, J. J. Hayward, P. E. Cohen, J. M. Greally and J. Wang (2013). "Genetic recombination is targeted towards gene promoter regions in dogs." PLoS Genet 9(12): e1003984.

Axelsson, E., M. T. Webster, A. Ratnakumar, C. P. Ponting, K. Lindblad-Toh and L. Consortium (2012). "Death of PRDM9 coincides with stabilization of the recombination landscape in the dog genome." Genome research 22(1): 51-63.

Baker, C. L., S. Kajita, M. Walker, R. L. Saxl, N. Raghupathy, K. Choi, P. M. Petkov and K. Paigen (2015). "PRDM9 drives evolutionary erosion of hotspots in Mus musculus through haplotype-specific initiation of meiotic recombination." PLoS genetics 11(1): e1004916.

Baker, C. L., S. Kajita, M. Walker, R. L. Saxl, N. Raghupathy, K. Choi, P. M. Petkov and K. Paigen (2015). "PRDM9 drives evolutionary erosion of hotspots in Mus musculus through haplotype-specific initiation of meiotic recombination." PLoS Genet 11(1): e1004916.

Balcova, M., B. Faltusova, V. Gergelits, T. Bhattacharyya, O. Mihola, Z. Trachtulec, C. Knopf, V. Fotopulosova, I. Chvatalova and S. Gregorova (2016). "Hybrid sterility locus on chromosome X controls meiotic recombination rate in mouse." PLoS genetics 12(4): e1005906.

Barton, N. H. and B. Charlesworth (1998). "Why sex and recombination?" Science 281(5385): 1986-1990.

Baudat, F., J. Buard, C. Grey, A. Fledel-Alon, C. Ober, M. Przeworski, G. Coop and B. De Massy (2010). "PRDM9 is a major determinant of meiotic recombination hotspots in humans and mice." Science 327(5967): 836-840.

Berg, I. L., N. Rita, K. W. G. Lam, S. Shriparna, O. H. Linda, C. A. May and A. J. Jeffreys (2010). "PRDM9 variation strongly influences recombination hot-spot activity and meiotic instability in humans." Nature Genetics 42(10): 859-863.

Boitard, S., W. Rodriguez, F. Jay, S. Mona and F. Austerlitz (2016). "Inferring population size history from large samples of genome-wide molecular data-an approximate Bayesian computation approach." PLoS Genet 12(3): e1005877.

Buard, J., E. Rivals, D. D. de Segonzac, C. Garres, P. Caminade, B. De Massy and P. Boursot (2014). "Diversity of Prdm9 zinc finger array in wild mice unravels new facets of the evolutionary turnover of this coding minisatellite." PLoS One 9(1): e85021.

Clark, A. G., X. Wang and T. Matise (2010). "Contrasting methods of quantifying fine structure of human recombination." Annual review of genomics and human genetics 11: 45.

Coop, G. and M. Przeworski (2006). "An evolutionary view of human recombination." Nature Reviews Genetics 8(1): 23-34.

Davies, B., E. Hatton, N. Altemose, J. G. Hussin, F. Pratto, G. Zhang, A. G. Hinch, D. Moralli, D. Biggs and R. Diaz (2016). "Re-engineering the zinc fingers of PRDM9 reverses hybrid sterility in mice." Nature 530(7589): 171-176.

Fazakerley, D. J., S. Naghiloo, R. Chaudhuri, F. Koumanov, J. G. Burchfield, K. C. Thomas, J. R. Krycer, M. J. Prior, B. L. Parker and B. A. Murrow (2015). "Proteomic analysis of GLUT4 storage vesicles reveals tumor suppressor candidate 5 (TUSC5) as a novel regulator of insulin action in adipocytes." Journal of biological chemistry 290(39): 23528-23542.

Grey, C., P. Barthès, G. Chauveau-Le Friec, F. Langa, F. Baudat and B. De Massy (2011). "Mouse PRDM9 DNA-binding specificity determines sites of histone H3 lysine 4 trimethylation for initiation of meiotic recombination." PLoS Biol 9(10): e1001176.

Hubbard, T., D. Barker, E. Birney, G. Cameron, Y. Chen, L. Clark, T. Cox, J. Cuff, V. Curwen and T. Down (2002). "The Ensembl genome database project." Nucleic acids research 30(1): 38-41.

Hubert, R., M. MacDonald, J. Gusella and N. Arnheim (1994). "High resolution localization of recombination hot spots using sperm typing." Nature genetics 7(3): 420-424.

Hunter, C. M., W. Huang, T. F. Mackay and N. D. Singh (2016). "The genetic architecture of natural variation in recombination rate in Drosophila melanogaster." PLoS Genet 12(4): e1005951.

Jensen-Seaman, M. I., T. S. Furey, B. A. Payseur, Y. Lu, K. M. Roskin, C.-F. Chen, M. A. Thomas, D. Haussler and H. J. Jacob (2004). "Comparative recombination rates in the rat, mouse, and human genomes." Genome research 14(4): 528-538.

Kadri, N. K., C. Harland, P. Faux, N. Cambisano, L. Karim, W. Coppieters, S. Fritz, E. Mullaart, D. Baurain, D. Boichard, R. Spelman, C. Charlier, M. Georges and T. Druet (2016). "Coding and noncoding variants in HFM1, MLH3, MSH4, MSH5, RNF212, and RNF212B affect recombination rate in cattle." Genome Res 26(10): 1323-1332.

Kauppi, L., A. J. Jeffreys and S. Keeney (2004). "Where the crossovers are: recombination distributions in mammals." Nature Reviews Genetics 5(6): 413-424.

Kong, A., G. Thorleifsson, D. F. Gudbjartsson, G. Masson, A. Sigurdsson, A. Jonasdottir, G. B. Walters, A. Jonasdottir, A. Gylfason and K. T. Kristinsson (2010). "Fine-scale recombination rate differences between sexes, populations and individuals." Nature 467(7319): 1099-1103.

Kono, H., M. Tamura, N. Osada, H. Suzuki, K. Abe, K. Moriwaki, K. Ohta and T. Shiroishi (2014). "Prdm9 polymorphism unveils mouse evolutionary tracks." DNA Research 21(3): 315-326.

Lesecque, Y., S. Glémin, N. Lartillot, D. Mouchiroud and L. Duret (2014). "The Red Queen model of recombination hotspots evolution in the light of archaic and modern human genomes."

Loftus, R. T., D. E. MacHugh, D. G. Bradley, P. M. Sharp and P. Cunningham (1994). "Evidence for two independent domestications of cattle." Proceedings of the National Academy of Sciences 91(7): 2757-2761.

Lu, S., C. Zong, W. Fan, M. Yang, J. Li, A. R. Chapman, P. Zhu, X. Hu, L. Xu and L. Yan (2012). "Probing meiotic recombination and aneuploidy of single sperm cells by whole-genome sequencing." Science 338(6114): 1627-1630.

Ma, L., J. R. O'Connell, P. M. VanRaden, B. Shen, A. Padhi, C. Sun, D. M. Bickhart, J. B. Cole, D. J. Null, G. E. Liu, Y. Da and G. R. Wiggans (2015). "Cattle Sex-Specific Recombination and Genetic Control from a Large Pedigree Analysis." PLoS Genet 11(11): e1005387.

Myers, S., R. Bowden, A. Tumian, R. E. Bontrop, C. Freeman, T. S. MacFie, G. McVean and P. Donnelly (2010). "Drive against hotspot motifs in primates implicates the PRDM9 gene in meiotic recombination." Science 327(5967): 876-879.

Nachman, M. W. and B. A. Payseur (2012). "Recombination rate variation and speciation: theoretical predictions and empirical results from rabbits and mice." Phil. Trans. R. Soc. B 367(1587): 409-421.

Oliver, P. L., L. Goodstadt, J. J. Bayes, Z. Birtle, K. C. Roach, N. Phadnis, S. A. Beatson, G. Lunter, H. S. Malik and C. P. Ponting (2009). "Accelerated evolution of

the Prdm9 speciation gene across diverse metazoan taxa." PLoS Genet 5(12): e1000753.

Padhi, A., B. Shen, J. Jiang, Y. Zhou, G. E. Liu and L. Ma (2017). "Ruminant-specific multiple duplication events of PRDM9 before speciation." BMC evolutionary biology 17(1): 79.

Paigen, K. and P. Petkov (2010). "Mammalian recombination hot spots: properties, control and evolution." Nature Reviews Genetics 11(3): 221-233.

Paigen, K., J. P. Szatkiewicz, K. Sawyer, N. Leahy, E. D. Parvanov, S. H. Ng, J. H. Graber, K. W. Broman and P. M. Petkov (2008). "The recombinational anatomy of a mouse chromosome." PLoS genetics 4(7): e1000119.

Parvanov, E. D., P. M. Petkov and K. Paigen (2010). "Prdm9 controls activation of mammalian recombination hotspots." Science 327(5967): 835-835.

Payseur, B. A. (2016). "Genetic Links between Recombination and Speciation." PLoS Genet 12(6): e1006066.

Persikov, A. V. and M. Singh (2013). "De novo prediction of DNA-binding specificities for Cys2His2 zinc finger proteins." Nucleic acids research: gkt890.

Ponting, C. P. (2011). "What are the genomic drivers of the rapid evolution of PRDM9?" Trends in Genetics 27(5): 165-171.

Pratto, F., K. Brick, P. Khil, F. Smagulova, G. V. Petukhova and R. D. Camerini-Otero (2014). "Recombination initiation maps of individual human genomes." Science 346(6211): 1256442.

R, C. T. (2014). "R: A Language and Environment for Statistical Computing. Vienna, Austria: R Foundation for Statistical Computing."

Sandor, C., W. Li, W. Coppieters, T. Druet, C. Charlier and M. Georges (2012). "Genetic variants in REC8, RNF212, and PRDM9 influence male recombination in cattle." PLoS Genet 8(7): e1002854.

Schwartz, J. J., D. J. Roach, J. H. Thomas and J. Shendure (2014). "Primate evolution of the recombination regulator PRDM9." Nature communications 5.

Ségurel, L., E. M. Leffler and M. Przeworski (2011). "The case of the fickle fingers: how the PRDM9 zinc finger protein specifies meiotic recombination hotspots in humans." PLoS Biol 9(12): e1001211.

Shifman, S., J. T. Bell, R. R. Copley, M. S. Taylor, R. W. Williams, R. Mott and J. Flint (2006). "A high-resolution single nucleotide polymorphism genetic map of the mouse genome." PLoS biology 4(12): e395.

Smagulova, F., K. Brick, Y. Pu, R. D. Camerini-Otero and G. V. Petukhova (2016). "The evolutionary turnover of recombination hot spots contributes to speciation in mice." Genes & development 30(3): 266-280.

Stumpf, M. P. and G. A. McVean (2003). "Estimating recombination rates from population-genetic data." Nature Reviews Genetics 4(12): 959-968.

Wang, J., H. C. Fan, B. Behr and S. R. Quake (2012). "Genome-wide single-cell analysis of recombination activity and de novo mutation rates in human sperm." Cell 150(2): 402-412.

Ye, J., G. Coulouris, I. Zaretskaya, I. Cutcutache, S. Rozen and T. L. Madden (2012). "Primer-BLAST: a tool to design target-specific primers for polymerase chain reaction." BMC bioinformatics 13(1): 1.

Zimin, A. V., A. L. Delcher, L. Florea, D. R. Kelley, M. C. Schatz, D. Puiu, F. Hanrahan, G. Pertea, C. P. Van Tassell and T. S. Sonstegard (2009). "A whole-genome assembly of the domestic cow, Bos taurus." Genome biology 10(4): 1.

Tables

Table 2.1 Most significant allele 5 specific hotspot regions for males and females in Holstein cattle. Start, End = physical position of hotspot regions on chromosome. M0, M1 and M2 = recombination rates of males with non-allele 5, allele 5 heterozygote and allele 5 homozygote genotypes, respectively. F0, F1 and F2 = female recombination rates. P-value 1= allele 5 homozygote versus non-allele 5 genotype. P-value 2 = allele 5 heterozygote versus non-allele 5 genotype. Top 5 rows are from male results and bottom 5 rows are female hotspots.

Chr	Start	End	Recombination Rate						Comparison	
			M0	M1	M2	F0	F1	F2	*P*-value1	*P*-value2
23	40203865	40236175	0.00061	0.00200	0.01006	0.00061	0.00089	0.00076	1.9×10^{-17}	9.8×10^{-5}
13	14923596	15017558	0.00067	0.00281	0.00905	0.00078	0.00314	0.00383	9.4×10^{-13}	1.2×10^{-2}
4	21160109	21179159	0.00047	0.00050	0.00733	0.00056	0.00062	0.00082	1.0×10^{-11}	3.1×10^{-09}
1	152566977	152592454	0.00126	0.00252	0.00975	0.00023	0.00030	0.00017	7.6×10^{-8}	1.9×10^{-3}
18	55956772	55983042	0.00041	0.00073	0.00544	0.00048	0.00044	0.00023	3.6×10^{-7}	7.1×10^{-4}
13	14923596	15017558	0.00067	0.00281	0.00905	0.00078	0.00314	0.00383	2.7×10^{-2}	2.4×10^{-14}
18	7014208	7041735	0.00048	0.00220	0.00114	0.00032	0.00157	0.00308	3.4×10^{-3}	4.2×10^{-9}
23	40366428	40416708	0.00062	0.00193	0.00079	0.00059	0.00192	0.00346	2.0×10^{-2}	4.1×10^{-7}
11	24175823	24193422	0.00025	0.00056	0.00031	0.00029	0.00084	0.00441	5.5×10^{-7}	3.2×10^{-3}
7	34678536	34709628	0.00062	0.00091	0.00075	0.00074	0.00211	0.00349	4.8×10^{-2}	1.5×10^{-6}

Table 2.2 Most significant non-allele 5 hotspot regions for males and females in Holstein cattle. Start, End = physical position of hotspot regions on chromosome. M0, M1 and M2 = recombination rates of males with non-allele 5, allele 5 heterozygote and allele 5 homozygote genotypes, respectively. F0, F1 and F2 = female recombination rates. P-value = allele 5 heterozygote versus non-allele 5 genotype. Top 5 rows are from male results and bottom 5 rows are female hotspots.

Chr	Start	End	Recombination Rate						Comparison
			M0	M1	M2	F0	F1	F2	*P*-value
21	67159217	67195140	0.00196	0.00072	0.00011	0.00084	0.00038	0.00035	7.4×10^{-6}
19	54920324	54947863	0.00216	0.00069	0.00059	0.00072	0.00049	0.00063	3.4×10^{-7}
4	18380952	18447308	0.002	0.00108	0.00057	0.0016	0.00122	0.00095	1.3×10^{-3}
18	6246055	6275418	0.00155	0.00069	0.00020	0.00092	0.00030	0.00034	6.4×10^{-4}
2	12415870	12451708	0.00126	0.00068	9.45E-05	0.00072	0.00057	0.00014	1.2×10^{-2}
22	33218085	33245588	0.00048	0.00025	0.00027	0.00171	0.00072	0.00038	6.1×10^{-3}
5	8435987	8476871	0.00079	0.00036	0.00040	0.00119	0.00034	5.28E-05	4.5×10^{-3}
7	34269024	34324708	0.00084	0.00047	0.00075	0.00129	0.00062	0.00018	3.9×10^{-2}
26	26340616	26369697	0.00054	0.00033	0.00038	0.00159	0.00064	0.00051	6.3×10^{-3}
10	7791755	7830003	0.00148	0.00072	0.00043	0.0017	0.00083	0.00068	1.6×10^{-2}

Figures

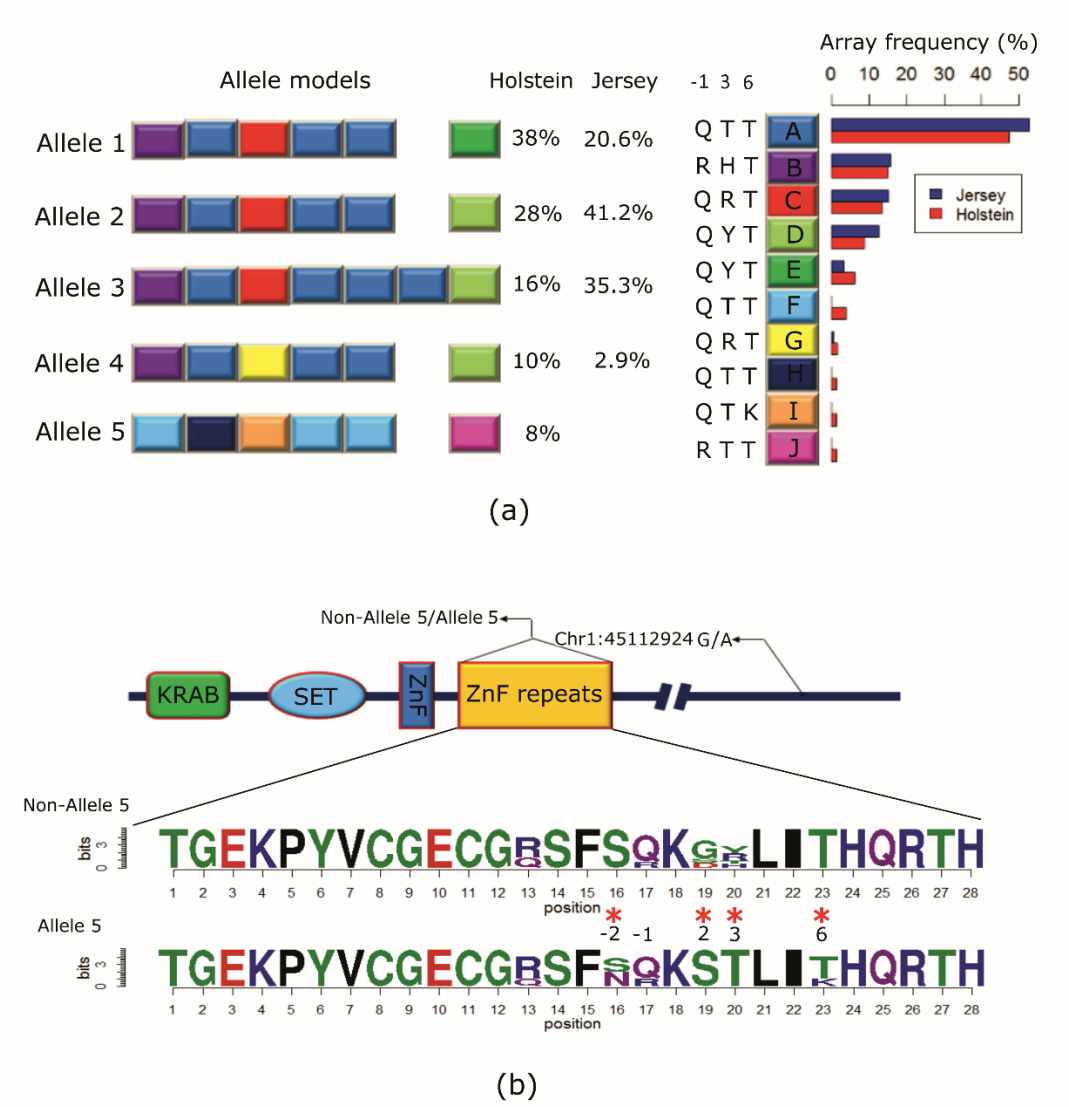


Figure 2.1 PRDM9 alleles and ZnF arrays in Holstein and Jersey cattle. (a) Allele and ZnF array information in Holstein and Jersey cattle. ZnF arrays were coded from A to J according to the full length amino acid composition, and the three important sites were shown (–1, 3 and 6) that may be same for different ZnF arrays. Alleles were coded from allele 1 to allele 5 based on their ZnF array composition. (b) PRDM9 gene structure and comparison between non-allele 5 and allele 5. Four domains (KRAB, SET, single ZnF, 6 to 7 ZnF repeats) were present in bovine PRDM9. Allele 5 was perfectly linked with the minor allele of the SNP at Chr1:45112924 (R2 = 1). Comparing between non-allele 5 and allele 5, amino acid frequencies at each positions were shown as proportional to the size of the letters (one ZnF includes 28 amino acids). Positions with different amino acid components between allele 5 and other alleles were highlighted by "*".

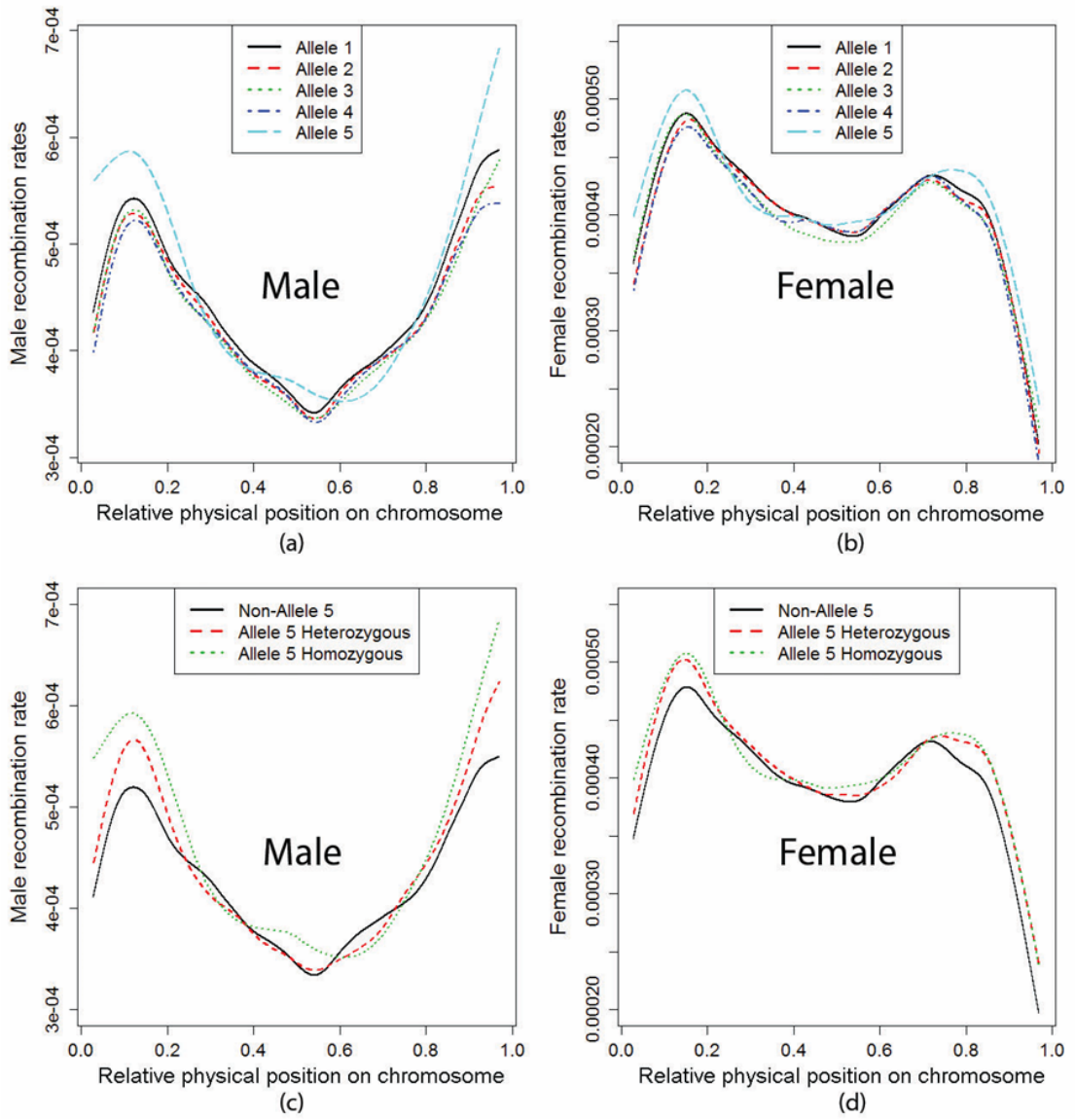


Figure 2.2 PRDM9 allele-specific distribution of recombination rate along a chromosome in males and females in Holstein. (a) Recombination patterns of five alleles in males. (b) Recombination patterns of five alleles in females. (c) Recombination patterns of three genotypes of allele 5 in males. (d) Recombination patterns of three genotypes of allele 5 in females. The relative physical position on a chromosome is used, where zero corresponding to the beginning of a chromosome and one the end. The smooth spline model was fitted across all of the 29 autosomes.

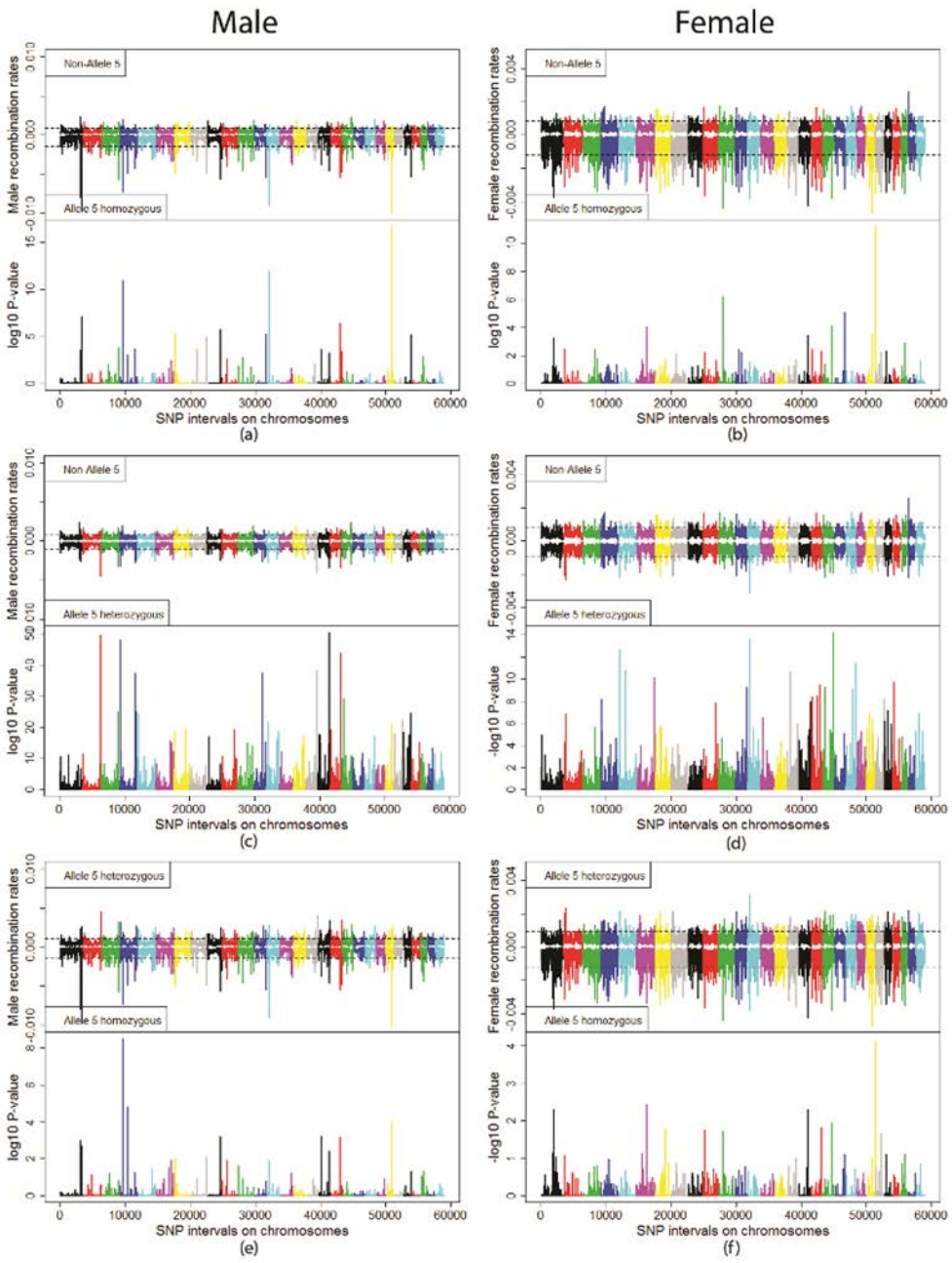


Figure 2.3 PRDM9 allele 5 dependent recombination hotspots in two sexes. (a) Non-allele 5 v.s. allele 5 homozygote in males. (b) Non-allele 5 v.s. allele 5 homozygote in females. (c) Non-allele 5 v.s. allele 5 heterozygote in males. (d) Non-allele 5 v.s. allele 5 heterozygote in females. (e) Allele 5 heterozygote v.s. allele 5 homozygote in males. (f) Allele 5 heterozygote v.s. allele 5 homozygote in females. For each panel, recombination rates in each SNP intervals of two groups were shown in the top half and corresponding P-values were shown in the bottom half. Different colors were used to distinguish the 29 chromosomes.

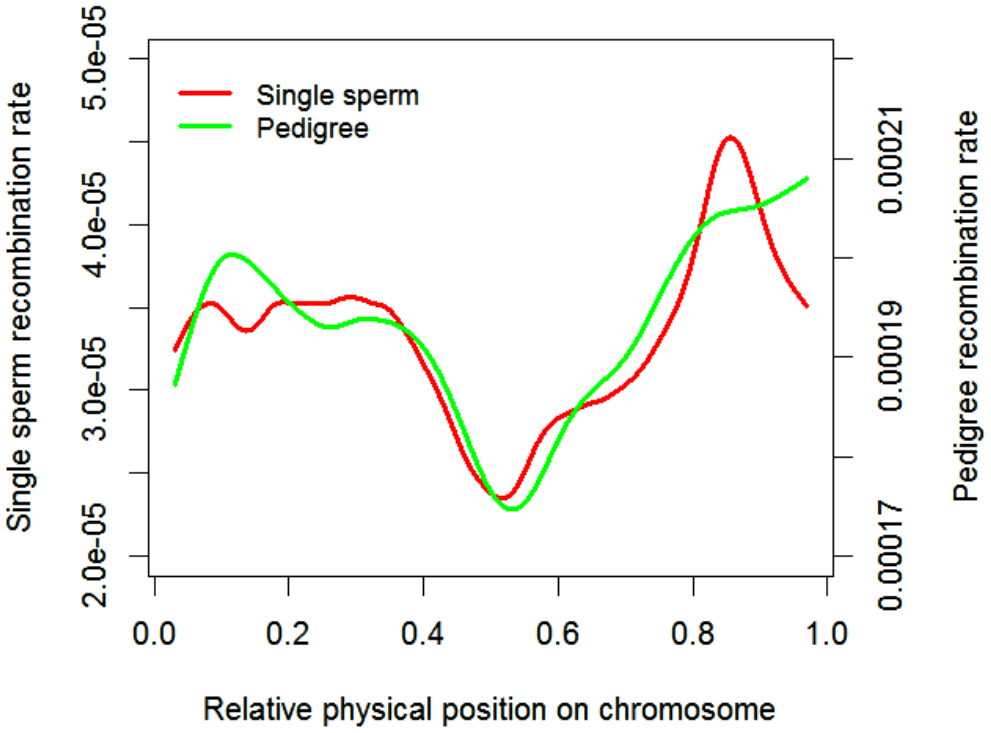


Figure 2.4 Spline-smoother plot of recombination rate along the chromosome from single sperm data and pedigree data. The relative physical position on a chromosome is used, where zero corresponding to the beginning of a chromosome and one the end. The smooth spline model was fitted across all of the 29 chromosomes. Due to the differences in SNP density, the single sperm and pedigree recombination rates were plotted in different scales.

Supplemental Materials

Supplemental Tables:

Supplementary Table 2.1 Primer information for bovine *PRDM9* ZnF amplification.

Primer aim	Primer name	Primer sequence
Primers for the two ZnF repeat regions	PRDM9_1_F	ATCGTGCACTGTCAACACCT
	PRDM9_1_R	CATGATTCACTCACACTCCCTG
	PRDM9_2_F	AGCCATCAGTTACTTGCCATCC
	PRDM9_2_R	TTGAAGGTATGGGTGGTGAATCT
Primers to amplify NCBI and Ensembl *PRDM9* gene ZnF repeat regions from testis tissue	PRDM9_PC_F	AAAGCAGCAGGAAAAGCGAG
	PRDM9_PC_R	CACGACCCTCAGGTTTGTCA
	PRDM9_NW_R	GGTGGCTGCAGTAATTCTCCT
	PRDM9_En_R	TTGAAGGTATGGGTGGTGAATCT
Primers for identifying *PRDM9* ZnF sequences	PRDM9_F	ATCGTGCACTGTCAACACCT
	PRDM9_R	TCCTCACACTCTTGGTGACTGTA

Supplementary Table 2.2 Clone and *PRDM9* allele information for Jesery cattle.

Animal ID	#Clones	Allele Combination		#Offspring	#Offspring 50K	Avg CO#	Hotspot Usage
JEUSA000114581918	9	Allele 3	Allele 4	1235	41	23.34	0.0491
JEUSA000117217618	7	Allele 2	Allele 3	1023	129	24.78	0.0439
JEUSA000115181456	7	Allele 1	Allele 3	1044	67	27.09	0.0432
JEUSA000114845461	6	Allele 3	Allele 3	1479	175	23.6	0.0412
JEUSA000067080468	9	Allele 1	Allele 2	1238	49	25.8	0.0486
JEUSA000114114336	8	Allele 2	Allele 3	1017	174	23.06	0.0415
JEUSA000117423084	8	Allele 2	Allele 2	1039	74	23.81	0.0497
JEUSA000117432987	10	Allele 2	Allele 2	1362	52	24	0.0458
JEUSA000117422971	6	Allele 2	Allele 3	1773	162	20.59	0.042
JEUSA000066857901	7	Allele 1	Allele 2	1190	86	24.78	0.0491
JEUSA000067129272	9	Allele 2	Allele 3	3160	180	25.08	0.0443
JEUSA000067107510	6	Allele 3	Allele 3	1332	147	24.37	0.0466
JEDNK000000301592	12	Allele 1	Allele 3	1252	211	24.93	0.0354
JEUSA000115883929	9	Allele 1	Allele 2	1759	277	25.59	0.0492
JEUSA000067029404	10	Allele 2	Allele 3	1389	10	24.9	0.0411
JEUSA000067332021	8	Allele 1	Allele 1	2390	127	24.17	0.0506
JEUSA000117222740	6	Allele 2	Allele 2	1040	141	27.6	0.046

Supplementary Table 2.3 *PRDM9* alleles tagged by Haplotypes and SNPs in Holstein. Colors are used to highlight the tagging markers for each of the five *PRDM9* alleles.

PRDM9 Genotype			Allele 1 Tagging		Allele 2 Tagging		Allele 3 Tagging		Allele 4 Tagging		Allele 5 Tagging
Animal ID	PRDM9 Allele		Hap 1	Hap 2	Hap 1	Hap 2	Hap 1	Hap 2	Hap 1	Hap 2	rs110661033
HOUSA00006189830 6	Allele 1	Allele 3	2	1	2	1	2	1	2	1	0
HOUSA00007219076 7	Allele 1	Allele 2	4	2	4	2	4	2	4	2	0
HOUSA00013680023 3	Allele 1	Allele 1	1	3	1	3	1	3	1	3	0
HOUSA00006663665 7	Allele 3	Allele 5	1	6	1	6	1	6	1	6	1
HOUSA00006873181 0	Allele 1	Allele 1	11	19	11	19	11	19	11	19	0
HOUSA00006099695 6	Allele 1	Allele 2	2	21	2	21	2	21	2	21	0
HOUSA00005280572 3	Allele 2	Allele 2	58	5	58	5	58	5	58	5	0
HOUSA00014014555 3	Allele 4	Allele 5	13	6	13	6	13	6	13	6	1
HOCAN000010705608	Allele 4	Allele 4	7	13	7	13	7	13	7	13	0
HOUSA00013443823 0	Allele 1	Allele 2	4	19	4	19	4	19	4	19	0
HOUSA00006487295 1	Allele 2	Allele 2	29	18	29	18	29	18	29	18	0
HOUSA00006999013 8	Allele 1	Allele 3	1	30	1	30	1	30	1	30	0
HOUSA00006154747 6	Allele 1	Allele 2	15	2	15	2	15	2	15	2	0
HO840003006972816	Allele 2	Allele 3	12	16	12	16	12	16	12	16	0
HOUSA00006206775 3	Allele 1	Allele 1	2	2	2	2	2	2	2	2	0
HOUSA00006591748 1	Allele 1	Allele 3	1	28	1	28	1	28	1	28	0
HOUSA00013574677 6	Allele 1	Allele 5	2	17	2	17	2	17	2	17	1
HOUSA00006496673 9	Allele 1	Allele 1	4	2	4	2	4	2	4	2	0
HOUSA00006897712 0	Allele 2	Allele 2	8	4	8	4	8	4	8	4	0
HOUSA00006113383 7	Allele 4	Allele 5	13	6	13	6	13	6	13	6	1
HOUSA00006276899 0	Allele 1	Allele 4	2	31	2	31	2	31	2	31	0
HOUSA00006998134 9	Allele 1	Allele 2	4	2	4	2	4	2	4	2	0
HOUSA00006206591 9	Allele 1	Allele 2	18	23	18	23	18	23	18	23	0
HOUSA00013297394 2	Allele 3	Allele 3	1	1	1	1	1	1	1	1	0
HOCAN000007816429	Allele 2	Allele 3	4	1	4	1	4	1	4	1	0

Note that PRDM9 alleles are tagged by the 36th haplotype segment. PRDM9 allele 1 is tagged by haplotype #2, allele 2 is tagged by haplotype #4, allele 3 is tagged by haplotype #1, allele 4 is tagged by haplotype #13, and allele 5 is tagged by SNP rs110661033.

Supplementary Table 2.5 Motif count ratio (allele 5 motif / allele 1 motif) in various Holstein recombination hotspots. Note the baseline ratio is 1.17 based on the cattle reference genome (UMD 3.1.1).

Sex	Non-Allele 5 specific	Non-Allele 5	Non-Allele5 and Allele 5 Heterozygous shared	Allele 5 Heterozygous	Genome-wide ratio	Allele5 Homozygous and Allele 5 Heterozygous shared	Allele 5 Homozygous	Allele 5 specific
Male	1.14	1.14	1.14	1.16	1.17	1.19	1.2	1.21
Female	1.15	1.15	1.17	1.17	1.17	1.19	1.18	1.18

Supplementary Table 2.6 Primer information for bovine *PRDM9* ZnF amplification.

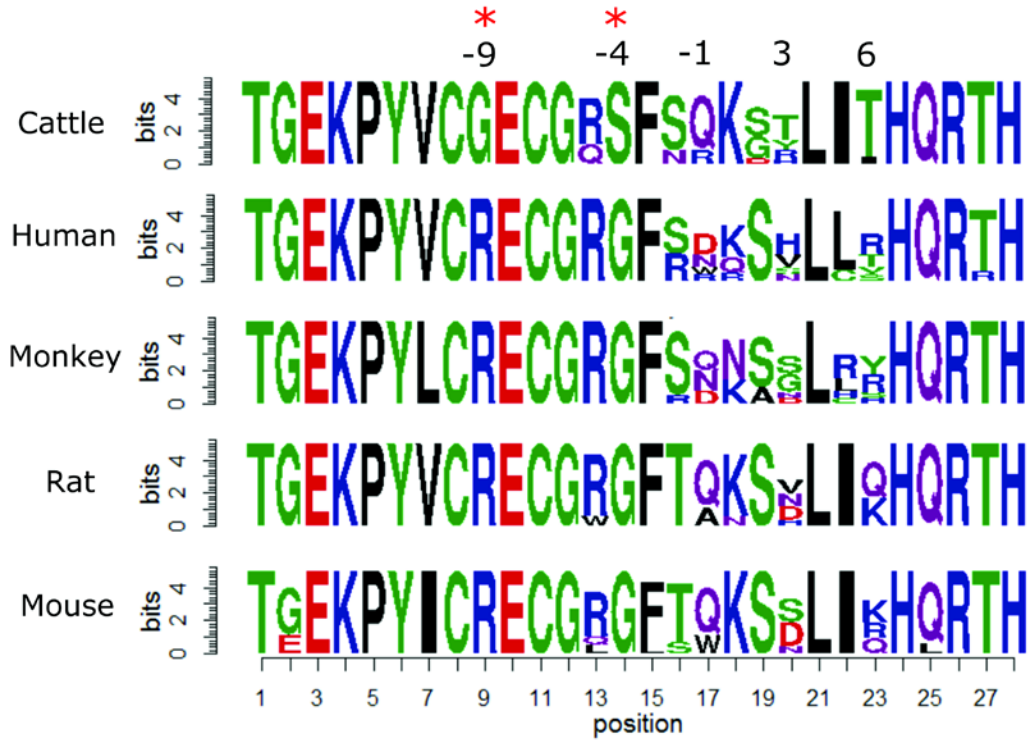
Primer aim	Primer name	Primer sequence
Primers for the two ZnF repeat regions	PRDM9_1_F	ATCGTGCACTGTCAACACCT
	PRDM9_1_R	CATGATTCACTCACACTCCCTG
	PRDM9_2_F	AGCCATCAGTTACTTGCCATCC
	PRDM9_2_R	TTGAAGGTATGGGTGGTGAATCT
Primers to amplify NCBI and Ensembl *PRDM9* gene ZnF repeat regions from testis tissue	PRDM9_PC_F	AAAGCAGCAGGAAAAGCGAG
	PRDM9_PC_R	CACGACCCTCAGGTTTGTCA
	PRDM9_NW_R	GGTGGCTGCAGTAATTCTCCT
	PRDM9_En_R	TTGAAGGTATGGGTGGTGAATCT
Primers for identifying *PRDM9* ZnF sequences	PRDM9_F	ATCGTGCACTGTCAACACCT
	PRDM9_R	TCCTCACACTCTTGGTGACTGTA

Supplementary Table 2.7 SNP data quality and recombination number for each single sperm.

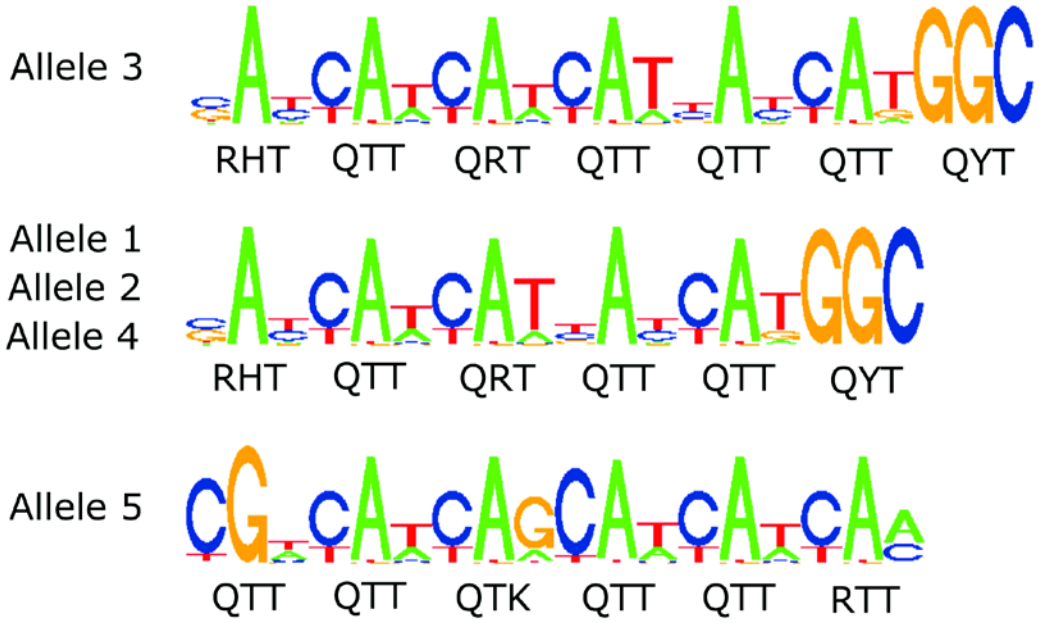
Sample ID	Recombination number	Call rate	Homozygote rate	SNP number for recombination analysis
S1	25	0.4444	0.9954	70,258
S2	19	0.4227	0.9966	66,198
S3	27	0.4141	0.9967	64,853
S4	35	0.4343	0.9966	64,491
S5	25	0.4114	0.9962	63,736
S6	27	0.4019	0.9953	63,640
S7	29	0.4213	0.9952	62,948
S8	27	0.4198	0.9957	62,583
S9	25	0.3916	0.9780	60,118
S10	23	0.3862	0.9955	57,520
S11	29	0.3972	0.9826	57,325
S12	34	0.3628	0.9956	56,755
S13	26	0.3573	0.9730	56,008
S14	28	0.3694	0.9643	55,704
S15	22	0.3555	0.9946	55,661
S16	23	0.3828	0.9730	55,618
S17	23	0.3717	0.9951	55,266
S18	36	0.3694	0.9945	55,000
S19	22	0.3527	0.9963	54,964
S20	25	0.3687	0.9857	54,462
S21	26	0.3762	0.9645	54,435
S22	21	0.3438	0.9943	53,876
S23	35	0.3440	0.9921	53,715
S24	24	0.3681	0.9731	53,379
S25	30	0.3376	0.9949	53,023
S26	29	0.3417	0.9939	52,795
S27	24	0.3650	0.9967	52,452
S28	19	0.3319	0.9969	51,796
S29	21	0.3386	0.9621	50,480
S30	30	0.3382	0.9958	50,288
S31	24	0.3162	0.9969	49,840
S32	45	0.4607	0.9575	49,346
S33	22	0.3134	0.9966	49,219
S34	21	0.3289	0.9949	49,082
S35	30	0.3289	0.9948	49,035
S36	23	0.3136	0.9938	48,818
S37	40	0.3136	0.9949	48,061
S38	34	0.3183	0.9973	47,325
S39	21	0.3174	0.9973	47,187

S40	27	0.3062	0.9686	47,152
S41	14	0.2946	0.9969	46,112
S42	31	0.3104	0.9887	45,110
S43	28	0.2802	0.9968	43,048
S44	25	0.2688	0.9942	42,186
S45	21	0.2674	0.9964	41,897
S46	29	0.2822	0.9622	40,580
S47	33	0.2576	0.9950	40,223
S48	17	0.2829	0.8623	38,929
S49	30	0.2490	0.9970	38,373
S50	31	0.2376	0.9934	36,486
S51	29	0.2220	0.9911	32,759
S52	29	0.2088	0.9938	30,812
S53	42	0.2079	0.9299	29,242
S54	27	0.1838	0.9719	27,366
S55	29	0.1421	0.9790	19,206
S56	35	0.5478	0.8610	11,762

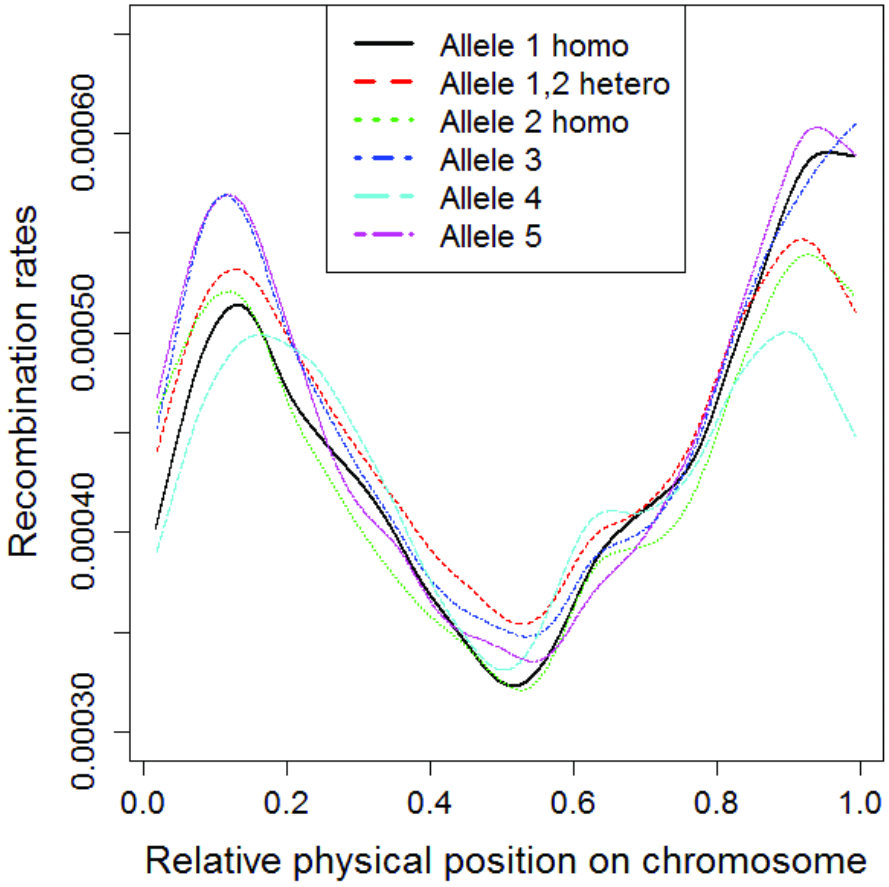
Supplemental Figures:



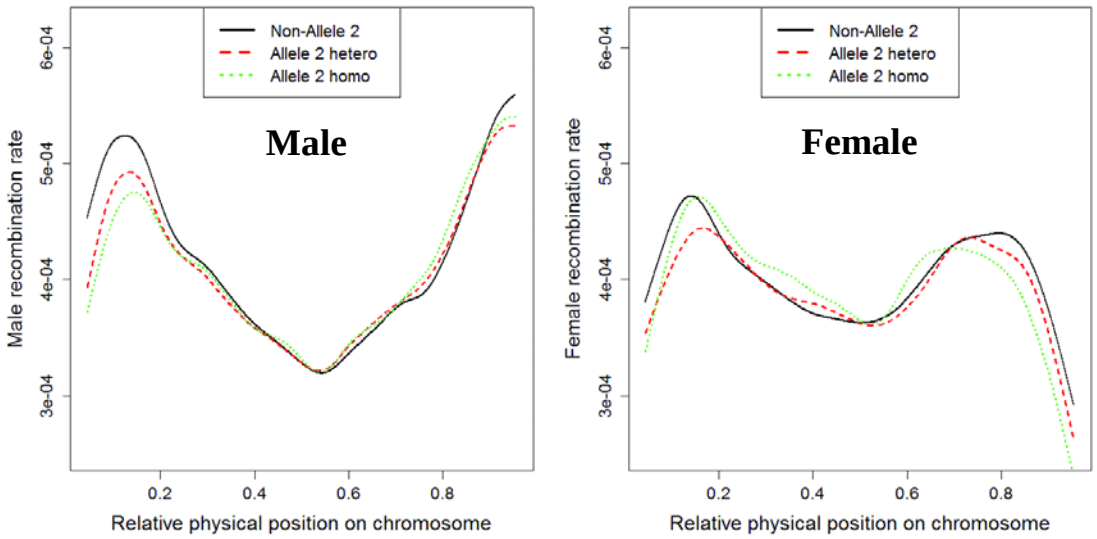
Supplementary Figure 2.1 *PRDM9* ZnF amino acid components of five species. *PRDM9* ZnF information of human (NP_064612.2), monkey(XP_001083675.2), rat (NP_001102373.2) and mouse (XP_011244657.1, NP_659058.2, XP_006524051.1, XP_006524052.1, XP_017172869.1 and XP_006524053.1) were downloaded from the NCBI database. Unique ZnF arrays from different species were used to calculate amino acid frequencies at each position (one ZnF has 28 amino acids). Coordinates with different amino acid components between cattle and other species were highlighted using "*".



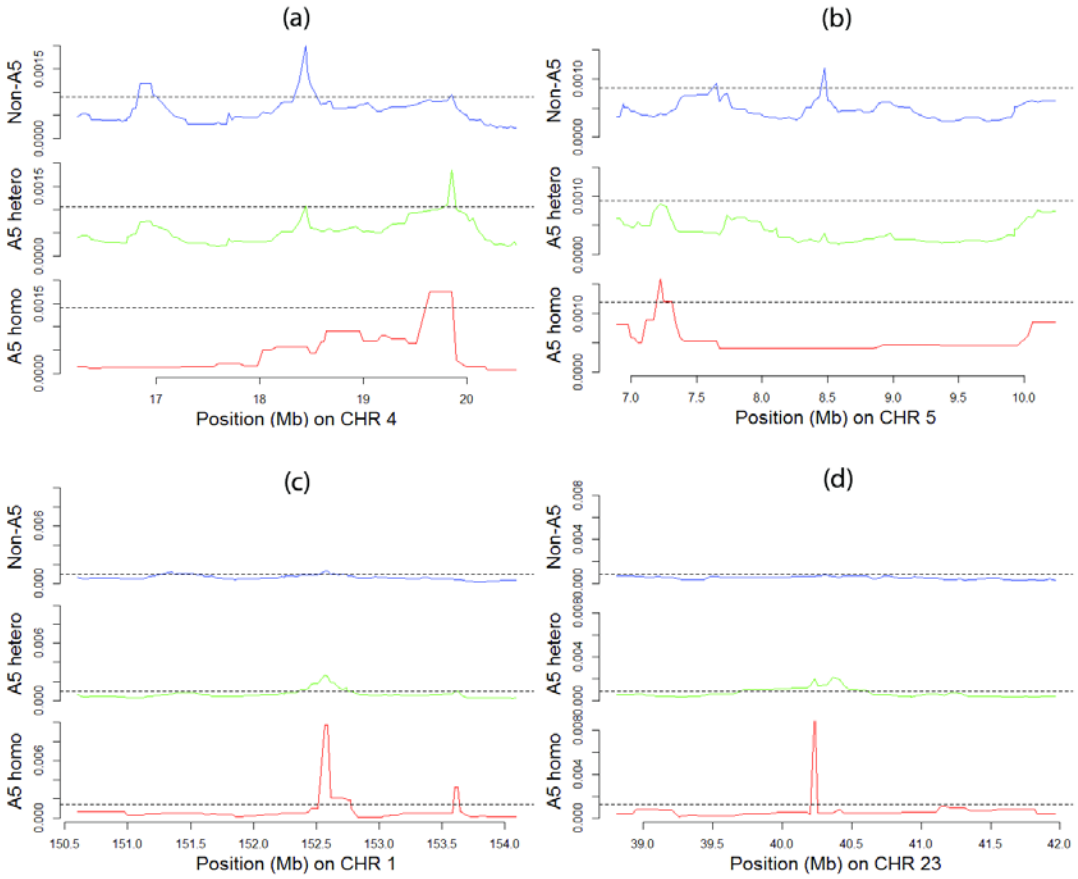
Supplementary Figure 2.2 Predicted DNA binding motifs for five *PRDM9* ZnF alleles. Allele 1, allele 2 and allele 4 had the same predicted motif. Allele 3 has one more ZnF array A, which lead to longer but similar binding motif as allele 1, allele 2 and allele 4. Allele 5 showed the most distinct binding motif compared with other alleles.



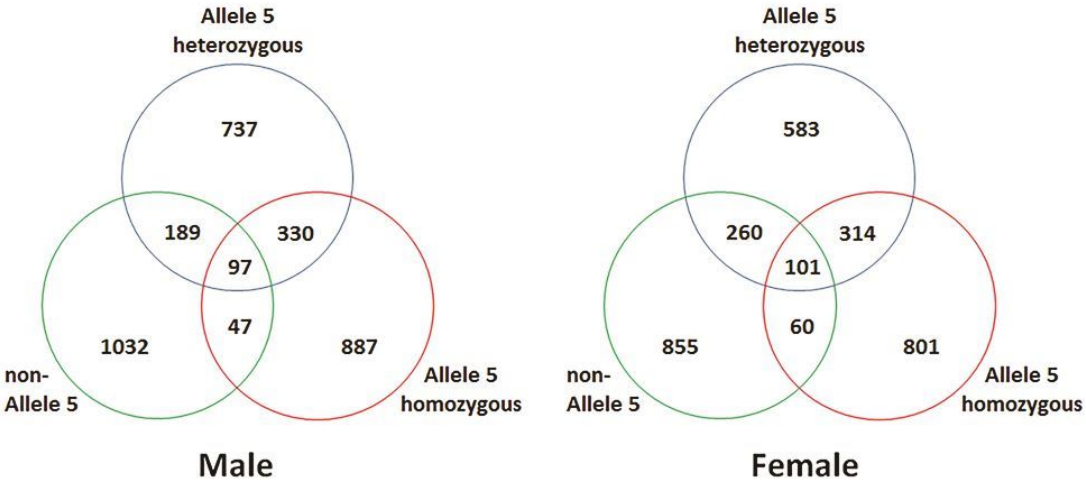
Supplementary Figure 2.3 *PRDM9* allele-specific recombination patterns along a chromosome in 25 Holstein bulls. The 25 Holstein bulls were divided to 6 groups based on their *PRDM9* genotypes: allele 1 homozygote, allele 2 homozygote, allele 1/2 heterozygote, allele 3 carrier (homozygote or heterozygote), allele 4 carrier (homozygote or heterozygote), and allele 5 carrier (heterozygote). The relative physical position on a chromosome is used, where zero corresponding to the beginning of a chromosome and one the end. The smooth spline model was fitted across all of the 29 autosomes



Supplementary Figure 2.4 *PRDM9* allele-specific recombination patterns along a chromosome in Jersey. The relative physical position on a chromosome is used, where zero corresponding to the beginning of a chromosome and one the end. The smooth spline model was fitted across all of the 29 autosomes.



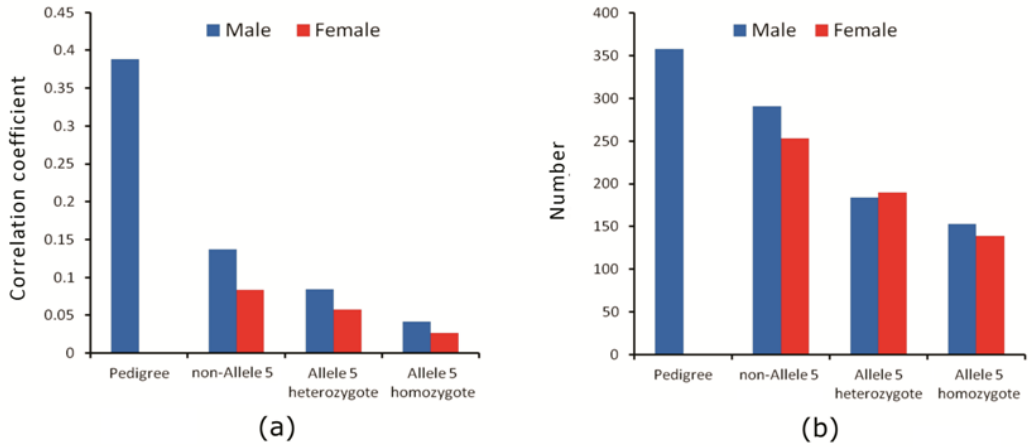
Supplementary Figure 2.5 Examples of *PRDM9* allele-specific regional recombination maps in Holstein. (**a**) Non-allele 5 hotspots in males. (**b**) Non-allele 5 hotspots in females. (**c**) Allele 5 hotspots in males. (**d**) Allele 5 hotspots in females.



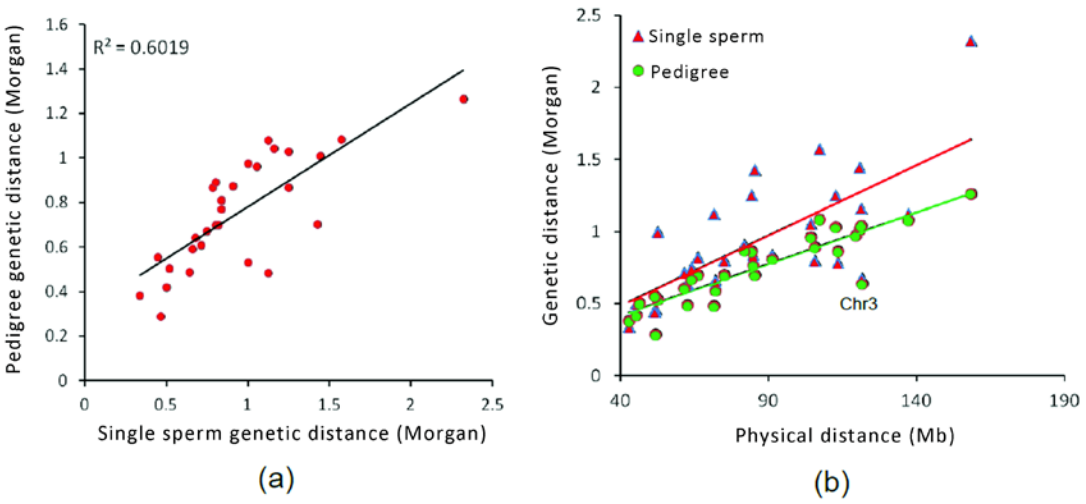
Supplementary Figure 2.6 Hotspots sharing between three *PRDM9* genotypes of allele 5 in Holstein bulls and cows. Recombination hotspots were tentatively defined as the SNP intervals with recombination rate 2.5 standard deviations greater than the genome-wide average.

Male / Female	Non-Allele 5	Allele 5 Hetero	Allele 5 Homo
Non-Allele 5	0.2498	0.2006	0.1114
Allele 5 Hetero	0.2678	0.3102	0.2927
Allele 5 Homo	0.1045	0.3107	0.1083

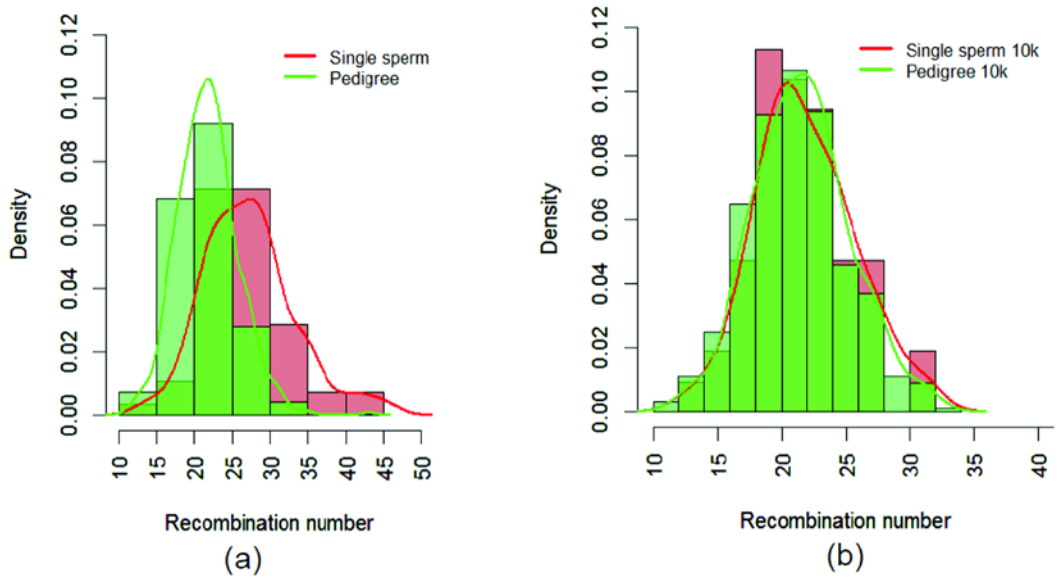
Supplementary Figure 2.7 Correlation matrix of recombination rate by *PRDM9* genotype and sex in Holstein. Top right = males. Bottom left = females. Diagonal = within genotype group and between two sexes.



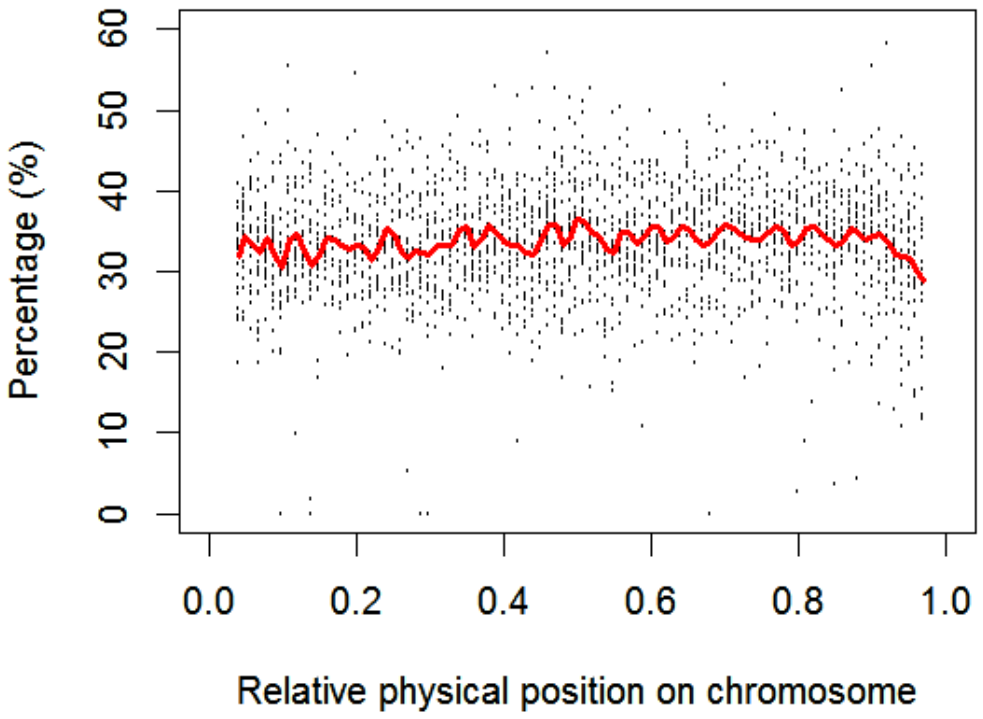
Supplementary Figure 2.8 Comparison of recombination between single sperm and pedigree data. (**a**) Correlation in recombination rate between single sperm and pedigree data. (**b**) Overlap number of top 5% recombination intervals. Pedigree = the same bull (allele 1/allele 2) as the sperm data.



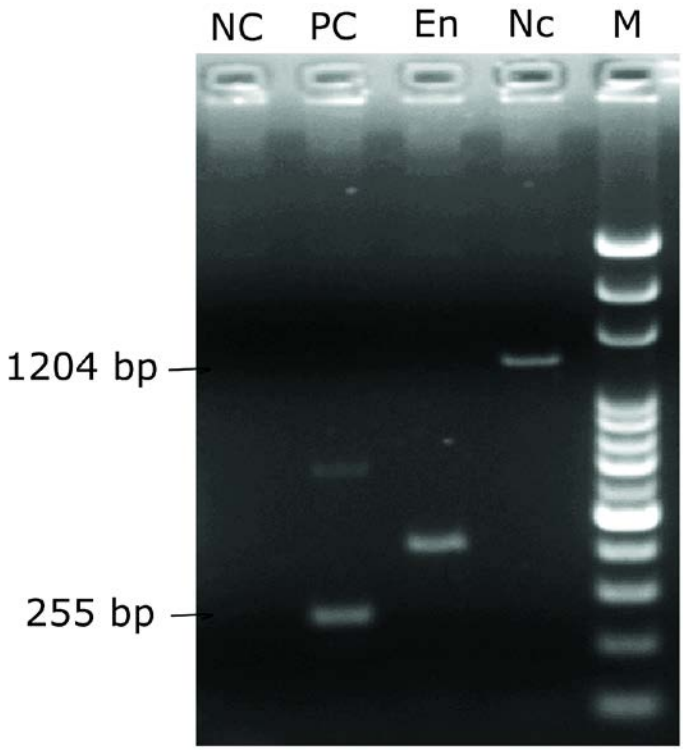
Supplementary Figure 2.9 Comparison between recombination of singe sperm and pedigree data in genetic and physical distances. (**a**) Linear regression of genetic distances between recombination of singe sperm and pedigree data. (**b**) Linear regression between genetic and physical distances for recombination of singe sperm and pedigree data. Note that chromosome 3 appeared to be an outlier in both sperm and pedigree data because of a long ROH (Run Of Homozygosity) region in the middle of the chromosome.



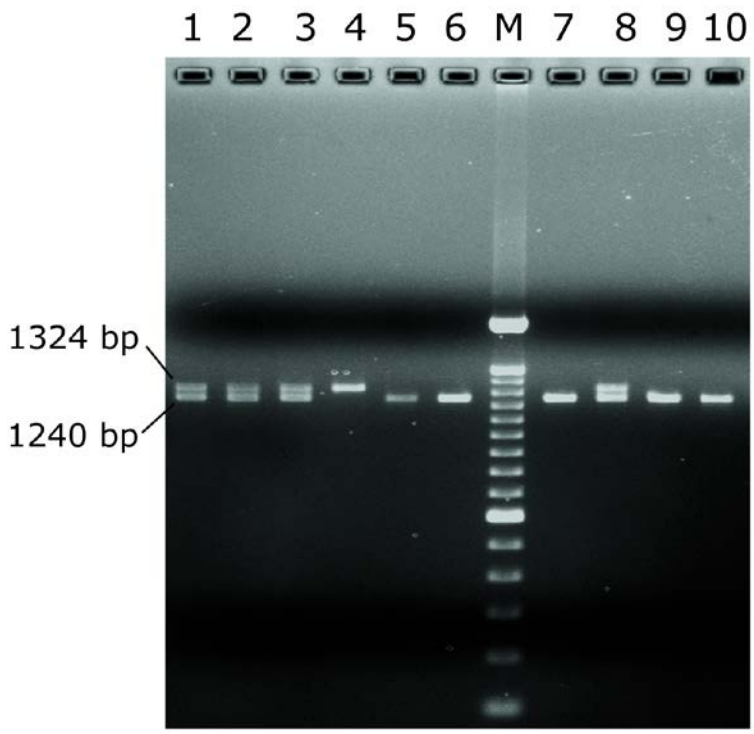
Supplementary Figure 2.10 Comparison of crossover numbers between single sperm and pedigree data. (**a**) Comparison of crossover numbers between single sperm and pedigree data with different density. (**b**) Comparison of crossover numbers between single sperm and pedigree data with same density. The SNP density of single sperm data was down sampled to a comparative level with 10k SNP microarray.



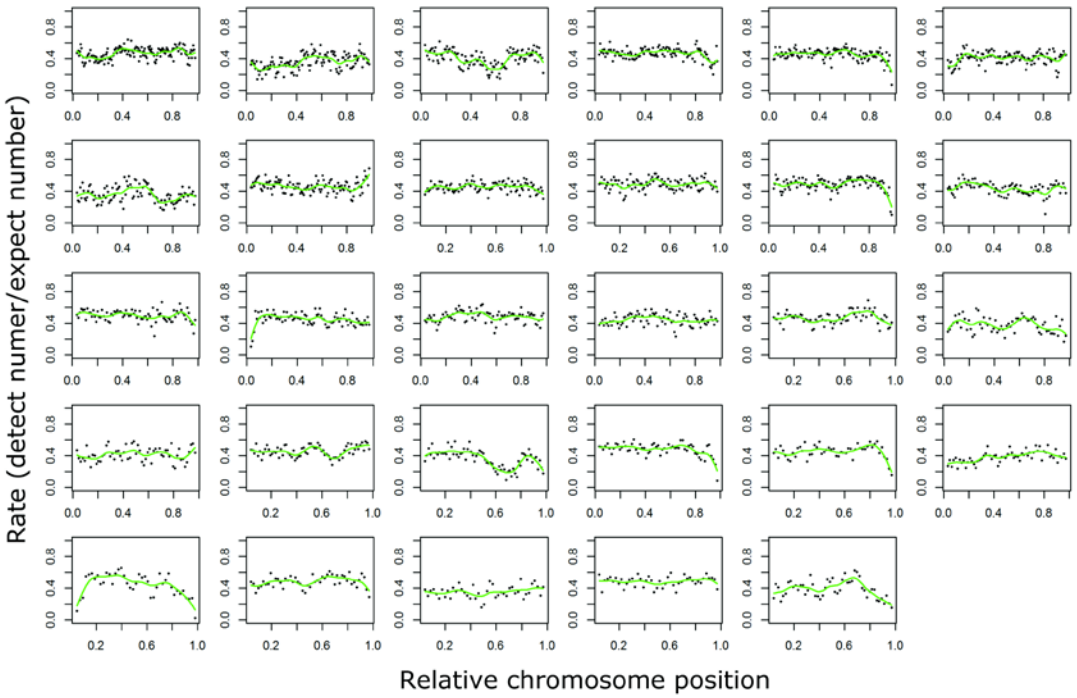
Supplementary Figure 2.11 Distribution of SNPs used for crossover detection on a chromosome. Each chromosome was separated to 100 windows. (the number of SNPs used for crossover detection in each windows) / (the number of SNPs can be used for crossover detection in each windows) was calculated for each window for the 56 single sperms. The smooth spline was fitted across all of the 29 chromosomes of the 56 single sperms. The plot illustrated an unbiased distribution of SNPs used for crossover detection on a chromosome.



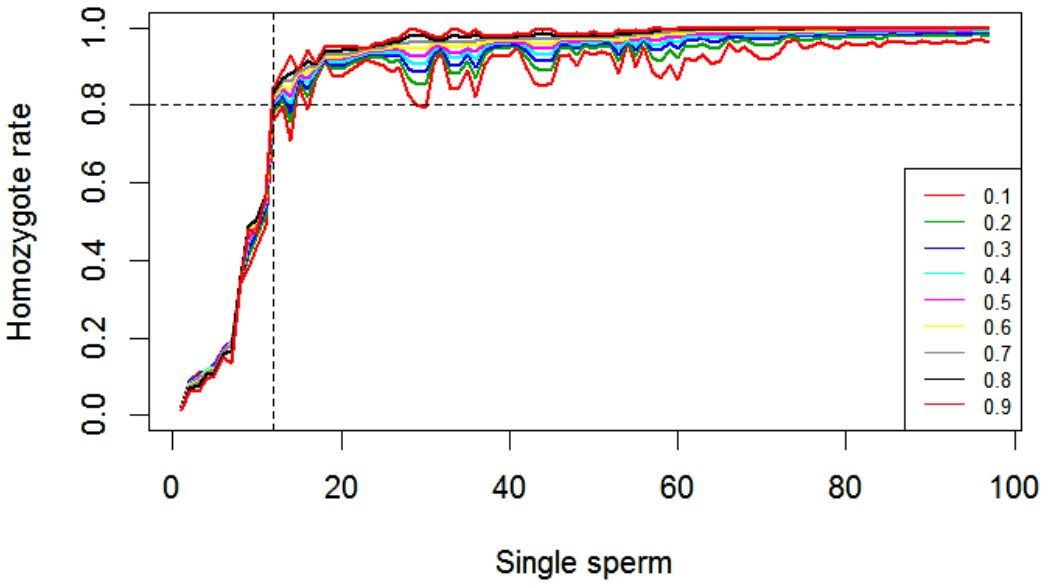
Supplementary Figure 2.12 Agarose gel electrophoresis result to validate *PRDM9* gene structure. NC = negative control; PC = positive control; En = the second region coding for ZnF in Ensembl database (http://useast.ensembl.org/Bos_taurus/Gene/Summary?db=core;g=ENSBTAG00000004538;r=1:45021274-45078993); Nc = the first region coding for ZnF both appear in Ensembl and NCBI databases (http://www.ncbi.nlm.nih.gov/gene/100190914); M = maker, 100 bp ladder. Primers were shown in table S1 and cDNA from testis was used as template for PCR amplification. In En lane, one non-specific amplification band appeared. In the Nc lane, the target band appeared and was confirmed by Sanger sequencing.



Supplementary Figure 2.13 Example result of agarose gel electrophoresis of the PCR products for the zinc finger regions of *PRDM9*. The different band sizes indicated different numbers of zinc finger arrays of *PRDM9*. Note one zinc finger array is 84 bp in length.



Supplementary Figure 2.14 Examination of PCR amplification bias. PCR amplification bias was examined for each sample by calculating the percentage of SNP number detected within 1- Mb windows. The percentages of SNP number detected were evenly distributed along all the chromosomes except a few where there is a slightly drop in the end of chromosome.



Supplementary Figure 2.15 Homozygote rate of single sperms using bovine HD microarray data with different GC score cutoffs. The x-axis was the number of single sperms sorted by the homozygote rate with GC score of 0.7.

Chapter 3: Characterization of Recombination Features and the Genetic Basis in Multiple Cattle Breeds

Abstract

Background: Crossover generated by meiotic recombination is a fundamental event which facilitates meiosis and sexual reproduction. Comparative studies have shown wide variation in recombination between species, but the characterization of recombination between bovine breeds remains elusive. Cattle populations in North America count millions, and the dairy industry has genotyped millions of individuals with pedigree information, providing a unique opportunity to study breed-level variations in recombination.

Results: Based on large pedigrees of Holstein, Jersey, Ayrshire and Brown Swiss cattle with genotype information, we identified over 8.9 million maternal and paternal crossover events within 446,373 three-generation families. We constructed eight genome-wide recombination maps for the two sexes in four cattle breeds. By examining the recombination patterns of different cattle breeds, we confirmed that male recombination map is 10% longer than the female map in all four breeds. When comparing recombination hotpot regions from four breeds, we found that 20% of the hotspots were shared between breeds with each breed exhibiting many breed-specific hotspots. Finally, our breed and sex-specific GWAS analyses confirmed previously reported seven loci that were associated with genome-wide recombination rate and the association of the PRDM9 gene with hotspot usage in both sexes and multiple cattle breeds.

Conclusions: Collectively, our results provided a comprehensive characterization of the meiotic recombination pattern in four cattle breeds and expanded our understanding of the breed differences in recombination within a mammalian species.

Introduction

In eukaryotes, meiotic recombination promotes genetic variation by reciprocal exchange of genetic materials between maternal and paternal homologs and introduction of new combinations of genetic variants into future generations. Aberrant meiotic recombination can cause aneuploidy and often lead to deleterious outcomes (Hassold and Hunt 2001, Lipkin, Moens et al. 2002). As a fundamental biological process, the genetic mechanisms of meiotic recombination are conserved across all eukaryotic species (Coop and Myers 2007).

Humans and chimpanzees showed little conservation on the high-resolution recombination landscape, suggesting a rapid evolution of recombination maps among species (Ptak, Hinds et al. 2005, Winckler, Myers et al. 2005). Pedigree-based studies have discovered considerable within-species variation in recombination rate in humans and mice (Shifman, Bell et al. 2006, Kong, Thorleifsson et al. 2010). Sex-specific recombination rates have been measured in several mammalian species with differences between the two sexes confirmed. Females were reported to have a higher recombination rate than males in humans (Kong, Gudbjartsson et al. 2002, Otto and Lenormand 2002), mice (Dietrich, Miller et al. 1996), dogs (Neff, Broman et al. 1999) and pigs (Tortereau, Servin et al. 2012). But males had more recombination events per meiosis in sheep (Maddox, Davies et al. 2001) and cattle (Ma, O'Connell et

al. 2015, Kadri, Harland et al. 2016). Despite the extensive variation in recombination rate between species and sexes, only a few studies have examined the within-species variation on recombination landscape, mostly in humans (Evans and Cardon 2005, Hinch, Tandon et al. 2011).

Genome-wide association studies (GWAS) have identified genes and genetic variants associated with recombination features in human (Kong, Thorleifsson et al. 2008, Chowdhury, Bois et al. 2009), mouse (Baudat, Buard et al. 2010), cattle (Sandor, Li et al. 2012, Ma, O'Connell et al. 2015) and sheep (Johnston, Berenos et al. 2016) studies. Several genes, including RNF212, CPLX1 and PRDM9, were reported to be associated with individual-level recombination rate across species. Recombination events are more likely to occur in short genomic regions known as hotspots, and many studies have shown that localization of recombination hotspots is associated with the PRDM9 gene in mammals, with the exception of canids that carry a non-functional copy of PRDM9 (Auton, Rui Li et al. 2013). Moreover, the fast-evolving PRDM9 gene is known as a speciation gene that causes hybrid sterility in multiple mouse subspecies (Gregorova and Forejt 2000). Taken together, these studies suggest the existence of genetic basis of recombination that may facilitate a quick response to selection in a short period of time.

The U.S. dairy population consists of many cattle breeds, with the most popular ones being Holstein, Jersey, Brown Swiss, and Ayrshire (VanRaden and Sanders 2003). These four dairy breeds were brought to the U.S. from Europe in the 17th century. The cattle domestication event was estimated to have begun approximately 10,000 to 11,000 years ago (Decker, McKay et al. 2014), but the

formation of diverse cattle breeds was far more recent. Given the fast evolution of recombination and close relationship between cattle breeds, it is questionable whether these cattle breeds will exhibit different recombination landscapes. Current breeding strategies in the cattle industry heavily replied on a small number of superior bulls, which will increase inbreeding level, decrease effective population size, and reduce genetic variations in the cattle population. Recombination may be used to increase genetic variation and address this growing issue of inbreeding in the cattle industry.

The USDA Animal Genomics and Improvement Laboratory (AGIL) maintains a large cattle database for millions of cattle of different breeds with both pedigree and genotype information. This provides a unique opportunity to study recombination features across multiple cattle breeds with high statistical power. Using this large cattle database, we generated eight recombination maps for the Holstein, Jersey, Brown Swiss, and Ayrshire cattle in the two sexes, respectively. We evaluated similarities and differences between the eight recombination maps and documented significant breed- and sex-specific recombination hotspots, revealing both broad- and fine-scale recombination features that differed between cattle breeds. Finally, we performed GWAS of recombination features to study the genetic basis of recombination in four cattle breeds and in two sexes.

Materials and methods

Estimation of Recombination Rates in Cattle Pedigree

We used an approach that was described previously (Ma, O'Connell et al. 2015). First, we extracted three-generation families from the pedigree of Holstein,

Jersey, Ayrshire, and Brown Swiss cattle. We required each three-generation family to have an offspring (first generation), at least one parent (second generation), and at least one grandparent (third generation) to be included in the analysis. We then phased the two haplotypes of an animal (second or third generation) based on the parental genotypes, and crossover locations were identified by comparing either a paternal or maternal haplotype of an offspring (third generation) to its corresponding parent's two haplotypes (second generation). Based on the location of a crossover, a recombination event was assigned to an interval flanked by two informative SNPs (phased heterozygote SNPs in the second generation). To construct recombination maps, we estimated recombination rate between consecutive SNPs as the average number of crossovers per meiosis by assigning a crossover event evenly to all SNP intervals between two informative SNPs. To ensure high-quality recombination maps, we only used those three-generation families genotyped by at least 50K SNP chips. For quality control purposes, we removed animals that had more than 45 crossover events genome-wide, based on the distribution of crossover events across all animals. The X chromosome was excluded from all analyses due to the poor quality of the assembly of chromosome X.

Global and Local Comparisons of Recombination Maps between Breeds

To show the global distribution of recombination rates along the chromosomes, we adopted a smooth spline model of recombination rates against relative physical locations of chromosomes using the smooth.spline function implemented in R 3.2.4 (R 2014). We divided the recombination data into subgroups based on breed and sex to evaluate the patterns of recombination map in each

subgroup. To identify breed-specific recombination hotspot regions, we locally compared recombination rate in a SNP interval between four cattle breeds. We applied a Chi-square test to determine if the proportion of crossover events in a SNP interval per meiosis is independent between pairs of cattle breeds. There were unequal numbers of animals for the four breeds due to different popularity in the dairy industry, which may reduce the power of the Chi-square test.

GWAS of Genome-Wide Recombination Rate and Hotspot Usage Using A Linear Mixed Model

From each three-generation families, we estimated the total number of crossover events per meiosis of the sire or dam (second generation). We adjusted the number of crossover events by SNP density and the number of informative markers (phased heterozygote SNPs) of each animal, and used the adjusted numbers of crossovers for further analyses. Each sire or dam may have multiple crossover measurements if they had multiple offspring, in which case we calculated the average adjusted crossovers as the phenotype of recombination rate. Hotspot regions were tentatively defined as SNP intervals with recombination rate >2.5 standard deviations above the mean. Hotspot usage was calculated as the proportion of crossover events that occurred in the hotspot regions per meiosis. The average hotspot usage was used when multiple measurements were available. To ensure data quality for GWAS, we only included three-generation families where all animals were genotyped by at least 50K SNP chips. Using genome-wide recombination rate and hotspot usage as phenotypes, we tested for the association between a phenotype and each SNP by a linear mixed model. The model equation was fitted as following,

$$\mathbf{y} = \mathbf{Xg} + \mathbf{Za} + \mathbf{e}$$

where **y** refers to the phenotypic value of individuals, **X** is the design matrix of the fixed effects **g**, which include a population mean and the additive effect of the candidate SNP. **Z** is a design matrix for the random animal effect **a**, and **e** represents residuals. The MMAP software was used for all GWAS analyses (O'Connell 2013).

Results

Identification of Crossover Events Using Genotyped Cattle Pedigree

Using a similar approach to our previous studies (Ma, O'Connell et al. 2015, Wang, Shen et al. 2016), we constructed three-generation families that included an offspring, parents, and grandparents from large pedigrees of four dairy cattle breeds. Within a three-generation family, we phased the SNP genotypes of the offspring and parents. By comparing phased genotypes between a sire-offspring or dam-offspring pair, we inferred paternal or maternal crossover events. We used a total of 1,073,914 genotyped animals with pedigree information across four cattle breeds, with Holsteins accounting for 87%, Jersey 11%, Brown Swiss 1.8%, and Ayrshire 0.4% of the data, respectively (Table 3.1). In total, we identified 5.4 million crossover events for Holstein, 3 million for Jersey, 0.41 million for Brown Swiss, and 51,982 for Ayrshire cattle (Table 3.2). To ensure data quality, we excluded the X chromosome and used the USDA-AGIL SNP coordinates that removed likely mapping errors in the UMD 3.1 Bovine genome assembly (Zimin, Delcher et al. 2009, Zimin, Kelley et al. 2012).

Global Recombination Patterns in Four Cattle Breeds and Two Sexes

To capture global recombination patterns, we assigned a crossover event evenly to all consecutive SNP intervals between two informative SNPs and generated recombination maps for Holstein, Jersey, Brown Swiss, and Ayrshire in the two sexes, respectively. To account for different SNP densities, we only included those crossovers identified from 50K SNP data. In Holstein cattle, the average number of crossovers was 25.10 and 22.74 respectively for males and females. This is consistent with the previously reported higher recombination rate in bulls than in cows (Sandor, Li et al. 2012, Ma, O'Connell et al. 2015, Wang, Shen et al. 2016). The male-biased recombination pattern was also confirmed in all three other breeds: an average of 23.7 male and 22.2 female crossovers in Jersey, 24.4 and 22.5 in Brown Swiss, and 24.8 and 22.5 in Ayrshire (Table 3.2). Compared between breeds, Jersey cattle had slightly less crossovers than other breeds in both sexes (decreased by 3% ~ 5% in males and 1.3% ~ 2.3% in females). To visualize the breed-specific recombination patterns along the genome, we generated smooth-spline plots of recombination rate versus chromosomal location in four breeds and two sexes, respectively (Figure 3.1). Overall, cattle recombination rate along the genome exhibited larger variations between the two sexes than between the four breeds. Males and females showed different recombination rates across the chromosomes with the higher male recombination rates most significant near the end of chromosomes (telomeres in cattle). All four breeds showed a similar trend across the chromosomes: males had a considerably higher recombination rate in the telomeric regions (15% of chromosomes to the end), a lower recombination rate in the middle of chromosomes,

and a slightly higher rate at the beginning of chromosomes (centromeres in cattle). In both sexes, Holsteins had the highest recombination rate and Jerseys showed the lowest recombination rates along the chromosomes, except for the telomeric regions.

Regional Recombination Patterns in Four Cattle Breeds and Two Sexes

We generated breed-specific recombination maps by calculating recombination rate between consecutive SNPs across the genome for four cattle breeds in two sexes. To find breed-specific recombination locations, we applied a Chi-square test to identify SNP intervals with significantly different recombination rate between the four breeds in two sexes. Using a genome-wide significance level of P-value $< 8.3 \times 10^{-7}$ after Bonferroni correction, we identified 21 SNP intervals with different recombination rate between Holstein and Jersey in males and 43 SNP intervals in females (Figure 3.2). The most Holstein favored recombination interval was located on chromosome 22, showing a 3.86-fold increase in recombination rate between Holstein and Jersey, a 3.18-fold increase between Holstein and Ayrshire, and a 2.09-fold increase between Holstein and Brown Swiss. However, we didn't find any SNP intervals with different recombination rate between other pairs of breeds, mainly due to the small sample sizes of Brown Swiss and Ayrshire data. More detailed differences in recombination pattern between breeds were revealed as we zoomed into regional recombination maps of the four cattle breeds.

Sharing of Hotspot Regions between Cattle Breeds in Two Sexes

To further characterize local recombination patterns, we tentatively defined hotspot regions as SNP intervals with recombination rate >2.5 standard deviations

above the mean. We herein used the term “hotspot region” instead of “hotspot” because our SNP intervals were much larger (average 44 Kb) than typical human or mouse recombination hotspots (a few Kb or smaller). In males, we identified 1345, 1378, 1295, and 1317 hotspot regions for Holstein, Jersey, Brown Swiss, and Ayrshire, respectively. Similar numbers of hotspot regions were found in females: 1355, 1289, 1421, and 1327 for the four breeds, respectively. A total of 320 (24%) hotspot regions were shared between sexes in Holstein cattle, but this number dropped to 256, 128, and 115 for Jersey, Brown Swiss, and Ayrshire, respectively. This relatively low sharing of hotspot regions between sexes was consistent with the identified sex differences in global recombination patterns.

To evaluate breed-specific distributions of recombination hotspots, we compared hotspot regions across four cattle breeds in two sexes. In males, Holstein, Jersey, Ayrshire, and Brown Swiss each had 258, 394, 480, and 708 unique hotspot regions, with 205 hotspots shared by all four breeds. In addition to the 205 common hotspots, Holstein and Jersey shared 268 hotspot regions, Holstein and Ayrshire shared 68 hotspot regions, Holstein and Brown Swiss shared 133 hotspot regions, Jersey and Ayrshire shared 40 hotspot regions, Jersey and Brown Swiss shared 102 hotspot regions, and Ayrshire and Brown Swiss shared 65 hotspot regions. In females, Holstein, Jersey, Ayrshire, and Brown Swiss each had 618, 714, 914, and 1092 unique hotspot regions, with 37 hotspot regions shared by all breeds. We observed the same trend of hotspot sharing in cows as in bulls: excluding the common hotspots, Holstein and Jersey shared 281 hotspot regions, Holstein and Ayrshire shared 153 hotspot regions, Holstein and Brown Swiss shared 97 hotspot regions,

Jersey and Ayrshire shared 65 hotspot regions, Jersey and Brown Swiss shared 46 hotspot regions, and Ayrshire and Brown Swiss shared 47 hotspot regions. These hotspot sharing results were consistent with the phylogenetic relationships that were reported in diverse cattle populations based on 50K SNP chips (Decker, McKay et al. 2014). We also conducted a correlation analysis of recombination maps between four cattle breeds in two sexes (Table 3.3). Consistent with hotspot sharing results, Holstein and Jersey had the highest correlations in both males and females, while Brown Swiss and Ayrshire had the lowest correlations in the two sexes.

GWAS of Genome-Wide Recombination Rate in Four Breeds and Two Sexes

To reduce biases caused by SNP density differences, recombination rate was adjusted by the number of informative SNP markers in the three-generation families. Using the adjusted genome-wide recombination rate as phenotype, our GWAS analysis included 12,348, 2,237, 1,217, and 340 bulls and 108,391, 18,029, 817, and 791 cows for Holstein, Jersey, Brown Swiss, and Ayrshire, respectively. Compared to the previous GWAS in Holsteins (Ma, O'Connell et al. 2015), this study had more cattle breeds and larger sample sizes. We used a genome-wide significance level of *P*-value $<7.3 \times 10^{-7}$ after Bonferroni correction.

In Holsteins, we successfully validated our previous GWAS results, including four and seven loci that were significantly associated with male and female recombination rates, respectively (Figure 3.3 and Table 3.4). This was expected because the two studies overlapped and this study had a larger sample size. While the previous study identified three associated loci shared between sexes, this study found two additional shared loci between sexes, one on chromosome 1 near the *PRDM9* gene

and the other on chromosome 3 near *MSH4*. In the previous study, both loci were associated only with female recombination rate but not in males. In total, five of the seven associated loci were shared between sexes with the same effect direction in the Holstein cattle. In Jerseys, although the genome-wide recombination rate was lower compared to Holsteins, we were able to confirm two major Holstein associated loci on chromosome 6 and chromosome 10 (Figure 3.3). GWAS for Ayrshire and Brown Swiss found no associations passing the genome-wide significance threshold because of the small sample sizes and low statistical power in these two breeds. However, most of the associated loci in Holsteins showed the same effect direction in Ayrshire and Brown Swiss, and many of them reached nominal significance levels (Table 3.4). Considering the different sample sizes and statistical power, these associations were likely shared between all four cattle breeds.

GWAS of Recombination Hotspot Usage in Two Sexes

Using the hotspot regions identified in each of the four cattle breeds, we measured hotspot usage as the proportion of recombination occurred in hotspot regions for individual animals. To increase accuracy of this measurement, we used only those three-generation families that were genotyped by 50K or higher density SNP chips. In males, the GWAS sample sizes were 2,375, 923, 728 and 994 for Holstein, Jersey, Ayrshire, and Brown Swiss, respectively. The female sample sizes were 18,784, 986, 343, and 165 for the four breeds, respectively. Consistent with previous GWAS studies, we identified a single locus near *PRDM9* to be associated with hotspot usage in Holstein and Jersey cattle, indicating hotspot usage to be a much less polygenic trait compared to recombination rate (Figure 3.4). In Holsteins, the top associated SNP

ARS-BFGL-NGS-83544 (P_{female} = 2.8×10^{-57}; P_{male} = 2.4×10^{-22}) was located downstream of *PRDM9*. In Jersey cattle, we found the same association peak SNP in females but not in males (Figure 3.4). Although this association was not confirmed in Ayrshire or Brown Swiss, the effect direction was consistent across all four cattle breeds in both sexes (Table 3.5).

Discussion

In this study, we took advantage of the large-scale pedigree data maintained by AGIL and the Council of Dairy Cattle Breeding (CDCB) to characterize recombination landscapes of four dairy cattle breeds and to provide comprehensive comparisons between breed-specific recombination maps in two sexes. Our study confirmed the sex difference in recombination rate in four cattle breeds with male recombination map being >10% longer than that of the females, whereas females have higher recombination rates than males in most other mammals. The majority of the cattle sex difference in recombination was found near the telomeres, which is consistent with other mammalian species. Both sexes had a decreased recombination rate around the center of chromosomes in cattle, possibly due to crossover interference (Wang, Shen et al. 2016).

While the four cattle breeds showed similar global recombination patterns across chromosomes, each breed had specific features in the distribution of recombination rate and hotspot regions. Interestingly, the relatedness derived from recombination features was consistent with phylogenic relationships among the four cattle breeds. Holstein and Jersey shared the most recombination features, whereas Ayrshire had the most distinct recombination pattern among the four breeds. The

pairwise resemblance was ordered from the highest to the lowest as following, HO-JE > HO-BS > JE-BS > HO-AY > JE-AY > BS-AY, in both sexes.

Our GWAS results in four cattle breeds confirmed all the loci previously reported to be associated with genome-wide recombination rate in Holsteins. With increased sample sizes and statistical power, we found two more associated loci to be shared between sexes, located on chromosome 1 (*PRDM9*) and chromosome 3 (*MSH4*). Although the Jersey samples had a smaller sample size compared to Holsteins, the Jersey GWAS confirmed major loci associated with recombination features in Holsteins, indicating a shared genetic basis of recombination across cattle breeds. Although we found no genome-wide significant associations in Brown Swiss or Ayrshire, the effect direction of the top associated SNPs were consistent across all four breeds.

Conclusions

Taken together, we characterized the cattle recombination landscape in four dairy breeds and generated breed-specific global and local recombination maps in cattle. We discovered breed specific recombination hotspot regions and identified genetic variants associated with recombination features in two sexes. These results will provide useful insights into the genetic mechanisms and evolution of recombination between breeds and within a mammalian species.

Acknowledgements

Council of Dairy Cattle Breeding is acknowledged for providing the access to the genotype and pedigree data used in this study.

References

Auton, A., Y. Rui Li, J. Kidd, K. Oliveira, J. Nadel, J. K. Holloway, J. J. Hayward, P. E. Cohen, J. M. Greally, J. Wang, C. D. Bustamante and A. R. Boyko (2013). "Genetic recombination is targeted towards gene promoter regions in dogs." PLoS Genet 9(12): e1003984.

Baudat, F., J. Buard, C. Grey, A. Fledel-Alon, C. Ober, M. Przeworski, G. Coop and B. de Massy (2010). "PRDM9 is a major determinant of meiotic recombination hotspots in humans and mice." Science 327(5967): 836-840.

Chowdhury, R., P. R. Bois, E. Feingold, S. L. Sherman and V. G. Cheung (2009). "Genetic analysis of variation in human meiotic recombination." PLoS Genet 5(9): e1000648.

Coop, G. and S. R. Myers (2007). "Live hot, die young: transmission distortion in recombination hotspots." PLoS genetics 3(3): e35.

Decker, J. E., S. D. McKay, M. M. Rolf, J. Kim, A. Molina Alcala, T. S. Sonstegard, O. Hanotte, A. Gotherstrom, C. M. Seabury, L. Praharani, M. E. Babar, L. Correia de Almeida Regitano, M. A. Yildiz, M. P. Heaton, W. S. Liu, C. Z. Lei, J. M. Reecy, M. Saif-Ur-Rehman, R. D. Schnabel and J. F. Taylor (2014). "Worldwide patterns of ancestry, divergence, and admixture in domesticated cattle." PLoS Genet 10(3): e1004254.

Dietrich, W. F., J. Miller, R. Steen, M. A. Merchant, D. Damron-Boles, Z. Husain, R. Dredge, M. J. Daly, K. A. Ingalls and T. J. O'Connor (1996). "A comprehensive genetic map of the mouse genome." Nature 380(6570): 149-152.

Evans, D. M. and L. R. Cardon (2005). "A comparison of linkage disequilibrium patterns and estimated population recombination rates across multiple populations." The American Journal of Human Genetics 76(4): 681-687.

Gregorova, S. and J. Forejt (2000). "PWD/Ph and PWK/Ph inbred mouse strains of Mus m. musculus subspecies--a valuable resource of phenotypic variations and genomic polymorphisms." Folia Biol (Praha) 46(1): 31-41.

Hassold, T. and P. Hunt (2001). "To err (meiotically) is human: the genesis of human aneuploidy." Nat Rev Genet 2(4): 280-291.

Hinch, A. G., A. Tandon, N. Patterson, Y. Song, N. Rohland, C. D. Palmer, G. K. Chen, K. Wang, S. G. Buxbaum and E. L. Akylbekova (2011). "The landscape of recombination in African Americans." Nature 476(7359): 170-175.

Johnston, S. E., C. Berenos, J. Slate and J. M. Pemberton (2016). "Conserved Genetic Architecture Underlying Individual Recombination Rate Variation in a Wild Population of Soay Sheep (Ovis aries)." Genetics 203(1): 583-598.

Kadri, N. K., C. Harland, P. Faux, N. Cambisano, L. Karim, W. Coppieters, S. Fritz, E. Mullaart, D. Baurain, D. Boichard, R. Spelman, C. Charlier, M. Georges and T. Druet (2016). "Coding and noncoding variants in HFM1, MLH3, MSH4, MSH5, RNF212, and RNF212B affect recombination rate in cattle." Genome Res 26(10): 1323-1332.

Kong, A., D. F. Gudbjartsson, J. Sainz, G. M. Jonsdottir, S. A. Gudjonsson, B. Richardsson, S. Sigurdardottir, J. Barnard, B. Hallbeck, G. Masson, A. Shlien, S. T. Palsson, M. L. Frigge, T. E. Thorgeirsson, J. R. Gulcher and K. Stefansson (2002). "A high-resolution recombination map of the human genome." Nat Genet 31(3): 241-247.

Kong, A., G. Thorleifsson, D. F. Gudbjartsson, G. Masson, A. Sigurdsson, A. Jonasdottir, G. B. Walters, A. Jonasdottir, A. Gylfason and K. T. Kristinsson (2010). "Fine-scale recombination rate differences between sexes, populations and individuals." Nature 467(7319): 1099-1103.

Kong, A., G. Thorleifsson, H. Stefansson, G. Masson, A. Helgason, D. F. Gudbjartsson, G. M. Jonsdottir, S. A. Gudjonsson, S. Sverrisson, T. Thorlacius, A. Jonasdottir, G. A. Hardarson, S. T. Palsson, M. L. Frigge, J. R. Gulcher, U. Thorsteinsdottir and K. Stefansson (2008). "Sequence variants in the RNF212 gene associate with genome-wide recombination rate." Science 319(5868): 1398-1401.

Lipkin, S. M., P. B. Moens, V. Wang, M. Lenzi, D. Shanmugarajah, A. Gilgeous, J. Thomas, J. Cheng, J. W. Touchman, E. D. Green, P. Schwartzberg, F. S. Collins and P. E. Cohen (2002). "Meiotic arrest and aneuploidy in MLH3-deficient mice." Nat Genet 31(4): 385-390.

Ma, L., J. R. O'Connell, P. M. VanRaden, B. Shen, A. Padhi, C. Sun, D. M. Bickhart, J. B. Cole, D. J. Null, G. E. Liu, Y. Da and G. R. Wiggans (2015). "Cattle Sex-Specific Recombination and Genetic Control from a Large Pedigree Analysis." PLoS Genet 11(11): e1005387.

Maddox, J. F., K. P. Davies, A. M. Crawford, D. J. Hulme, D. Vaiman, E. P. Cribiu, B. A. Freking, K. J. Beh, N. E. Cockett, N. Kang, C. D. Riffkin, R. Drinkwater, S. S. Moore, K. G. Dodds, J. M. Lumsden, T. C. van Stijn, S. H. Phua, D. L. Adelson, H. R. Burkin, J. E. Broom, J. Buitkamp, L. Cambridge, W. T. Cushwa, E. Gerard, S. M. Galloway, B. Harrison, R. J. Hawken, S. Hiendleder, H. M. Henry, J. F. Medrano, K. A. Paterson, L. Schibler, R. T. Stone and B. van Hest (2001). "An enhanced linkage map of the sheep genome comprising more than 1000 loci." Genome Res 11(7): 1275-1289.

Neff, M. W., K. W. Broman, C. S. Mellersh, K. Ray, G. M. Acland, G. D. Aguirre, J. S. Ziegle, E. A. Ostrander and J. Rine (1999). "A second-generation genetic linkage map of the domestic dog, Canis familiaris." Genetics 151(2): 803-820.

O'Connell, J. R. (2013). MMAP: a comprehensive mixed model program for analysis of pedigree and population data. 63th Annual Meeting of The American Society of Human Genetics. Oct 22–262013.

Otto, S. P. and T. Lenormand (2002). "Resolving the paradox of sex and recombination." Nat Rev Genet 3(4): 252-261.

Ptak, S. E., D. A. Hinds, K. Koehler, B. Nickel, N. Patil, D. G. Ballinger, M. Przeworski, K. A. Frazer and S. Paabo (2005). "Fine-scale recombination patterns differ between chimpanzees and humans." Nat Genet 37(4): 429-434.

R, C. T. (2014). "R: A Language and Environment for Statistical Computing. Vienna, Austria: R Foundation for Statistical Computing."

Sandor, C., W. Li, W. Coppieters, T. Druet, C. Charlier and M. Georges (2012). "Genetic variants in REC8, RNF212, and PRDM9 influence male recombination in cattle." PLoS Genet 8(7): e1002854.

Shifman, S., J. T. Bell, R. R. Copley, M. S. Taylor, R. W. Williams, R. Mott and J. Flint (2006). "A high-resolution single nucleotide polymorphism genetic map of the mouse genome." PLoS biology 4(12): e395.

Tortereau, F., B. Servin, L. Frantz, H. J. Megens, D. Milan, G. Rohrer, R. Wiedmann, J. Beever, A. L. Archibald, L. B. Schook and M. A. Groenen (2012). "A high density recombination map of the pig reveals a correlation between sex-specific recombination and GC content." BMC Genomics 13: 586.

VanRaden, P. M. and A. H. Sanders (2003). "Economic merit of crossbred and purebred US dairy cattle." J Dairy Sci 86(3): 1036-1044.

Wang, Z., B. Shen, J. Jiang, J. Li and L. Ma (2016). "Effect of sex, age and genetics on crossover interference in cattle." Scientific reports 6.

Winckler, W., S. R. Myers, D. J. Richter, R. C. Onofrio, G. J. McDonald, R. E. Bontrop, G. A. McVean, S. B. Gabriel, D. Reich, P. Donnelly and D. Altshuler (2005). "Comparison of fine-scale recombination rates in humans and chimpanzees." Science 308(5718): 107-111.

Zimin, A. V., A. L. Delcher, L. Florea, D. R. Kelley, M. C. Schatz, D. Puiu, F. Hanrahan, G. Pertea, C. P. Van Tassell and T. S. Sonstegard (2009). "A whole-genome assembly of the domestic cow, Bos taurus." Genome biology 10(4): R42.

Zimin, A. V., D. R. Kelley, M. Roberts, G. Marçais, S. L. Salzberg and J. A. Yorke (2012). "Mis-assembled “segmental duplications” in two versions of the Bos taurus genome." PLoS One 7(8): e42680.

Tables

Table 3.1 Number of genotyped animals by breed, sex, and SNP density across four cattle breeds.

Breed	Male		Female		Total
	<50K SNP	≥50K SNP	<50K SNP	≥50K SNP	
Holstein	69,846	85,572	700,724	73,693	929,835(87%)
Jersey	8,409	9,409	99,487	2,787	120,092(11%)
Brown Swiss	1,441	14,191	3,241	382	19,255(1.8%)
Ayrshire	160	1,355	2,049	1,168	4,732(0.4%)

Table 3.2 Summary statistics for the number of meiosis and crossovers, and genome-wide recombination rate by breed and sex.

Breed	Crossovers		Meiosis		Recombination Rate	
	Male	Female	Male	Female	Male	Female
Holstein	>3.7 M	>1.7 M	147,930	71,687	25.5	23.2
Jersey	>2.3 M	>0.7 M	108,163	37,008	23.7	22.2
Brown Swiss	328,653	9,804	13,556	436	24.4	22.5
Ayrshire	40,161	11,821	1,620	526	24.8	22.5

Table 3.3 Correlation coefficient between recombination maps of four cattle breeds in two sexes. Correlations in males were presented in the top-right triangle and female correlations in the bottom-left.

	HO	**JE**	**BS**	**AY**
HO		0.87	0.83	0.67
JE	0.70		0.78	0.63
BS	0.52	0.38		0.31
AY	0.45	0.34	0.27	

Table 3.4 Top SNPs associated with genome-wide recombination rate in two sexes and four cattle breeds.

SNP	Chr	Pos	Holstein				Jersey				Brown Swiss				Ayrshire			
			Male		Female		Male		Female		Male		Female		Male		Female	
			Beta	*P*	Beta	*P*	Beta	*P*	Beta	*P*	Beta	*P*	Beta	*P*	Beta	*P*	Beta	*P*
ARS-BFGL-NGS-83544	1	158140250	0.57	1.4×10^{-12}	0.51	5.0×10^{-31}	0.53	0.30	0.24	0.46	0.80	6.5×10^{-4}	0.58	5.0×10^{-3}	0.75	0.05	0.37	2.0×10^{-3}
Hapmap58808-rs29017431	3	52612595	0.52	1.5×10^{-7}	0.36	1.8×10^{-11}	0.14	4.9×10^{-2}	0.32	1.3×10^{-5}	0.57	4.6×10^{-5}	0.15	4.8×10^{-1}	0.12	0.07	0.27	0.74
BovineHD0600030770	6	109176815	-0.78	2.8×10^{-39}	-0.38	4.4×10^{-35}	-0.64	2.1×10^{-11}	-0.32	6.2×10^{-8}	-0.73	4.3×10^{-5}	-0.15	1.8×10^{-1}	-0.67	5.0×10^{-5}	-0.26	2.5×10^{-1}
ARS-BFGL-NGS-117763	6	129108015	0.92	1.3×10^{-47}	0.57	3.0×10^{-63}	0.61	1.1×10^{-11}	0.47	2.1×10^{-17}	0.64	2.7×10^{-4}	0.65	2.8×10^{-1}	0.58	5.4×10^{-3}	0.49	0.85
ARS-BFGL-BAC-10975	10	21225382	-0.74	1.4×10^{-32}	-0.42	3.8×10^{-38}	-0.35	1.3×10^{-3}	-0.46	8.1×10^{-13}	-0.30	7.6×10^{-2}	-0.32	5.2×10^{-2}	-0.31	0.08	-0.39	0.66
BovineHD1000009771	10	29619086	-0.03	0.6	0.34	4.5×10^{-11}	-0.17	1.3×10^{-2}	0.31	5.1×10^{-6}	0.03	0.85	0.83	1.4×10^{-2}	-0.09	0.87	0.39	0.06
BTA-78285-no-rs	10	86717378	-0.57	1.9×10^{-20}	-0.41	8.0×10^{-37}	-0.23	1.9×10^{-2}	-0.14	2.0×10^{-2}	-0.48	5.8×10^{-3}	-0.33	9.4×10^{-2}	-0.26	4.8×10^{-3}	-0.37	0.56
ARS-BFGL-NGS-23945	26	31439013	-0.15	1.6×10^{-2}	-0.18	2.7×10^{-8}	0.007	0.94	-0.18	4.2×10^{-3}	-0.13	0.57	-0.80	0.78	0.08	0.18	-0.11	0.46

Table 3.5 Top SNP associated with hotspot usage in two sexes and four cattle breeds.

SNP	Chr	Pos	Holstein				Jersey				Brown Swiss				Ayrshire			
			Male		Female		Male		Female		Male		Female		Male		Female	
			Beta	*P*	Beta	*P*	Beta	*P*	Beta	*P*	Beta	*P*	Beta	*P*	Beta	*P*	Beta	*P*
ARS-BFGL-NGS-83544	1	15814 0250	-0.01	2.4×10^{-22}	-0.01	2.8×10^{-57}	-0.005	3.1×10^{-3}	-0.005	3.7×10^{-7}	-0.003	0.4	-0.005	0.5	-0.002	0.74	-0.002	0.6

Figures

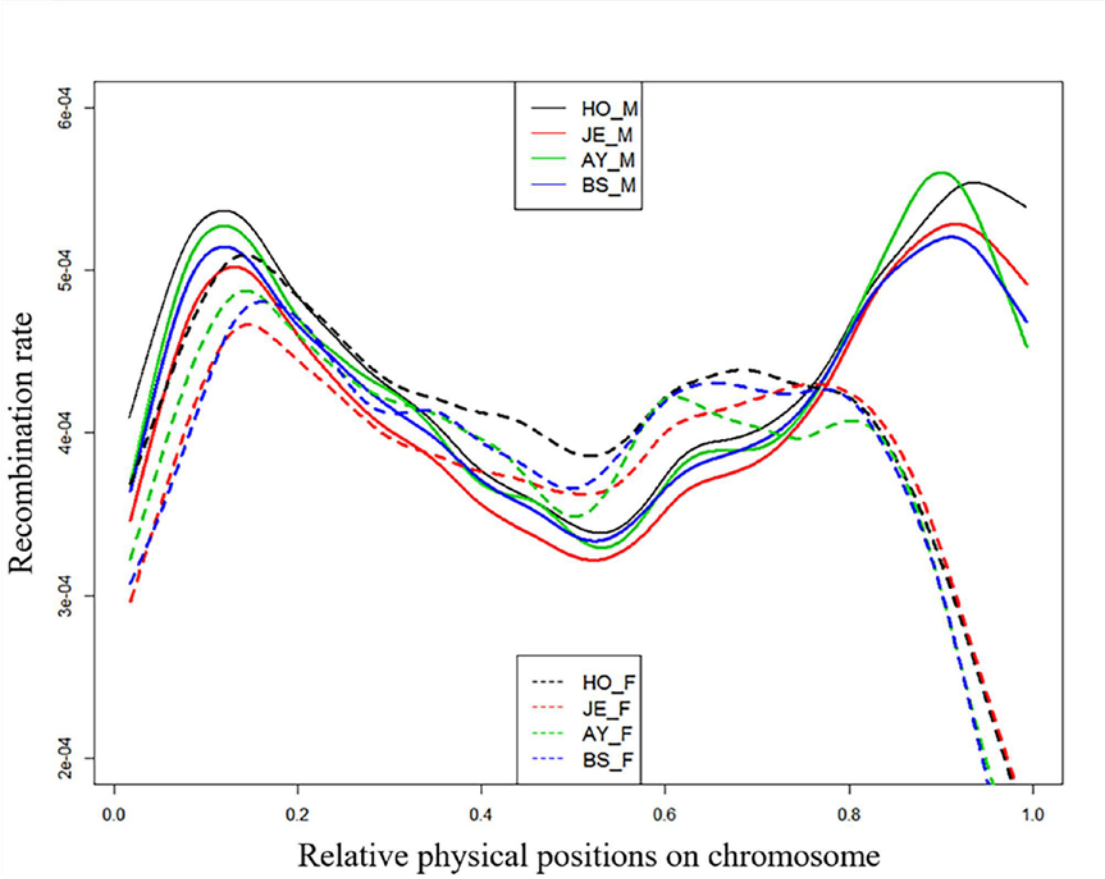


Figure 3.1 Smooth-spline plotting of male and female recombination rates along a chromosome in four cattle breeds. The relative physical position for each SNP interval on a chromosome was calculated by standardizing the original physical position by the chromosome length: a value of zero corresponds to the beginning of a chromosome and a value of one corresponds to the end. Solid lines: males, dotted lines: females. The smooth-spline model was fitted across all 29 autosomes

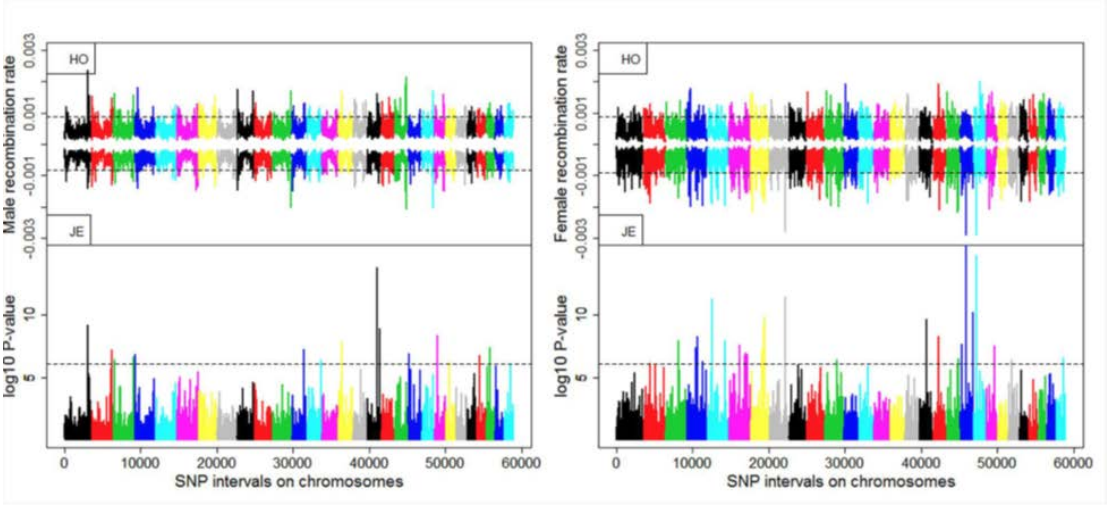


Figure 3.2 Breed-specific recombination locations between Holstein and Jersey in males (left) and females (right). For each panel, recombination rates in each SNP intervals of two groups were shown in the top half and corresponding *P*-values were shown in bottom. Different colors were used to distinguish the 29 chromosomes. The dash line shows the significance level of P-value < 8.3×10^{-7} after Bonferroni correction

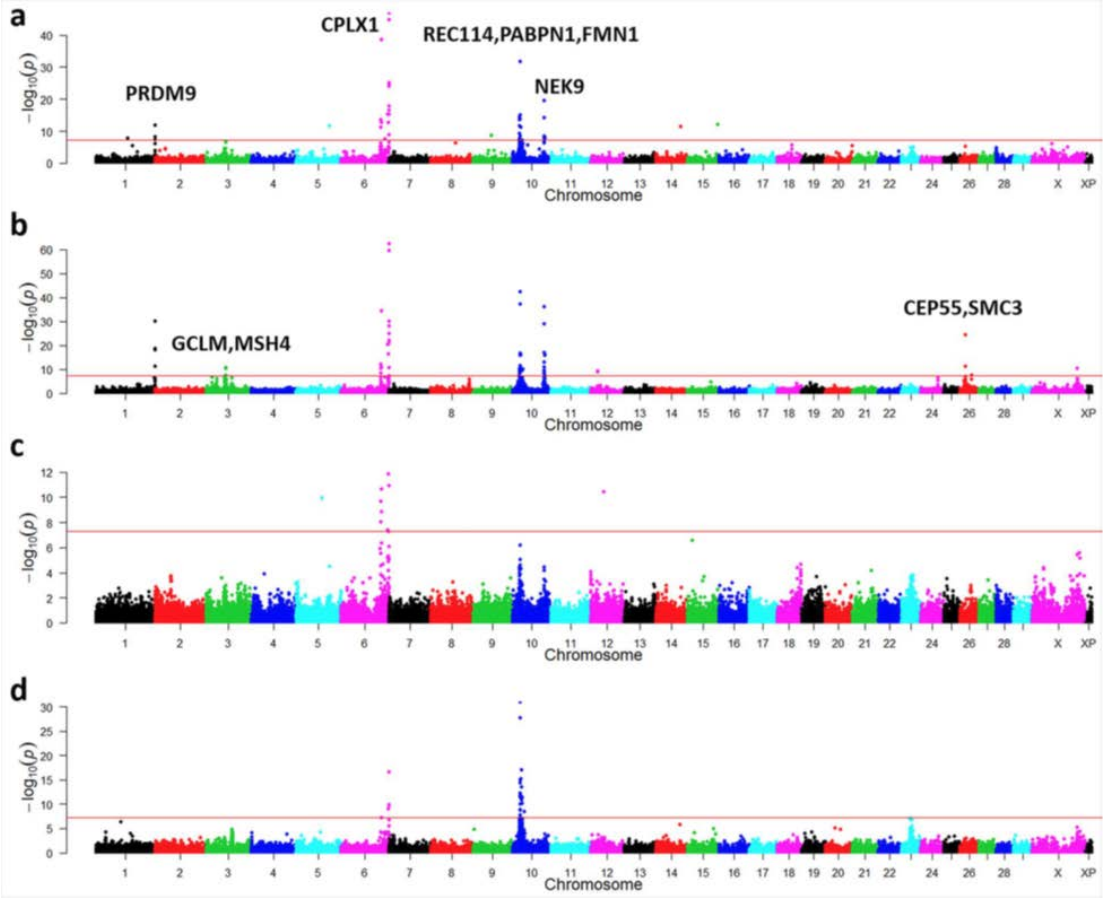


Figure 3.3 Manhattan plot of the GWAS of genome-wide recombination rates for Holsteins and Jerseys in two sexes. **a**: Holstein males, **b**: Holstein females , **c**: Jersey males, and **d**: Jersey females. Different colors were used to distinguish the 29 chromosomes. The genome-wide significance level of 1.6×10^{-7} was shown by the horizontal dotted line. USDA-AGIL SNP coordinates were used for plotting, which placed *PRDM9*-linked SNPs to the end of Chromosome 1

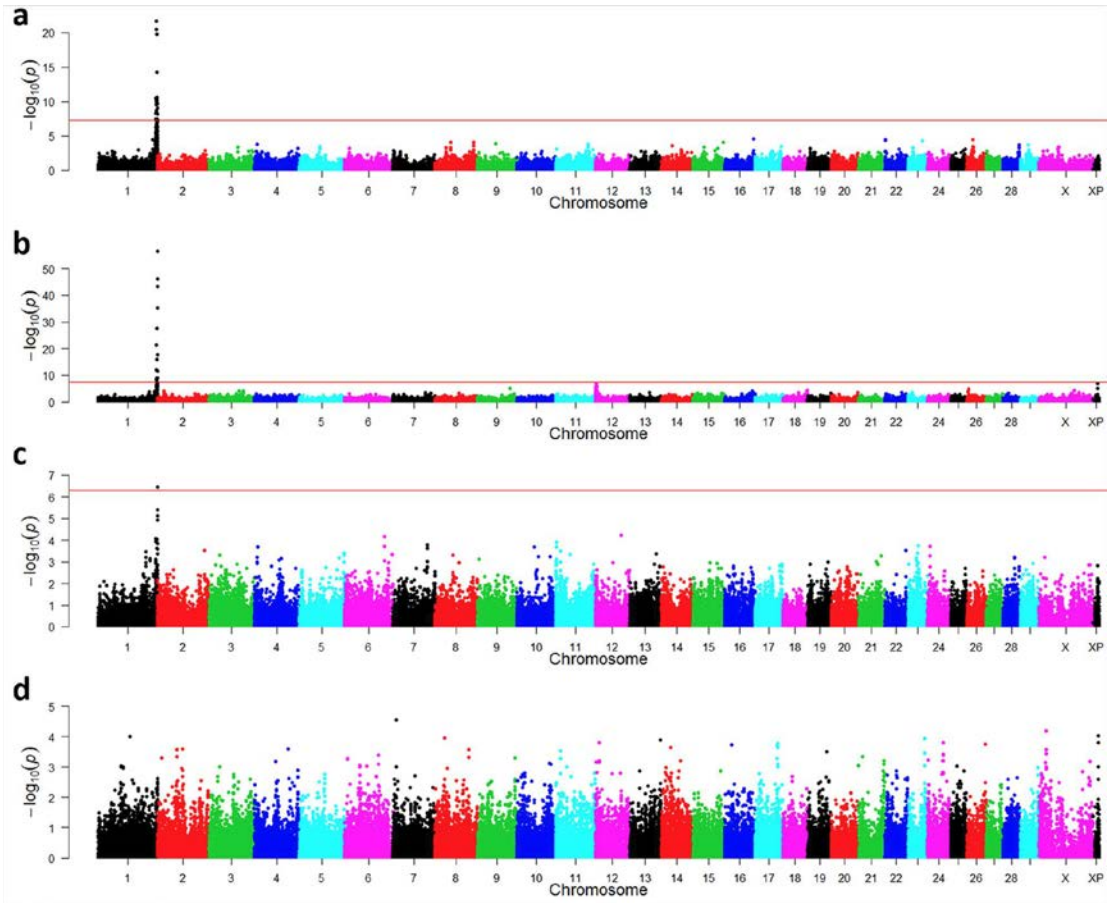
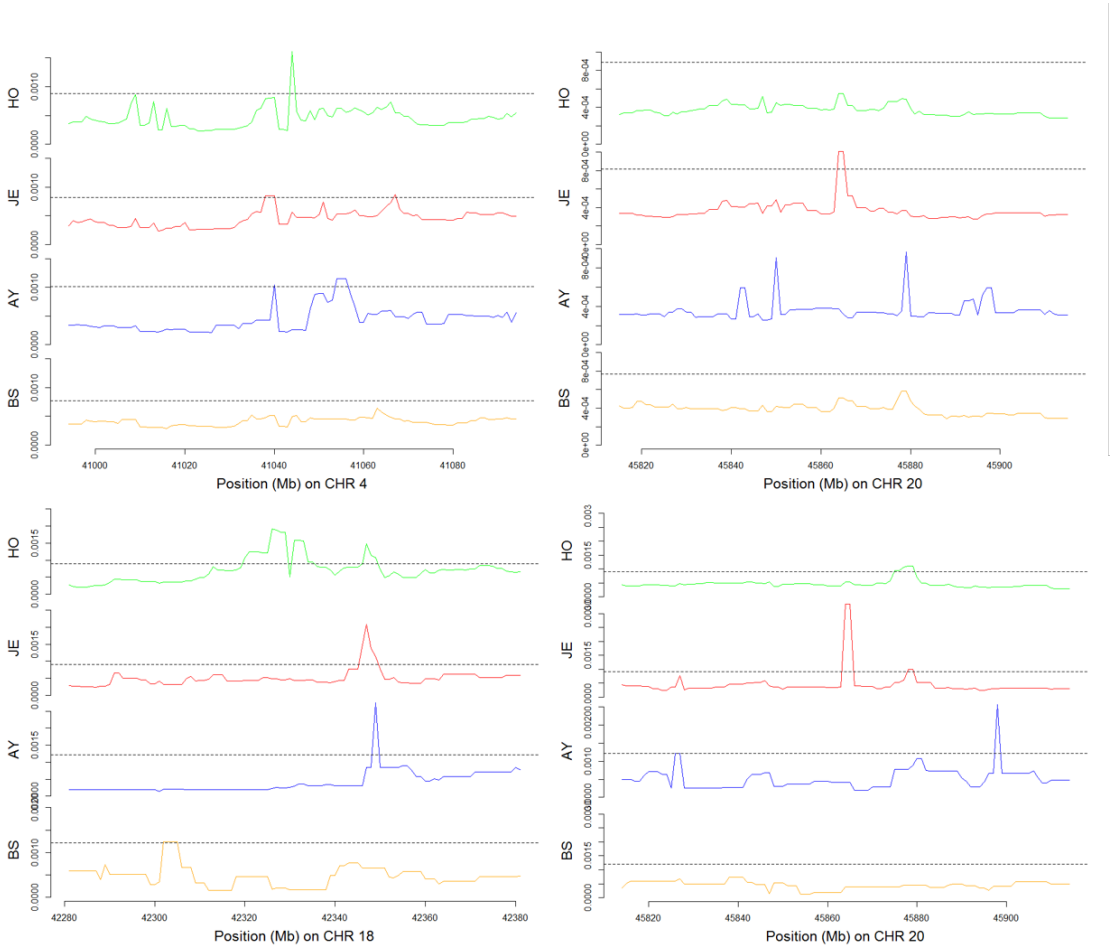


Figure 3.4 Manhattan plot of GWAS of hotspot usage for Holstein and Jersey cattle in two sexes. **a**: Holstein males, **b**: Holstein females , **c**: Jersey males, and **d**: Jersey females. Different colors were used to distinguish the 29 chromosomes. The genome-wide significance level of 1.6×10^{-7} was shown by the horizontal dotted line. USDA-AGIL SNP coordinates were used for plotting, which placed *PRDM9*-linked SNPs to the end of Chromosome 1

Supplemental Materials



Supplementary Figure 3.1. Example regions showing different recombination patterns between four cattle breeds. Top 2: males; Bottom 2: females

Chapter 4: Characterization of Maternal Age and Temperature Effects on Recombination Features in Holstein

Abstract

Background: Meiotic recombination is a fundamental biological event which facilitates genetic diversion. It is known to be phenotypically plastic and affected by intrinsic and extrinsic conditions. Maternal age effect on meiotic recombination rates has been characterized in wide ranges of species excluding bovine, and the direction of the effect remains inconclusive. The temperature effect on meiotic recombination rates were extensively studied in small organisms, here we seek the temperature effect in a large mammal, cattle.

Results: Based on large pedigrees of Holstein, we identified maternal crossover events within 36,999 three-generation families. We tested the correlation between maternal age and meiotic recombination rates in cattle and found a non-linear relationship. We also found that increasing environment temperature is associated with elevated recombination rates in Holstein dams.

Conclusions: Collectively, our results provided fist hand information regarding the plastic nature of meiotic recombination in cattle. We characterized the maternal age and temperature effect on recombination rates and suggested the adaptation of meiotic recombination to environmental factor stimuli in a mammalian species.

Background

Meiotic recombination occurs in all sexually reproducing organisms, it facilitates the paring and alignment of homologue chromosomes during prophase. It also plays a powerful role in driving the evolution of genome structure by introducing genetic variation into the new generation. The recombination process promotes the exchange of genetic materials between maternal and paternal homologs and thus enable each child to receive a unique combination of parental genomes from infinite possible outcomes. Aberrant meiotic recombination can lead to aneuploidy and often deleterious outcomes (Hassold and Hunt 2001, Lipkin, Moens et al. 2002). There is tremendous variability in meiotic recombination between individuals, populations, and species. There are also a number of factors that influence the meiotic recombination within the genome. Genome-wide association studies (GWAS) have identified genes and genetic variants associated with recombination features in human (Kong, Thorleifsson et al. 2008, Chowdhury, Bois et al. 2009), mouse (Baudat, Buard et al. 2010), cattle (Sandor, Li et al. 2012, Ma, O'Connell et al. 2015) and sheep (Johnston, Berenos et al. 2016) studies. Several genes, including RNF212, CPLX1 and PRDM9, were reported to be associated with individual-level recombination rate across species.

It is also known that recombination rates are phenotypically plastic. The meiotic recombination rates may present variations in response to different environmental conditions. Factors that could affect recombination rates could be grouped into two categories: intrinsic factors, such as age or sex, and extrinsic factors, such as temperature, parasitism, mating and social stress. Some studies suggested that

maternal age had a significant effect on the meiotic recombination rates, a recent multicohort analysis in human suggested a small and significant positive maternal age effect on recombination rates, which contradicted previous studies in smaller human populations (Kong, Barnard et al. 2004, Coop, Wen et al. 2008, Bleazard, Ju et al. 2013, Martin, Christ et al. 2015). A negative effect of maternal age was also observed in mice and hamster (Polani and Jagiello 1976, Sugawara and Mikamo 1983). However, neither maternal age nor any other non-genetic factors were found to affect the recombination rates in wild sheep (Johnston, Berenos et al. 2016). Extensive studies of maternal age effect were conducted in Drosophila, worms, plants and yeast, yet no consistent conclusions have been reached (Hunter, Robinson et al. 2016, Modliszewski and Copenhaver 2017). As for paternal age effect, many studies reported no effect of paternal age on meiotic recombination (Griffin, Abruzzo et al. 1995, Hussin, Roy-Gagnon et al. 2011),

Meiotic recombination rates can also be modulated by extrinsic factors, such as nutrient condition, chemical treatments or temperature. In Drosophila, environmental stressors such as EDTA chemical addition or lack of nutrients would increase the recombination rates (Levine 1955). In Saccharomyces cerevisiae, the meiotic recombination rates would also increase when facing scarce resources (Abdullah and Borts 2001). However, the effect of temperature on meiotic recombination rates is complex. Some found a positive correlation in Arabidopsis thaliana, Caenorhabditis elegans and the Melanoplus femurrubrum (Church and Wimber 1969, Rose and Baillie 1979, Francis, Lam et al. 2007). Some found a negative correlation in Allium ursinum (Loidl 1989) and some found positive then

negative correlation in Drosophila melanogaster (Stern 1926). A recent study in Drosophila found that increased exposure to heat shock conditions is associated with a non-linear increase in meiotic recombination rates, suggesting modulating recombination frequency is one mechanism by which organisms can rapidly respond to environmental cues and confer increased adaptive potential to their offspring (Jackson, Nielsen et al. 2015). Heat stress has been found as a major factor which decrease the fertility in dairy cattle. In fact the conception rate for Holstein in summer season is 20-30% less than it in winter seasons (Cavestany, el-Wishy et al. 1985).

The USDA Animal Genomics and Improvement Laboratory (AGIL) maintains a large dairy cattle database with both pedigree and genotype information for millions of cattle of different breeds. This provides a unique opportunity to study the recombination features across multiple cattle breeds with high statistical power. Using this unique cattle database, we characterized the recombination feature for the Holstein dams. As mounting evidence have shown meiotic recombination rates response to both intrinsic and extrinsic stress, the primary motivation for this study was to determine how recombination rates varies in relation to advancing maternal age and common environment factor such as temperature.

Methods

Estimation of Recombination Rate in Cattle Pedigree

We used an approach that was described previously (Ma, O'Connell et al. 2015). First, we extracted three-generation families from the pedigree of Holstein, and Jersey cattle. We required each three-generation family to have an offspring (first

generation), at least one parent (second generation), and at least one grandparent (third generation) to be genotyped. We then phased the two haplotypes of an animal (second or third generation) based on the parental genotypes, and crossover locations were identified by comparing either a paternal or maternal haplotype of an offspring (third generation) to its corresponding parent's two haplotypes (second generation). Based on the location of a crossover, a recombination event was assigned to an interval flanked by two informative SNPs (phased heterozygote SNPs in the second generation). To construct recombination maps, we estimated recombination rate between consecutive SNPs as the average number of crossovers per meiosis by assigning a crossover event evenly to all SNP intervals between two informative SNPs. To ensure high-quality recombination maps, we only used those three-generation families genotyped by at least 50K SNP chips. For quality control purposes, we removed animals that had more than 45 crossover events genome-wide, based on the distribution of crossover events across all animals. The X chromosome was excluded from all analyses due to the poor quality of the assembly of chromosome X.

Accessing Temperature Information from NOAA Database

NOAA (National Oceanic and Atmospheric Administration) is an American scientific agency that focuses on the conditions of the oceans and atmosphere. It's also the largest database that contains the weather records of most US cities since 1970s. By accessing the NOAA database, we extracted the weather condition during the month when each calf was conceived. We then combined the temperature information with our recombination records for our association study.

A Mixed Model for Temperature Effect Analysis

From each of the 36,009 three-generation families in Holstein, we estimated the total number of crossover events per meiosis of the dam (second generation). We then adjusted the number of crossover events by SNP density and the number of informative markers (phased heterozygote SNPs) of each animal, and used the adjusted numbers of crossovers for further analyses. Using the corrected recombination rate residuals as phenotypes, we adopted a linear mixed model to test for the temperature effect. The model equation was fitted as following,

$$\boldsymbol{Y_{ij}} = \mu + \boldsymbol{T_i} + \mathbf{A} + \boldsymbol{A^2} + \mathbf{B} + \boldsymbol{B^2} + \boldsymbol{\varepsilon_j}$$

where Y refers to the recombination rate residuals of individuals, Ti is the fixed effects of temperature at level i, A represents maternal age, and B represents dams' birthday. Statistical differences were declared as significant at $P < 0.05$.

Results

Identification of crossover events using genotyped cattle pedigree

Adopting a similar method to our previous studies (Ma, O'Connell et al. 2015, Wang, Shen et al. 2016), we identified meiotic crossovers events by first constructing three-generation families including an offspring, parents, and grandparents from large Holstein pedigrees. We then phased the SNP genotypes of the offspring and parents within each three-generation family. By comparing phased genotypes between a dam-offspring pair, we inferred maternal crossover events. In this study, we used a total of 305,545 three-generation families and identified 6,677,618 maternal crossover events in Holstein. Temperature and location information were available for 36,999 of these

three-generation families. To ensure data quality, we excluded the X chromosome and used the USDA-AGIL SNP coordinates that corrected likely mapping errors in the Bovine genome assembly UMD3.1 (Zimin, Delcher et al. 2009, Zimin, Kelley et al. 2012).

Effects of Maternal Age on Recombination Rates in Holstein

Many studies has suggested a correlation between maternal age and the number of recombination events in plants, mice and human. However, the direction of this correlation is inconsistent, even in the same species. Recent study in humans observed that recombination rates increase with age in females in contrast with previous decreasing trend in human studies (Hussin, Roy-Gagnon et al. 2011, Bleazard, Ju et al. 2013, Martin, Christ et al. 2015). Mice also demonstrate a negative correlation between recombination rates and maternal age (Polani and Jagiello 1976). As for large mammals, a study in wild sheep stated that there were no effect of maternal age on recombination rates (Johnston, Berenos et al. 2016). To investigate the relationship of recombination rates with maternal age in cattle, we first modeled the relationship between a continuous maternal age variable and the recombination rates residuals in Holstein (Figure 4.1a). The recombination rates residuals were obtained by correcting the recombination rates with SNP chip density and the number of informative markers within each of the 305,545 three-generation families. We observed a check shaped trend which the recombination rates first decreases from 20 to 65 month old dams (Figure 4.1b). Then an elevation in recombination rates were discovered as the dam continues to grow older from 65 to 200 month. However, the decreasing recombination rates trend we found from 20 to 65 month has more

statistical power as it consists of 91.8% of our records. Dams that gave birth between 65 and 100 month old consists of 21,798 (7.1%) cases of our data, and we have 3,321 (1.1%) cases of dam giving birth above 100 month age. The increasing tend we observed when maternal age is over 65 month old couldn't be considered random since the large sample size (~ 25,000) in this range. The decreasing trend resembled the negative maternal age effect on recombination rates in mouse and some human studies, and the increasing trend resembled the positive effect maternal age on recombination rates in the latest multicohort human study (Martin, Christ et al. 2015). In this study, 65 months is the age which separated two groups of cows by the direction of maternal age effect. We then divided the Holstein cows into ten groups based on maternal age starting from 20 month old, with an increment of 10 month and plotted the relationship (Figure 4.1c). We noted that the last age group consists of all the records of dam giving birth with maternal age over 120 month. The trend in the boxplot matched our previous finding, with age groups over 60 month exhibiting an increasing trend with gradually increased variance. A decreasing tread before 65 months old maternal age is still clear.

We did the same study in Jersey for additional evidence and utilized 34,877 records in Jersey dams (Figure 4.1d). There were only 1,597 (4.6%) records of Jersey dams giving birth after 100 month old. A similar trend was observed with steeper slopes and a different turning point at 100 month maternal age. No previous studies have reported bell shaped relationship between maternal age and recombination rate. However, studies have shown bell shaped relationship regarding the temperature effect on recombination rates in plants, fruit fly and grasshopper, although the

underlying mechanism of meiotic recombination are totally different between cattle and these animals.

Effects of Temperature on Maternal Recombination Rates in Holstein

Studies have shown that temperature affects meiotic recombination rates in many poikilothermic organisms, including yeast, plants, worms, grasshopper and large reptiles such as crocodile. However, the direction of the effect remains inconclusive. Utilizing the large cattle pedigree information from USDA database, we characterized the meiotic recombination features in Holstein dams and integrated them with the environmental temperature information. Through NOAA national weather database, we obtained the temperature information for the month when calves were conceived. In total, we have 36,999 records with both the environmental factor information and the meiotic recombination outcome. We then fit a model to explore the temperature effect on maternal recombination rates in cattle. THI (temperature humidity index) was wildly used to indicate heat or cold stress in cattle. An environment with THI exceeds 78 should be considered as heat stress condition for cattle, when both the productive and reproductive performance of cow would be seriously affected. There are no predetermined THI index for cold stress condition as the wellness of the cattle in cold environment also depend on their hair coats. We adopted the THI index into this study and chose the most extreme temperature as the threshold of heat and cold stress for cattle. Temperature above 80 Fahrenheit would be considered to be heat stress condition for cattle and temperature below 40 Fahrenheit would be considered to be cold stress condition for cattle.

We divided the cows into three groups based on the temperature condition of the conceiving environments. Over 6K cows conceived under heat stress, over 25K cows conceived in mild temperature environment and 6K cows conceived under cold stress. To characterize the temperature effect, we plotted the recombination rates residuals against the three conditions (Figure 4.2). Cows under heat stress showed an elevation of recombination rates and cow under cold stress showed a decreased recombination rates. A rising of recombination events under heat stress is consistent with recent study on fruit flies which found increase of recombination frequency under heat shock condition (Jackson, Nielsen et al. 2015).

To comprehensive dissect the temperature effect, we fitted a mixed model which includes the temperature condition as fixed effect with three levels. We also included maternal age, the birth date of the dam and the quadratic term of these factors in the model (Table 4.1). Result showed that heat stress condition would increase the recombination rates (P=0.0038) while cold stress condition may correlated to the recombination rate decrease (P=0.059). It also confirmed that maternal age have a negative effect on the recombination rates (P=1.2x10-11) while dam's birthday would slightly influence the recombination rate (P=5.41x10-33). Both the quadratic terms of the factors also correlated with recombination rates with minute effective size in Holstein. We then tested if the temperature effect on recombination rates could be inherited into the next generation and found no connection.

Conclusions

Taken together, we correlated the meiotic recombination features with non-genetic factors as maternal age and temperature. We characterized their relationships with recombination rates and discovered a non-linear correlation between maternal age and meiotic recombination in cattle. We also found elevated recombination rates with increasing conceiving environmental temperature. These results will provide useful insights into both the intrinsic and extrinsic effect on meiotic recombination and evolution of plastic recombination in cattle.

Acknowledgements

Council of Dairy Cattle Breeding is acknowledged for providing the access to the genotype and pedigree data used in this study.

References

Abdullah, M. F. and R. H. Borts (2001). "Meiotic recombination frequencies are affected by nutritional states in Saccharomycescerevisiae." Proc Natl Acad Sci U S A 98(25): 14524-14529.

Baudat, F., J. Buard, C. Grey, A. Fledel-Alon, C. Ober, M. Przeworski, G. Coop and B. de Massy (2010). "PRDM9 is a major determinant of meiotic recombination hotspots in humans and mice." Science 327(5967): 836-840.

Bleazard, T., Y. S. Ju, J. Sung and J. S. Seo (2013). "Fine-scale mapping of meiotic recombination in Asians." BMC Genet 14: 19.

Cavestany, D., A. B. el-Wishy and R. H. Foote (1985). "Effect of season and high environmental temperature on fertility of Holstein cattle." J Dairy Sci 68(6): 1471-1478.

Chowdhury, R., P. R. Bois, E. Feingold, S. L. Sherman and V. G. Cheung (2009). "Genetic analysis of variation in human meiotic recombination." PLoS Genet 5(9): e1000648.

Church, K. and D. E. Wimber (1969). "Meiosis in the grasshopper: chiasma frequency after elevated temperature and x-rays." Can J Genet Cytol 11(1): 209-216.

Coop, G., X. Wen, C. Ober, J. K. Pritchard and M. Przeworski (2008). "High-resolution mapping of crossovers reveals extensive variation in fine-scale recombination patterns among humans." Science 319(5868): 1395-1398.

Francis, K. E., S. Y. Lam, B. D. Harrison, A. L. Bey, L. E. Berchowitz and G. P. Copenhaver (2007). "Pollen tetrad-based visual assay for meiotic recombination in Arabidopsis." Proc Natl Acad Sci U S A 104(10): 3913-3918.

Griffin, D. K., M. A. Abruzzo, E. A. Millie, L. A. Sheean, E. Feingold, S. L. Sherman and T. J. Hassold (1995). "Non-disjunction in human sperm: evidence for an effect of increasing paternal age." Hum Mol Genet 4(12): 2227-2232.

Hassold, T. and P. Hunt (2001). "To err (meiotically) is human: the genesis of human aneuploidy." Nat Rev Genet 2(4): 280-291.

Hunter, C. M., M. C. Robinson, D. L. Aylor and N. D. Singh (2016). "Genetic Background, Maternal Age, and Interaction Effects Mediate Rates of Crossing Over in Drosophila melanogaster Females." G3 (Bethesda) 6(5): 1409-1416.

Hussin, J., M. H. Roy-Gagnon, R. Gendron, G. Andelfinger and P. Awadalla (2011). "Age-dependent recombination rates in human pedigrees." PLoS Genet 7(9): e1002251.

Jackson, S., D. M. Nielsen and N. D. Singh (2015). "Increased exposure to acute thermal stress is associated with a non-linear increase in recombination frequency and an independent linear decrease in fitness in Drosophila." Bmc Evolutionary Biology 15.

Johnston, S. E., C. Berenos, J. Slate and J. M. Pemberton (2016). "Conserved Genetic Architecture Underlying Individual Recombination Rate Variation in a Wild Population of Soay Sheep (Ovis aries)." Genetics 203(1): 583-598.

Kong, A., J. Barnard, D. F. Gudbjartsson, G. Thorleifsson, G. Jonsdottir, S. Sigurdardottir, B. Richardsson, J. Jonsdottir, T. Thorgeirsson, M. L. Frigge, N. E. Lamb, S. Sherman, J. R. Gulcher and K. Stefansson (2004). "Recombination rate and reproductive success in humans." Nat Genet 36(11): 1203-1206.

Kong, A., G. Thorleifsson, H. Stefansson, G. Masson, A. Helgason, D. F. Gudbjartsson, G. M. Jonsdottir, S. A. Gudjonsson, S. Sverrisson, T. Thorlacius, A. Jonasdottir, G. A. Hardarson, S. T. Palsson, M. L. Frigge, J. R. Gulcher, U. Thorsteinsdottir and K. Stefansson (2008). "Sequence variants in the RNF212 gene associate with genome-wide recombination rate." Science 319(5868): 1398-1401.

Levine, R. P. (1955). "Chromosome Structure and the Mechanism of Crossing Over." Proceedings of the National Academy of Sciences of the United States of America 41(10): 727-730.

Lipkin, S. M., P. B. Moens, V. Wang, M. Lenzi, D. Shanmugarajah, A. Gilgeous, J. Thomas, J. Cheng, J. W. Touchman, E. D. Green, P. Schwartzberg, F. S. Collins and P. E. Cohen (2002). "Meiotic arrest and aneuploidy in MLH3-deficient mice." Nat Genet 31(4): 385-390.

Loidl, J. (1989). "Effects of Elevated-Temperature on Meiotic Chromosome Synapsis in Allium-Ursinum." Chromosoma 97(6): 449-458.

Ma, L., J. R. O'Connell, P. M. VanRaden, B. Shen, A. Padhi, C. Sun, D. M. Bickhart, J. B. Cole, D. J. Null, G. E. Liu, Y. Da and G. R. Wiggans (2015). "Cattle Sex-Specific Recombination and Genetic Control from a Large Pedigree Analysis." PLoS Genet 11(11): e1005387.

Martin, H. C., R. Christ, J. G. Hussin, J. O'Connell, S. Gordon, H. Mbarek, J. J. Hottenga, K. McAloney, G. Willemsen, P. Gasparini, N. Pirastu, G. W. Montgomery, P. Navarro, N. Soranzo, D. Toniolo, V. Vitart, J. F. Wilson, J. Marchini, D. I. Boomsma, N. G. Martin and P. Donnelly (2015). "Multicohort analysis of the maternal age effect on recombination." Nat Commun 6: 7846.

Modliszewski, J. L. and G. P. Copenhaver (2017). "Meiotic recombination gets stressed out: CO frequency is plastic under pressure." Curr Opin Plant Biol 36: 95-102.

Polani, P. E. and G. M. Jagiello (1976). "Chiasmata, meiotic univalents, and age in relation to aneuploid imbalance in mice." Cytogenet Cell Genet 16(6): 505-529.

Rose, A. M. and D. L. Baillie (1979). "The Effect of Temperature and Parental Age on Recombination and Nondisjunction in CAENORHABDITIS ELEGANS." Genetics 92(2): 409-418.

Sandor, C., W. Li, W. Coppieters, T. Druet, C. Charlier and M. Georges (2012). "Genetic variants in REC8, RNF212, and PRDM9 influence male recombination in cattle." PLoS Genet 8(7): e1002854.

Stern, C. (1926). "An Effect of Temperature and Age on Crossing-Over in the First Chromosome of Drosophila Melanogaster." Proc Natl Acad Sci U S A 12(8): 530-532.

Sugawara, S. and K. Mikamo (1983). "Absence of Correlation between Univalent Formation and Meiotic Nondisjunction in Aged Female Chinese Hamsters." Cytogenetics and Cell Genetics 35(1): 34-40.

Wang, Z., B. Shen, J. Jiang, J. Li and L. Ma (2016). "Effect of sex, age and genetics on crossover interference in cattle." Sci Rep 6: 37698.

Zimin, A. V., A. L. Delcher, L. Florea, D. R. Kelley, M. C. Schatz, D. Puiu, F. Hanrahan, G. Pertea, C. P. Van Tassell, T. S. Sonstegard, G. Marcais, M. Roberts, P. Subramanian, J. A. Yorke and S. L. Salzberg (2009). "A whole-genome assembly of the domestic cow, Bos taurus." Genome Biol 10(4): R42.

Zimin, A. V., D. R. Kelley, M. Roberts, G. Marcais, S. L. Salzberg and J. A. Yorke (2012). "Mis-assembled "segmental duplications" in two versions of the Bos taurus genome." PLoS One 7(8): e42680.

Tables

Table 4.1 Results of the mixed model analyses of factors related to the recombintion rates in Holstein.

Factor	Beta	P-value
Mean	**-1.12**	**2.21×10^{-16}**
Cold Stress	**-1.43×10^{-1}**	**5.89×10^{-2}**
Heat Stress	**2.03×10^{-1}**	**3.88×10^{-3}**
Maternal age	**-7.73×10^{-2}**	**1.20×10^{-11}**
Maternal age^2	**4.07×10^{-4}**	**1.75×10^{-3}**
Dam birthday	**5.05×10^{-3}**	**5.41×10^{-33}**
Dam birthday^2	**-4.44×10^{-7}**	**1.71×10^{-42}**

Figures

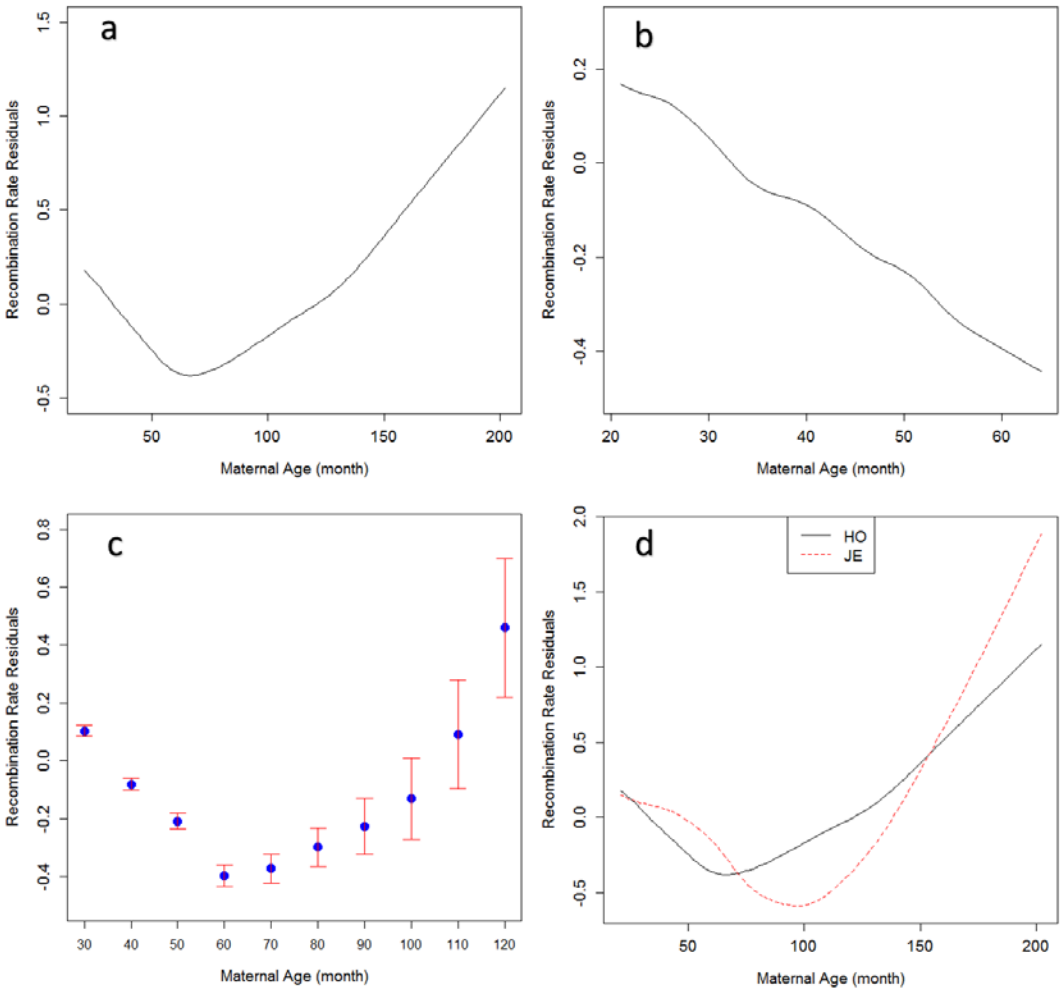


Figure 4.1 Smooth-spline and box plotting of recombination rate residuals against maternal age in cattle. a: Recombination rates residuals along with maternal age in 305,545 Holstein three-generation families. b: Recombination rates residuals along with maternal age from 20 to 65 month in Holstein. c: Recombination rates residuals in different maternal age groups. d: Comparison between Holstein and Jersey.

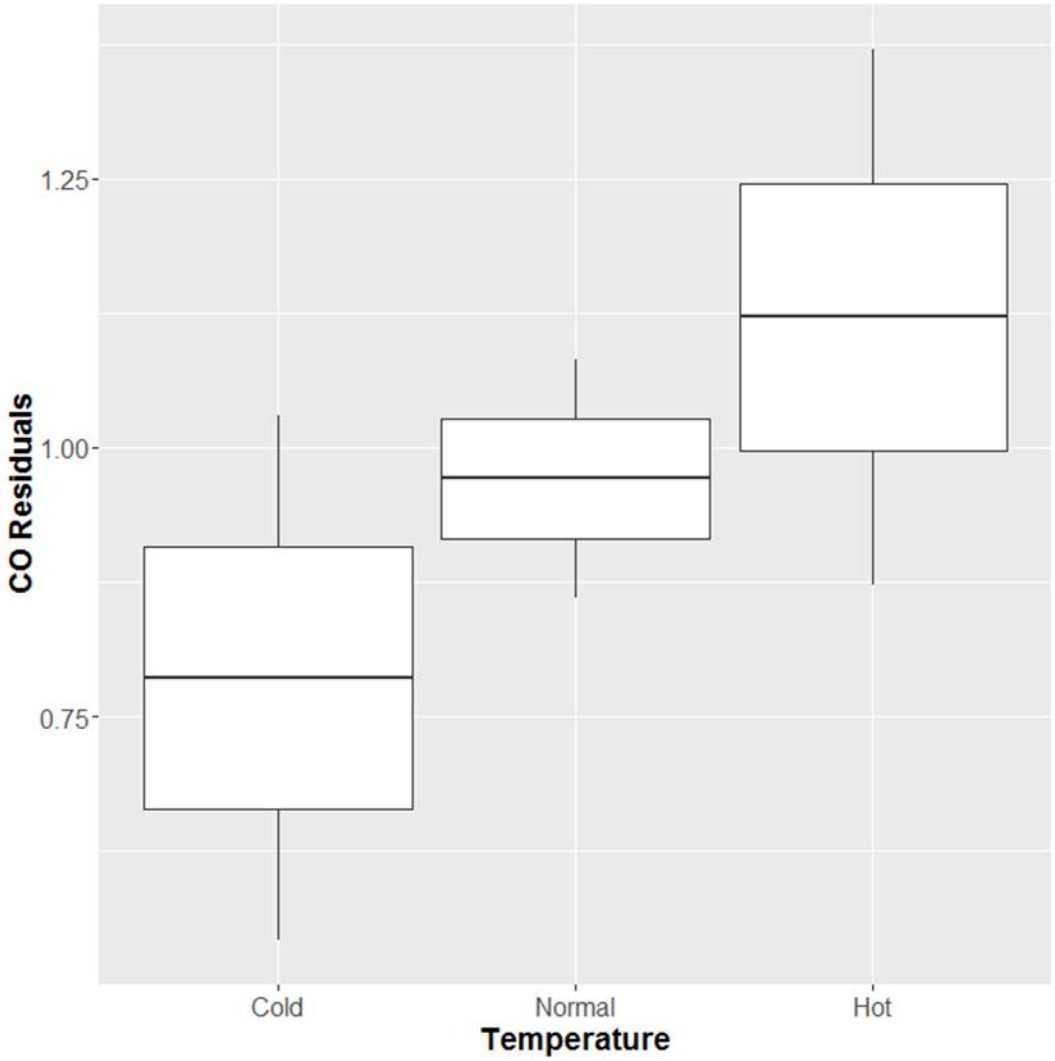


Figure 4.2 Boxplot of recombination rate residuals in three temperature conditions in Holstein. Cold: Cold stress condition. Hot: Heat stress condition

Chapter 5: Summary and Future Directions

This thesis aims to commit a comprehensive study on the effect of genetics, sex, and age on recombination and its evolution using a uniquely large cattle database hosted at the USDA, where over a million animals with full pedigree information have been genotyped and new data are being generated at an increasing speed. More specifically, the three main objectives were 1) to characterize PRDM9 alleles in cattle and construct allele-specific recombination maps; 2) to study the genetic basis of recombination as well as the sex and breed effect on recombination; 3) to characterize the effect of maternal age and stress factors that may shape recombination patterns in cattle.

In Chapter 2, five distinct PRDM9 alleles were identified and characterized in multiple cattle breeds. Allele-specific recombination maps were generated using over 239,000 meiosis in Holstein. One allele of PRDM9 was discovered to have distinct zinc-finger repeats than other alleles and shape the recombination landscape differently in both frequency and placement, especially towards the chromosomal ends. Based on the composition of zinc finger repeats, we predicted two 17-bp motifs 'CGNNANNAGNANNANNA' and 'ANNANNANNANNANGGC' of the most common allele (i.e., allele 1) and allele 5, respectively. Consistent enrichment patterns of these binding motifs were found in corresponding recombination hotspot regions. In addition, we constructed recombination maps of the same animal through both the pedigree methods and single sperm genotyping. The two maps were consistent, and the quality of pedigree-based results was validated.

In Chapter 3, we extended our analysis in recombination patterns to 4 major U.S. dairy cattle breeds, Holstein, Jersey, Ayrshire and Brown Swiss. We identified over 8.9 million crossover events within 446,373 three-generation families. A total of eight genome-wide recombination maps were constructed for the two sexes in four cattle breeds. We found that the male genetic map is 10% longer than the female map in four breeds. In males, 15% of the hotspots are shared among 4 breeds while breed-specific hotspots constitute up to 55% of total hotspots, suggesting a tremendous variation of recombination feature between breeds. With power of detection in four cattle breeds, our GWAS analyses of genome-wide recombination rate discovered two loci to be shared between sexes in cattle, one variant near PRDM9 and one variant in MSH4. As for the recombination placement represented by hotspot usage, a single variant near PRDM9 was found to have positive effect in both sex in Holstein (P_{female} = 2.8×10-57; P_{male} = 2.4×10-22), whereas the association was only found in Jersey females but not male. It is suggested that the genetic basis of recombination features are partially conserved among cattle breeds.

In Chapter 4, we characterize the effect of maternal age in cattle as the third mammalian species (in addition to humans and mice) to be studied regarding the age effect. Based on large pedigrees of Holstein, we identified 6,677,618 maternal crossover events in a total of 305,545 three-generation families. Contrary to the positive correlation in human or the negative correlation in mice, our results show a quadratic trend between recombination rate and age: a decreasing trend of recombination in young dam population before 60 month old, and an increasing trend

in the old dam population over 60 month old. This trend is further validated in Jersey with an increased turning point at 100 month.

There was no previous study of temperature effect on recombination in a mammalian species. Here by using recombination data derived from 36,999 three-generation families in Holstein that have temperature data available during pregnancy, we presented a positive correlation between temperature and recombination rate in cattle.

Collectively, by analyzing genomic data in cattle with unprecedentedly large pedigree information, our studies come together to illustrate the recombination properties and broaden our understanding of recombination in a large mammalian species.

There are multiple future research directions that can be extended beyond these current studies. First of all, as the 1,000 Bull Genomes project continues to generate whole genome sequence (WGS) data and a new bovine assembly by PacBio sequencing is about to be published, a finer-scale recombination map from even larger pedigree could reveal more details of the recombination features in cattle.

Second, as we have shown the maternal age effect in cattle, a maternal age-specific recombination maps could be constructed to show the effect of age at fine scale, which would be the first study of this kind.

Third, since the male map is longer in cattle, which is in contrary to most mammalian species except sheep, it is interesting to dig deeper into the biological difference of recombination between cattle and other species. This could contribute to the understanding of sexual dimorphism in recombination.

Fourth, it is promising to link the recombination with the reproduction and production traits in cattle and consider it in the evaluation of breeding values of cattle.

Last, it is valuable to sequence more PRDM9 alleles in multiple cattle breeds and across generations to study the evolution of PRDM9 and recombination maps together in a species under intense artificial selection.

Compiled References

Abdullah, M. F. and R. H. Borts (2001). "Meiotic recombination frequencies are affected by nutritional states in Saccharomycescerevisiae." Proc Natl Acad Sci U S A **98**(25): 14524-14529.

Ahlawat, S., P. Sharma, R. Sharma, R. Arora and S. De (2016). "Zinc finger domain of the PRDM9 gene on chromosome 1 exhibits high diversity in ruminants but its paralog PRDM7 contains multiple disruptive mutations." PloS one **11**(5): e0156159.

Ahlawat, S., P. Sharma, R. Sharma, R. Arora, N. Verma, B. Brahma, P. Mishra and S. De (2016). "Evidence of positive selection and concerted evolution in the rapidly evolving PRDM9 zinc finger domain in goats and sheep." Animal Genetics **47**(6): 740-751.

Arbeithuber, B., A. J. Betancourt, T. Ebner and I. Tiemann-Boege (2015). "Crossovers are associated with mutation and biased gene conversion at recombination hotspots." Proc Natl Acad Sci U S A **112**(7): 2109-2114.

Archibald, A. L., C. S. Haley, J. F. Brown, S. Couperwhite, H. A. McQueen, D. Nicholson, W. Coppieters, A. Van de Weghe, A. Stratil, A. K. Wintero and et al. (1995). "The PiGMaP consortium linkage map of the pig (Sus scrofa)." Mamm Genome **6**(3): 157-175.

Auton, A., Y. R. Li, J. Kidd, K. Oliveira, J. Nadel, J. K. Holloway, J. J. Hayward, P. E. Cohen, J. M. Greally and J. Wang (2013). "Genetic recombination is targeted towards gene promoter regions in dogs." PLoS Genet **9**(12): e1003984.

Axelsson, E., M. T. Webster, A. Ratnakumar, C. P. Ponting, K. Lindblad-Toh and L. Consortium (2012). "Death of PRDM9 coincides with stabilization of the recombination landscape in the dog genome." Genome research **22**(1): 51-63.

Baker, C. L., S. Kajita, M. Walker, R. L. Saxl, N. Raghupathy, K. Choi, P. M. Petkov and K. Paigen (2015). "PRDM9 drives evolutionary erosion of hotspots in Mus musculus through haplotype-specific initiation of meiotic recombination." PLoS Genet **11**(1): e1004916.

Balcova, M., B. Faltusova, V. Gergelits, T. Bhattacharyya, O. Mihola, Z. Trachtulec, C. Knopf, V. Fotopulosova, I. Chvatalova and S. Gregorova (2016). "Hybrid sterility locus on chromosome X controls meiotic recombination rate in mouse." PLoS genetics **12**(4): e1005906.

Barton, N. H. and B. Charlesworth (1998). "Why sex and recombination?" Science **281**(5385): 1986-1990.

Baudat, F., J. Buard, C. Grey, A. Fledel-Alon, C. Ober, M. Przeworski, G. Coop and B. de Massy (2010). "PRDM9 is a major determinant of meiotic recombination hotspots in humans and mice." Science **327**(5967): 836-840.

Baudat, F., Y. Imai and B. de Massy (2013). "Meiotic recombination in mammals: localization and regulation." Nat Rev Genet **14**(11): 794-806.

Bell, S. P. and A. Dutta (2002). "DNA replication in eukaryotic cells." Annu Rev Biochem **71**: 333-374.

Berg, I. L., R. Neumann, S. Sarbajna, L. Odenthal-Hesse, N. J. Butler and A. J. Jeffreys (2011). "Variants of the protein PRDM9 differentially regulate a set of human meiotic recombination hotspots highly active in African populations." Proc Natl Acad Sci U S A **108**(30): 12378-12383.

Berg, I. L., N. Rita, K. W. G. Lam, S. Shriparna, O. H. Linda, C. A. May and A. J. Jeffreys (2010). "PRDM9 variation strongly influences recombination hot-spot activity and meiotic instability in humans." Nature Genetics **42**(10): 859-863.

Birtle, Z. and C. P. Ponting (2006). "Meisetz and the birth of the KRAB motif." Bioinformatics **22**(23): 2841-2845.

Bleazard, T., Y. S. Ju, J. Sung and J. S. Seo (2013). "Fine-scale mapping of meiotic recombination in Asians." BMC Genet **14**: 19.

Boitard, S., W. Rodriguez, F. Jay, S. Mona and F. Austerlitz (2016). "Inferring Population Size History from Large Samples of Genome-Wide Molecular Data - An Approximate Bayesian Computation Approach." PLoS Genet **12**(3): e1005877.

Buard, J., E. Rivals, D. D. de Segonzac, C. Garres, P. Caminade, B. De Massy and P. Boursot (2014). "Diversity of Prdm9 zinc finger array in wild mice unravels new facets of the evolutionary turnover of this coding minisatellite." PLoS One **9**(1): e85021.

Campbell, C. L., C. Bherer, B. E. Morrow, A. R. Boyko and A. Auton (2016). "A Pedigree-Based Map of Recombination in the Domestic Dog Genome." G3 (Bethesda).

Cavestany, D., A. B. el-Wishy and R. H. Foote (1985). "Effect of season and high environmental temperature on fertility of Holstein cattle." J Dairy Sci **68**(6): 1471-1478.

Chowdhury, R., P. R. Bois, E. Feingold, S. L. Sherman and V. G. Cheung (2009). "Genetic analysis of variation in human meiotic recombination." PLoS Genet **5**(9): e1000648.

Church, K. and D. E. Wimber (1969). "Meiosis in the grasshopper: chiasma frequency after elevated temperature and x-rays." Can J Genet Cytol **11**(1): 209-216.

Clark, A. G., X. Wang and T. Matise (2010). "Contrasting methods of quantifying fine structure of human recombination." Annual review of genomics and human genetics **11**: 45.

Coop, G. and M. Przeworski (2006). "An evolutionary view of human recombination." Nature Reviews Genetics **8**(1): 23-34.

Coop, G., X. Wen, C. Ober, J. K. Pritchard and M. Przeworski (2008). "High-resolution mapping of crossovers reveals extensive variation in fine-scale recombination patterns among humans." Science **319**(5868): 1395-1398.

Cox, A., C. L. Ackert-Bicknell, B. L. Dumont, Y. Ding, J. T. Bell, G. A. Brockmann, J. E. Wergedal, C. Bult, B. Paigen, J. Flint, S. W. Tsaih, G. A. Churchill and K. W. Broman (2009). "A new standard genetic map for the laboratory mouse." Genetics **182**(4): 1335-1344.

Crawford, A. M., K. G. Dodds, A. J. Ede, C. A. Pierson, G. W. Montgomery, H. G. Garmonsway, A. E. Beattie, K. Davies, J. F. Maddox, S. W. Kappes and et al. (1995). "An autosomal genetic linkage map of the sheep genome." Genetics **140**(2): 703-724.

Crow, J. F. (2000). "The origins, patterns and implications of human spontaneous mutation." Nat Rev Genet **1**(1): 40-47.

Davies, B., E. Hatton, N. Altemose, J. G. Hussin, F. Pratto, G. Zhang, A. G. Hinch, D. Moralli, D. Biggs and R. Diaz (2016). "Re-engineering the zinc fingers of PRDM9 reverses hybrid sterility in mice." Nature **530**(7589): 171-176.

de Massy, B. (2013). "Initiation of meiotic recombination: how and where? Conservation and specificities among eukaryotes." Annu Rev Genet **47**: 563-599.

Decker, J. E., S. D. McKay, M. M. Rolf, J. Kim, A. Molina Alcala, T. S. Sonstegard, O. Hanotte, A. Gotherstrom, C. M. Seabury, L. Praharani, M. E. Babar, L. Correia de Almeida Regitano, M. A. Yildiz, M. P. Heaton, W. S. Liu, C. Z. Lei, J. M. Reecy, M. Saif-Ur-Rehman, R. D. Schnabel and J. F. Taylor (2014). "Worldwide patterns of ancestry, divergence, and admixture in domesticated cattle." PLoS Genet **10**(3): e1004254.

Dietrich, W. F., J. Miller, R. Steen, M. A. Merchant, D. Damron-Boles, Z. Husain, R. Dredge, M. J. Daly, K. A. Ingalls and T. J. O'Connor (1996). "A comprehensive genetic map of the mouse genome." Nature **380**(6570): 149-152.

Fazakerley, D. J., S. Naghiloo, R. Chaudhuri, F. Koumanov, J. G. Burchfield, K. C. Thomas, J. R. Krycer, M. J. Prior, B. L. Parker and B. A. Murrow (2015). "Proteomic analysis of GLUT4 storage vesicles reveals tumor suppressor candidate 5 (TUSC5) as a novel regulator of insulin action in adipocytes." Journal of biological chemistry **290**(39): 23528-23542.

Fledel-Alon, A., E. M. Leffler, Y. Guan, M. Stephens, G. Coop and M. Przeworski (2011). "Variation in human recombination rates and its genetic determinants." PLoS One **6**(6): e20321.

Francis, K. E., S. Y. Lam, B. D. Harrison, A. L. Bey, L. E. Berchowitz and G. P. Copenhaver (2007). "Pollen tetrad-based visual assay for meiotic recombination in Arabidopsis." Proc Natl Acad Sci U S A **104**(10): 3913-3918.

Gerton, J. L. and R. S. Hawley (2005). "Homologous chromosome interactions in meiosis: diversity amidst conservation." Nat Rev Genet **6**(6): 477-487.

Gratten, J., N. R. Wray, W. J. Peyrot, J. J. McGrath, P. M. Visscher and M. E. Goddard (2016). "Risk of psychiatric illness from advanced paternal age is not predominantly from de novo mutations." Nat Genet **48**(7): 718-724.

Gregorova, S. and J. Forejt (2000). "PWD/Ph and PWK/Ph inbred mouse strains of Mus m. musculus subspecies--a valuable resource of phenotypic variations and genomic polymorphisms." Folia Biol (Praha) **46**(1): 31-41.

Grelon, M. (2016). "Meiotic recombination mechanisms." C R Biol **339**(7-8): 247-251.

Grey, C., P. Barthès, G. Chauveau-Le Friec, F. Langa, F. Baudat and B. De Massy (2011). "Mouse PRDM9 DNA-binding specificity determines sites of histone H3 lysine 4 trimethylation for initiation of meiotic recombination." PLoS Biol **9**(10): e1001176.

Griffin, D. K., M. A. Abruzzo, E. A. Millie, L. A. Sheean, E. Feingold, S. L. Sherman and T. J. Hassold (1995). "Non-disjunction in human sperm: evidence for an effect of increasing paternal age." Hum Mol Genet **4**(12): 2227-2232.

Groenen, M. A., P. Wahlberg, M. Foglio, H. H. Cheng, H. J. Megens, R. P. Crooijmans, F. Besnier, M. Lathrop, W. M. Muir, G. K. Wong, I. Gut and L. Andersson (2009). "A high-density SNP-based linkage map of the chicken genome reveals sequence features correlated with recombination rate." Genome Res **19**(3): 510-519.

Gruhn, J. R., C. Rubio, K. W. Broman, P. A. Hunt and T. Hassold (2013). "Cytological studies of human meiosis: sex-specific differences in recombination originate at, or prior to, establishment of double-strand breaks." PLoS One **8**(12): e85075.

Hassold, T. and P. Hunt (2001). "To err (meiotically) is human: the genesis of human aneuploidy." Nat Rev Genet **2**(4): 280-291.

Hayashi, K., K. Yoshida and Y. Matsui (2005). "A histone H3 methyltransferase controls epigenetic events required for meiotic prophase." Nature **438**(7066): 374-378.

Heil, C. S. and M. A. Noor (2012). "Zinc finger binding motifs do not explain recombination rate variation within or between species of Drosophila." PLoS One **7**(9): e45055.

Hubbard, T., D. Barker, E. Birney, G. Cameron, Y. Chen, L. Clark, T. Cox, J. Cuff, V. Curwen and T. Down (2002). "The Ensembl genome database project." Nucleic acids research **30**(1): 38-41.

Hubert, R., M. MacDonald, J. Gusella and N. Arnheim (1994). "High resolution localization of recombination hot spots using sperm typing." Nature genetics **7**(3): 420-424.

Hunter, C. M., W. Huang, T. F. Mackay and N. D. Singh (2016). "The genetic architecture of natural variation in recombination rate in Drosophila melanogaster." PLoS Genet **12**(4): e1005951.

Hunter, C. M., M. C. Robinson, D. L. Aylor and N. D. Singh (2016). "Genetic Background, Maternal Age, and Interaction Effects Mediate Rates of Crossing Over in Drosophila melanogaster Females." G3 (Bethesda) **6**(5): 1409-1416.

Hussin, J., M. H. Roy-Gagnon, R. Gendron, G. Andelfinger and P. Awadalla (2011). "Age-dependent recombination rates in human pedigrees." PLoS Genet **7**(9): e1002251.

Jackson, S., D. M. Nielsen and N. D. Singh (2015). "Increased exposure to acute thermal stress is associated with a non-linear increase in recombination frequency and an independent linear decrease in fitness in Drosophila." Bmc Evolutionary Biology **15**.

Jensen-Seaman, M. I., T. S. Furey, B. A. Payseur, Y. Lu, K. M. Roskin, C.-F. Chen, M. A. Thomas, D. Haussler and H. J. Jacob (2004). "Comparative recombination rates in the rat, mouse, and human genomes." Genome research **14**(4): 528-538.

Johnston, S. E., C. Berenos, J. Slate and J. M. Pemberton (2016). "Conserved Genetic Architecture Underlying Individual Recombination Rate Variation in a Wild Population of Soay Sheep (Ovis aries)." Genetics **203**(1): 583-598.

Kadri, N. K., C. Harland, P. Faux, N. Cambisano, L. Karim, W. Coppieters, S. Fritz, E. Mullaart, D. Baurain, D. Boichard, R. Spelman, C. Charlier, M. Georges and T. Druet (2016). "Coding and noncoding variants in HFM1, MLH3, MSH4, MSH5, RNF212, and RNF212B affect recombination rate in cattle." Genome Res **26**(10): 1323-1332.

Kappes, S. M., J. W. Keele, R. T. Stone, R. A. McGraw, T. S. Sonstegard, T. P. Smith, N. L. Lopez-Corrales and C. W. Beattie (1997). "A second-generation linkage map of the bovine genome." Genome Res **7**(3): 235-249.

Kauppi, L., A. J. Jeffreys and S. Keeney (2004). "Where the crossovers are: recombination distributions in mammals." Nature Reviews Genetics **5**(6): 413-424.

Kong, A., J. Barnard, D. F. Gudbjartsson, G. Thorleifsson, G. Jonsdottir, S. Sigurdardottir, B. Richardsson, J. Jonsdottir, T. Thorgeirsson, M. L. Frigge, N. E. Lamb, S. Sherman, J. R. Gulcher and K. Stefansson (2004). "Recombination rate and reproductive success in humans." Nat Genet **36**(11): 1203-1206.

Kong, A., D. F. Gudbjartsson, J. Sainz, G. M. Jonsdottir, S. A. Gudjonsson, B. Richardsson, S. Sigurdardottir, J. Barnard, B. Hallbeck, G. Masson, A. Shlien, S. T. Palsson, M. L. Frigge, T. E. Thorgeirsson, J. R. Gulcher and K. Stefansson (2002). "A high-resolution recombination map of the human genome." Nat Genet **31**(3): 241-247.

Kong, A., G. Thorleifsson, D. F. Gudbjartsson, G. Masson, A. Sigurdsson, A. Jonasdottir, G. B. Walters, A. Jonasdottir, A. Gylfason and K. T. Kristinsson (2010).

"Fine-scale recombination rate differences between sexes, populations and individuals." Nature **467**(7319): 1099-1103.

Kong, A., G. Thorleifsson, H. Stefansson, G. Masson, A. Helgason, D. F. Gudbjartsson, G. M. Jonsdottir, S. A. Gudjonsson, S. Sverrisson, T. Thorlacius, A. Jonasdottir, G. A. Hardarson, S. T. Palsson, M. L. Frigge, J. R. Gulcher, U. Thorsteinsdottir and K. Stefansson (2008). "Sequence variants in the RNF212 gene associate with genome-wide recombination rate." Science **319**(5868): 1398-1401.

Kono, H., M. Tamura, N. Osada, H. Suzuki, K. Abe, K. Moriwaki, K. Ohta and T. Shiroishi (2014). "Prdm9 polymorphism unveils mouse evolutionary tracks." DNA Research **21**(3): 315-326.

Lam, I. and S. Keeney (2015). "Nonparadoxical evolutionary stability of the recombination initiation landscape in yeast." Science **350**(6263): 932-937.

Lamb, N. E., S. L. Sherman and T. J. Hassold (2005). "Effect of meiotic recombination on the production of aneuploid gametes in humans." Cytogenet Genome Res **111**(3-4): 250-255.

Lesecque, Y., S. Glémin, N. Lartillot, D. Mouchiroud and L. Duret (2014). "The Red Queen model of recombination hotspots evolution in the light of archaic and modern human genomes."

Levine, R. P. (1955). "Chromosome Structure and the Mechanism of Crossing Over." Proceedings of the National Academy of Sciences of the United States of America **41**(10): 727-730.

Lipkin, S. M., P. B. Moens, V. Wang, M. Lenzi, D. Shanmugarajah, A. Gilgeous, J. Thomas, J. Cheng, J. W. Touchman, E. D. Green, P. Schwartzberg, F. S. Collins and P. E. Cohen (2002). "Meiotic arrest and aneuploidy in MLH3-deficient mice." Nat Genet **31**(4): 385-390.

Loftus, R. T., D. E. MacHugh, D. G. Bradley, P. M. Sharp and P. Cunningham (1994). "Evidence for two independent domestications of cattle." Proc Natl Acad Sci U S A **91**(7): 2757-2761.

Loidl, J. (1989). "Effects of Elevated-Temperature on Meiotic Chromosome Synapsis in Allium-Ursinum." Chromosoma **97**(6): 449-458.

Lu, S., C. Zong, W. Fan, M. Yang, J. Li, A. R. Chapman, P. Zhu, X. Hu, L. Xu and L. Yan (2012). "Probing meiotic recombination and aneuploidy of single sperm cells by whole-genome sequencing." Science **338**(6114): 1627-1630.

Lynn, A., T. Ashley and T. Hassold (2004). "Variation in human meiotic recombination." Annu Rev Genomics Hum Genet **5**: 317-349.

Ma, L., J. R. O'Connell, P. M. VanRaden, B. Shen, A. Padhi, C. Sun, D. M. Bickhart, J. B. Cole, D. J. Null, G. E. Liu, Y. Da and G. R. Wiggans (2015). "Cattle Sex-Specific Recombination and Genetic Control from a Large Pedigree Analysis." PLoS Genet **11**(11): e1005387.

Maddox, J. F., K. P. Davies, A. M. Crawford, D. J. Hulme, D. Vaiman, E. P. Cribiu, B. A. Freking, K. J. Beh, N. E. Cockett, N. Kang, C. D. Riffkin, R. Drinkwater, S. S. Moore, K. G. Dodds, J. M. Lumsden, T. C. van Stijn, S. H. Phua, D. L. Adelson, H. R. Burkin, J. E. Broom, J. Buitkamp, L. Cambridge, W. T. Cushwa, E. Gerard, S. M. Galloway, B. Harrison, R. J. Hawken, S. Hiendleder, H. M. Henry, J. F. Medrano, K. A. Paterson, L. Schibler, R. T. Stone and B. van Hest (2001). "An enhanced linkage map of the sheep genome comprising more than 1000 loci." Genome Res **11**(7): 1275-1289.

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**(12): 2827-2841.

Marklund, L., M. Johansson Moller, B. Hoyheim, W. Davies, M. Fredholm, R. K. Juneja, P. Mariani, W. Coppieters, H. Ellegren and L. Andersson (1996). "A comprehensive linkage map of the pig based on a wild pig-Large White intercross." Anim Genet **27**(4): 255-269.

Martin, H. C., R. Christ, J. G. Hussin, J. O'Connell, S. Gordon, H. Mbarek, J. J. Hottenga, K. McAloney, G. Willemsen, P. Gasparini, N. Pirastu, G. W. Montgomery, P. Navarro, N. Soranzo, D. Toniolo, V. Vitart, J. F. Wilson, J. Marchini, D. I. Boomsma, N. G. Martin and P. Donnelly (2015). "Multicohort analysis of the maternal age effect on recombination." Nat Commun **6**: 7846.

Modliszewski, J. L. and G. P. Copenhaver (2017). "Meiotic recombination gets stressed out: CO frequency is plastic under pressure." Curr Opin Plant Biol **36**: 95-102.

Morton, N. E. (1991). "Parameters of the human genome." Proc Natl Acad Sci U S A **88**(17): 7474-7476.

Munoz-Fuentes, V., A. Di Rienzo and C. Vila (2011). "Prdm9, a major determinant of meiotic recombination hotspots, is not functional in dogs and their wild relatives, wolves and coyotes." PLoS One **6**(11): e25498.

Myers, S., R. Bowden, A. Tumian, R. E. Bontrop, C. Freeman, T. S. MacFie, G. McVean and P. Donnelly (2010). "Drive against hotspot motifs in primates implicates the PRDM9 gene in meiotic recombination." Science **327**(5967): 876-879.

Nachman, M. W. and B. A. Payseur (2012). "Recombination rate variation and speciation: theoretical predictions and empirical results from rabbits and mice." Phil. Trans. R. Soc. B **367**(1587): 409-421.

Neff, M. W., K. W. Broman, C. S. Mellersh, K. Ray, G. M. Acland, G. D. Aguirre, J. S. Ziegle, E. A. Ostrander and J. Rine (1999). "A second-generation genetic linkage map of the domestic dog, Canis familiaris." Genetics **151**(2): 803-820.

O'Connell, J. R. (2013). MMAP: a comprehensive mixed model program for analysis of pedigree and population data. 63th Annual Meeting of The American Society of Human Genetics. **Oct 22–262013**.

Oliver, P. L., L. Goodstadt, J. J. Bayes, Z. Birtle, K. C. Roach, N. Phadnis, S. A. Beatson, G. Lunter, H. S. Malik and C. P. Ponting (2009). "Accelerated evolution of the Prdm9 speciation gene across diverse metazoan taxa." PLoS Genet **5**(12): e1000753.

Otto, S. P. and T. Lenormand (2002). "Resolving the paradox of sex and recombination." Nat Rev Genet **3**(4): 252-261.

Padhi, A., B. Shen, J. Jiang, Y. Zhou, G. E. Liu and L. Ma (2017). "Ruminant-specific multiple duplication events of PRDM9 before speciation." BMC evolutionary biology **17**(1): 79.

Paigen, K. and P. Petkov (2010). "Mammalian recombination hot spots: properties, control and evolution." Nature Reviews Genetics **11**(3): 221-233.

Paigen, K., J. P. Szatkiewicz, K. Sawyer, N. Leahy, E. D. Parvanov, S. H. Ng, J. H. Graber, K. W. Broman and P. M. Petkov (2008). "The recombinational anatomy of a mouse chromosome." PLoS genetics **4**(7): e1000119.

Pan, J., M. Sasaki, R. Kniewel, H. Murakami, H. G. Blitzblau, S. E. Tischfield, X. Zhu, M. J. Neale, M. Jasin, N. D. Socci, A. Hochwagen and S. Keeney (2011). "A

hierarchical combination of factors shapes the genome-wide topography of yeast meiotic recombination initiation." Cell **144**(5): 719-731.

Parvanov, E. D., P. M. Petkov and K. Paigen (2010). "Prdm9 controls activation of mammalian recombination hotspots." Science **327**(5967): 835-835.

Parvanov, E. D., P. M. Petkov and K. Paigen (2010). "Prdm9 controls activation of mammalian recombination hotspots." Science **327**(5967): 835.

Payseur, B. A. (2016). "Genetic Links between Recombination and Speciation." PLoS Genet **12**(6): e1006066.

Persikov, A. V. and M. Singh (2013). "De novo prediction of DNA-binding specificities for Cys2His2 zinc finger proteins." Nucleic acids research: gkt890.

Polani, P. E. and G. M. Jagiello (1976). "Chiasmata, meiotic univalents, and age in relation to aneuploid imbalance in mice." Cytogenet Cell Genet **16**(6): 505-529.

Ponting, C. P. (2011). "What are the genomic drivers of the rapid evolution of PRDM9?" Trends in Genetics **27**(5): 165-171.

Pratto, F., K. Brick, P. Khil, F. Smagulova, G. V. Petukhova and R. D. Camerini-Otero (2014). "Recombination initiation maps of individual human genomes." Science **346**(6211): 1256442.

Ptak, S. E., D. A. Hinds, K. Koehler, B. Nickel, N. Patil, D. G. Ballinger, M. Przeworski, K. A. Frazer and S. Paabo (2005). "Fine-scale recombination patterns differ between chimpanzees and humans." Nat Genet **37**(4): 429-434.

R, C. T. (2014). "R: A Language and Environment for Statistical Computing. Vienna, Austria: R Foundation for Statistical Computing."

Rose, A. M. and D. L. Baillie (1979). "The Effect of Temperature and Parental Age on Recombination and Nondisjunction in CAENORHABDITIS ELEGANS." Genetics **92**(2): 409-418.

Sandor, C., W. Li, W. Coppieters, T. Druet, C. Charlier and M. Georges (2012). "Genetic variants in REC8, RNF212, and PRDM9 influence male recombination in cattle." PLoS Genet **8**(7): e1002854.

Schwartz, J. J., D. J. Roach, J. H. Thomas and J. Shendure (2014). "Primate evolution of the recombination regulator PRDM9." Nature communications **5**.

Ségurel, L., E. M. Leffler and M. Przeworski (2011). "The case of the fickle fingers: how the PRDM9 zinc finger protein specifies meiotic recombination hotspots in humans." PLoS Biol **9**(12): e1001211.

Shifman, S., J. T. Bell, R. R. Copley, M. S. Taylor, R. W. Williams, R. Mott and J. Flint (2006). "A high-resolution single nucleotide polymorphism genetic map of the mouse genome." PLoS biology **4**(12): e395.

Singer, A., H. Perlman, Y. Yan, C. Walker, G. Corley-Smith, B. Brandhorst and J. Postlethwait (2002). "Sex-specific recombination rates in zebrafish (Danio rerio)." Genetics **160**(2): 649-657.

Singhal, S., E. M. Leffler, K. Sannareddy, I. Turner, O. Venn, D. M. Hooper, A. I. Strand, Q. Li, B. Raney, C. N. Balakrishnan, S. C. Griffith, G. McVean and M. Przeworski (2015). "Stable recombination hotspots in birds." Science **350**(6263): 928-932.

Smagulova, F., K. Brick, Y. Pu, R. D. Camerini-Otero and G. V. Petukhova (2016). "The evolutionary turnover of recombination hot spots contributes to speciation in mice." Genes & development **30**(3): 266-280.

Speed, R. M. (1977). "The effects of ageing on the meiotic chromosomes of male and female mice." Chromosoma **64**(3): 241-254.

Stern, C. (1926). "An Effect of Temperature and Age on Crossing-Over in the First Chromosome of Drosophila Melanogaster." Proc Natl Acad Sci U S A **12**(8): 530-532.

Stumpf, M. P. and G. A. McVean (2003). "Estimating recombination rates from population-genetic data." Nature Reviews Genetics **4**(12): 959-968.

Sugawara, S. and K. Mikamo (1983). "Absence of Correlation between Univalent Formation and Meiotic Nondisjunction in Aged Female Chinese Hamsters." Cytogenetics and Cell Genetics **35**(1): 34-40.

Szostak, J. W., T. L. Orr-Weaver, R. J. Rothstein and F. W. Stahl (1983). "The double-strand-break repair model for recombination." Cell **33**(1): 25-35.

Tortereau, F., B. Servin, L. Frantz, H. J. Megens, D. Milan, G. Rohrer, R. Wiedmann, J. Beever, A. L. Archibald, L. B. Schook and M. A. Groenen (2012). "A high density recombination map of the pig reveals a correlation between sex-specific recombination and GC content." BMC Genomics **13**: 586.

VanRaden, P. M. and A. H. Sanders (2003). "Economic merit of crossbred and purebred US dairy cattle." J Dairy Sci **86**(3): 1036-1044.

Venn, O., I. Turner, I. Mathieson, N. de Groot, R. Bontrop and G. McVean (2014). "Nonhuman genetics. Strong male bias drives germline mutation in chimpanzees." Science **344**(6189): 1272-1275.

Wang, J., H. C. Fan, B. Behr and S. R. Quake (2012). "Genome-wide single-cell analysis of recombination activity and de novo mutation rates in human sperm." Cell **150**(2): 402-412.

Wang, Z., B. Shen, J. Jiang, J. Li and L. Ma (2016). "Effect of sex, age and genetics on crossover interference in cattle." Sci Rep **6**: 37698.

Weng, Z. Q., M. Saatchi, R. D. Schnabel, J. F. Taylor and D. J. Garrick (2014). "Recombination locations and rates in beef cattle assessed from parent-offspring pairs." Genet Sel Evol **46**: 34.

Westergaard, M. and D. von Wettstein (1972). "The synaptinemal complex." Annu Rev Genet **6**: 71-110.

Winckler, W., S. R. Myers, D. J. Richter, R. C. Onofrio, G. J. McDonald, R. E. Bontrop, G. A. McVean, S. B. Gabriel, D. Reich, P. Donnelly and D. Altshuler (2005). "Comparison of fine-scale recombination rates in humans and chimpanzees." Science **308**(5718): 107-111.

Wong, A. K., A. L. Ruhe, B. L. Dumont, K. R. Robertson, G. Guerrero, S. M. Shull, J. S. Ziegle, L. V. Millon, K. W. Broman, B. A. Payseur and M. W. Neff (2010). "A comprehensive linkage map of the dog genome." Genetics **184**(2): 595-605.

Yang, F. and P. J. Wang (2009). "The Mammalian synaptonemal complex: a scaffold and beyond." Genome Dyn **5**: 69-80.

Ye, J., G. Coulouris, I. Zaretskaya, I. Cutcutache, S. Rozen and T. L. Madden (2012). "Primer-BLAST: a tool to design target-specific primers for polymerase chain reaction." BMC bioinformatics **13**(1): 1.

Zickler, D. (2006). "From early homologue recognition to synaptonemal complex formation." Chromosoma **115**(3): 158-174.

Zimin, A. V., A. L. Delcher, L. Florea, D. R. Kelley, M. C. Schatz, D. Puiu, F. Hanrahan, G. Pertea, C. P. Van Tassell and T. S. Sonstegard (2009). "A whole-genome assembly of the domestic cow, Bos taurus." Genome biology **10**(4): 1.

Zimin, A. V., A. L. Delcher, L. Florea, D. R. Kelley, M. C. Schatz, D. Puiu, F. Hanrahan, G. Pertea, C. P. Van Tassell, T. S. Sonstegard, G. Marcais, M. Roberts, P. Subramanian, J. A. Yorke and S. L. Salzberg (2009). "A whole-genome assembly of the domestic cow, Bos taurus." Genome Biol **10**(4): R42.

Zimin, A. V., D. R. Kelley, M. Roberts, G. Marcais, S. L. Salzberg and J. A. Yorke (2012). "Mis-assembled "segmental duplications" in two versions of the Bos taurus genome." PLoS One **7**(8): e42680.