Mammalian Meiotic Recombination: A Toolbox for Genome Evolution

Laia Capilla a  Montserrat Garcia Caldés a, b  Aurora Ruiz-Herrera a, b

a Genome Integrity and Instability Group, Institut de Biotecnologia i Biomedicina, and b Departament de Biologia Cellular, Fisiologia i Immunologia, Universitat Autònoma de Barcelona, Cerdanyola del Vallès, Spain

Key Words
Crossovers · Evolution · Meiosis · PRDM9 · Recombination

Abstract
Meiotic recombination is a process that increases genetic diversity and is fundamental for sexual reproduction. Determining by which mechanisms genetic variation is generated and maintained across different phylogenetic groups provides the basis for our understanding of biodiversity and evolution. In this review, we go through different aspects of this essential phenomenon, paying special attention to mammals. We provide a comprehensive view on the organization of meiotic chromosomes and the mechanisms involved in the formation and genomic distribution of recombination hotspots, focusing on the factors influencing the formation and repair of the massive amount of self-induced DNA breaks in early stages of meiosis. At the same time, we discuss the genetic and mechanistic factors that influence recombination landscapes in mammals, as reflected by several layers of regulation. These factors include the selective forces that affect the DNA sequence itself, which can be modulated by genome reshuffling and the evolutionary history of each taxon, and the forces that control how the DNA is packaged into chromosomes during meiosis.

Since Janssens’s initial chiasmatype theory, based on the first observation of chiasmata [Janssens, 1909; reviewed in Koszul et al., 2012], much has been learned about the mechanisms, characteristic features, and evolution of meiotic recombination. Homologous chromosomes (homologs) exchange genetic information through recombination, a process that increases genetic diversity in organisms with sexual reproduction. During prophase I of meiosis, both maternal and paternal homologous chromosomes engage in a chromosomal interplay that results in new combinations of alleles that will be inherited by the following generation. Moreover, meiotic recombination provides physical connections between homologs that are essential for faithful chromosomal pairing and segregation [Hassold and Hunt, 2001; Handel and Schimenti, 2010]. All of these aspects are relevant for our understanding of genetic variation and evolutionary change. Recent investigations have revealed the importance of recombination modifiers (i.e., genetic and mechanistic factors) in modeling genomic architecture and evolution. In this review, we focus on different aspects of meiotic recombination, paying special attention to mammals. The general view of the organization of meiotic chromosomes and the mechanisms involved in the formation and genomic distribution of recombination hotspots is discussed.
Recombination and the Repair of Double-Strand Breaks

Organization of Meiotic Chromosomes: The Importance of Chromosomal Axes

Recombination between homologs is an intricate, highly regulated process in which chromosomal movements are fine-tuned with molecular events during prophase I, a long phase that is divided into 4 differentiated substages: leptonema, zygonema, pachynema, and diplonema. At leptonema, chromatids start to condense, and homologs begin to pair, a process that is facilitated by telomere clustering at the nuclear envelope [Reig-Viader et al., 2016 and references therein]. Homologous chromosomes initiate synopsis at zygonema, which is completed at pachynema, along with recombination. Later on, at diplonema, homologs set themselves apart from each other to complete meiosis. During this process, crucial mechanisms take place, which, in turn, determine the successful progression of meiosis: the assembly of the chromatin into chromosomal axes and the formation and repair of double-strand breaks (DSBs) [Keeney et al., 1997; Romanienko and Camerini-Otero, 2000; Longhese et al., 2009] (fig. 1).

All throughout prophase I, chromosomes are organized into large DNA loops attached to the lateral elements of the synaptonemal complex (SC), a proteinaceous scaffold which consists of a tripartite structure with axial elements composed of SC proteins 2 and 3 (SYCP2 and SYCP3) that are connected by filament proteins with overlapping central elements (SYCP1) [Heyting, 1996; Page and Hawley, 2004; Henderson and Keeney, 2005]. It is known that the repair of DSBs occurs in the context of this axial chromosomal structure [reviewed in de Massy, 2013 and Lam and Keeney, 2014] (fig. 1). Therefore, although DSBs occur genome-wide, their formation and repair is mediated by a set of proteins from the DNA damage response mechanism that are physically and functionally associated with the chromosome structural axes [Zickler and Kleckner, 1999; de Boer and Heyting, 2006; Storlazzi et al., 2010; de Massy, 2013; Lam and Keeney, 2014]. However, the mechanism(s) by which this translocation (i.e., the recruitment of the 3' strand overhangs to the SC) occurs is not well understood. DNA loops are organized in the SC, and their density along the chromosomal length is highly conserved [Zickler and Kleckner, 1999; Kleckner, 2006]. As a consequence, variation in SC length is expected to result in variation in both the number and length of DNA loops that are anchored to the SC, which is consistent with previous observations of an inverse correlation between loop size and axis length in mice [Zickler and Kleckner, 1999; Kleckner et al., 2003; Kauppi et al., 2011]. Furthermore, a close interplay has been described among the assembly of SCs, the organization of DNA loops, and the formation of DSBs [Kleckner et al., 2003; Wang et al., 2015]. This fact has important implications, since it suggests that DSBs are necessary not only for the formation of crossovers (COs) but also for homolog pairing and synapsis [De Vries et al., 1999; Baudat et al., 2000; Romanienko and Camerini-Otero, 2000; Kleckner et al., 2003; Kauppi et al., 2013].

Molecular Events Involved in the Formation and Repair of DSBs

The formation of DSBs is catalyzed by SPO11, an endonuclease orthologous to the subunit A of TopoVI DNA topoisomerase (type IIB DNA topoisomerase). This protein is highly conserved among eukaryotes with apparently no or little sequence specificity [Keeney et al., 1997; Malik et al., 2008; Lam and Keeney, 2014]. In mice and humans, SPO11 presents 2 isoforms (SPO11a and SPO11β), SPO11β being responsible for the majority of induced meiotic DSBs. Recent results suggest that SPO11β and the newly described TOP6BL (for TopoVIB-like) protein work as a heterotetramer resembling ancestral topoisomerases [Bouuaert and Keeney, 2016; Robert et al., 2016]. Along with SPO11, other proteins have been reported as being involved in the formation of DSBs, such as MEI4, REC114, and HORMAD1 in the case of mice [Kumar et al., 2010; de Massy, 2013; Lam and Keeney, 2014]. Such proteins catalyze the repair of DSBs on chromosomal axes [de Massy, 2013] (fig. 1).

In an orchestrated manner, the formation of DSBs induces the phosphorylation of histone H2AX on serine 139 (γH2AX) by proteins ATM (ataxia telangiectasia mutated) and ATR (ATM-Rad3-related) [Rogakou et al., 1998; Burma et al., 2001; Kuo and Yang, 2008], which recruit different proteins involved in DNA damage response. After an extensive resection of DSBs, RPA (replication protein A) binds to the 3' strand overhangs [He et al., 1995]. RPA is then displaced by RAD51 (radiation sensitive 51) and/or DMC1 (disrupted meiotic cDNA 1) [Pittman et al., 1998; Yoshida et al., 1998], which form nucleoprotein filaments that catalyze strand invasion. All of these proteins catalyze DNA intermediates of meiotic recombination, with 2 possible final outcomes, COs or non-COs (NCOs), which are catalyzed by different DSB repair pathways. In the case of CO formation, a D-loop is created capturing the second 3' end of the homologous chromosome, and after DNA synthesis and ligation, a
A double Holliday junction is produced, with heteroduplex DNA flanking the DSB site [Collins and Newlon, 1994; Schwacha and Kleckner, 1995] (fig. 1). This structure is then resolved by the cleavage and ligation of strands of identical polarities, which generates COs. Thus, the formation of COs involves the invasion of the homologous chromosome by the single end created by DSBs, leading to the formation of chiasmata connecting both homologous chromosomes. NCO products, on the other hand, are thought to be generated by either a double Holliday junction or by a mechanism called synthesis-dependent strand annealing [McMahill et al., 2007]. In synthesis-dependent strand annealing, the D-loop is disassembled by displacement of the newly synthesized strand, which anneals with the other DSB end. DSB repair is completed with DNA synthesis by a DNA polymerase (using the homologous chromosome as a template).
mologous sequence as template) followed by a ligation to the original strand. In this case, the mismatch repair of the heteroduplex generates gene conversion without resulting in a CO. Later on, at pachynema, events that will become COs are resolved by the repair pathway directed by proteins MSH4 and MSH5 (MutS homologs 4 and 5, that appear earlier at zygonema) and MLH1 and MLH3 (MutL homologs 1 and 3) [Baker et al., 1996; Kneitz et al., 2000; Santucci-Darmanin et al., 2000; Lipkin et al., 2002; Snowden et al., 2004].

The mismatch repair of heteroduplexes can lead to either gene conversion or restoration. However, in most organisms, the number of DSBs largely exceeds the number of resulting COs, indicating that the majority of DSBs are resolved as NCOs. In mice, about one-tenth of initial DSBs are finally processed into COs [Koeleher et al., 2002a; Moens et al., 2007; Murakami and Keeney, 2008; Cole et al., 2012]. Importantly, CO assurance in mammals is carefully controlled by homeostasis [Cole et al., 2012]. Whereas the cell-to-cell variation of RAD51 and DMC1 foci (a proxy of early recombination events) is relatively high and dynamic (between 31.8 and 36.5% of variation), the final numbers of COs (measured as the number of MLH1 foci) are relatively constant among cells (only 11.5% of variation) [Cole et al., 2012].

Since only a small fraction of DSBs are eventually processed as COs, a highly regulated genetic control determines both CO homeostasis and chromosomal distribution. In this way, if a CO occurs in a certain position, the probability for a new CO to take place nearby increases with chromosomal length. As a consequence, COs tend to follow an evenly spaced distribution across chromosome axes [Jones, 1967; Kleckner et al., 2003; Wang et al., 2015]. Importantly, this CO interference is influenced by the physical distance along the chromosomal axes (micrometers) rather than the genomic (Mb) or genetic distance (cM) [Wang et al., 2015]. However, not all COs are subject to interference, leading to recognition of interfering (class I) and non-interfering (class II) COs in different organisms [Hollingsworth and Brill, 2004; Phadnis et al., 2011]. Non-interfering COs are Mus81-Mms4 dependent and distribute themselves randomly along the chromosomes independent of each other, whereas interfering COs have been found to be distributed according to a gamma distribution. In mice, most COs manifest interference and are controlled by proteins Msh4-Msh5 [Berchowitz and Copenhaver, 2010], although some Mus81 activity has been detected during meiosis [Holloway et al., 2008]. Despite the evolutionary rationale of CO interference is still unknown, spaced COs ensure faithful chromosomal segregation and might facilitate linkage of functionally related genes [Wang et al., 2015 and references therein].

Methodological Approaches to the Study of Recombination

In order to understand the mechanisms involved in recombination, different approaches have been developed over the years to estimate the number and genome distribution of COs. These methodologies can be classified as direct (i.e., direct measure of recombination events or DSB sites on meiotic cells) or indirect (i.e., estimation of recombination rates) (fig. 2).

Direct analyses are based on the detection of either COs or their final products, the chiasmata, which are visible cytogenetically in meiocytes in later stages of the first meiotic division. Initial recombination studies using direct approaches were based on the detection of chiasmata by Giemsa staining in metaphase I [Templado et al., 1976; Rasmussen and Holm, 1984; Hultén, 1990; Moens and Spyropoulos, 1995; Hassold et al., 2000; Castiglia and Capanna, 2002]. The subsequent development of fluorescent antibody technologies initiated a novel approach that quantitatively improved recombination analysis. This technique permitted the in situ identification of the different proteins involved in the meiotic process in meiocytes, such as DSB repair-associated proteins (RAD51, DMC1, among others), DNA mismatch repair proteins (i.e., MLH1, which localize type I COs), as well as the proteins involved in SC formation (SYCP1, SYCP2, and SYCP3; fig. 2). This methodology is still broadly used in a wide range of organisms in order to understand the dynamics of the meiotic process and the proteins involved [Anderson et al., 1999; Froenicke et al., 2002; Lynn et al., 2002; Al-Jaru et al., 2014; Capilla et al., 2014]. However, the immunofluorescent localization of meiotic proteins presents some limitations that have prevented their widespread use in non-model species. First, it requires the use of testicular or fetal tissue, and, second, it presents a lower resolution (Mb scale) when compared to indirect techniques (see below). Finally, recombination rates estimated using cytological techniques are not linked to the DNA sequence and, therefore, cannot be used to estimate intergenic recombination rates.

More recently, new approaches have been developed to directly estimate recombination rates in germ cells at the finest scale. These include the detection of DNA re-
gions linked to proteins associated with DSB repair (such as DMC1 or RAD51) by combining chromatin immunoprecipitation (ChIP) assays with high-throughput sequencing, known as ChIP-sequencing [Smagulova et al., 2011; Brick et al., 2012] or the SPO11-oligonucleotide sequencing [Pan et al., 2011; Fowler et al., 2014]. In the latter case, SPO11 remains covalently linked to the 5′ ends of broken DNA, being released later on by an endonuclease complex. These SPO11 oligonucleotides can be isolated and sequenced.

Fig. 2. Examples of different methodological approaches for the detection of recombination events. 

a Immunofluorescence. At pachynema, the position of COs along the chromosomal axes can be detected by immunodetection of the MLH1 protein in spermatocyte spreads. This image shows a spread of a lemur (Lemur catta) spermatocyte stained with anti-SYCP3 (lateral element of the SC), anti-CEN (marker of centromeres), and anti-MLH1 (marker of COs). b SPO11-oligonucleotide sequencing quantitatively maps DSBs across the genome at the nucleotide resolution [Pan et al., 2011; Fowler et al., 2014]. During DSB formation, SPO11 remains covalently linked to the 5′ ends of broken DNA, being released later on by an endonuclease complex. c Estimates of LD. Fine-scale recombination maps can be obtained by genotyping SNPs (markers) across the entire genome in a segregating population of individuals. Here, recombination events through SNP genotyping over a population are represented. This results in LD-based genetic maps. Recombination is normally measured in units of $\rho = 4N_e r$, where $\rho$ refers to the population recombination rate, $N_e$ refers to the effective population size, and $r$ to the per-generation recombination rate. In this approach, $r$ can depend on local sequence or DNA structure, whereas $\rho$ depends on demographic history.
nome at the nucleotide resolution [Neale et al., 2005; Pan et al., 2011; Fowler et al., 2014].

Indirect methodologies, on the other hand, estimate recombination rates based on the analysis of DNA polymorphisms and include different approaches, such as pedigree analysis, linkage disequilibrium (LD) analysis, and sperm typing. Genotyping (using either microsatellites or single nucleotide polymorphisms, SNPs) a progeny or gametes from pedigree analyses on controlled crosses directly quantifies products of meiosis resulting in genetic linkage maps (normally expressed in centimorgans, cM) [Haldane, 1922]. In the case of humans, this approach has been traditionally based on family studies where each offspring can provide information on the product of 1 paternal and 1 maternal meiotic event. Thus, recombination rates can be estimated by examining the parent-to-offspring transmission of alleles at specific loci. Using this approach, genetic maps have been broadly established at different resolutions in several mammalian species, allowing accurate estimates of the recombination fraction between markers separated by several Mb [e.g., Robinson, 1996; Broman et al., 1998; Hawken et al., 1999; Yu and Feingold, 2002; Dumont et al., 2011].

But, although pedigree analysis has provided important advances in detecting the distribution of recombination hotspots genome-wide, it relies directly on the resolution of the genetic maps, and consequently, even strong hotspots (where recombination rates can be thousands of times greater than the genome average) could remain undetected. Moreover, this approach permits the study of a small number of meiotic cells, underscoring the detection of rare variants and limiting the potential resolution of recombination maps that resulted from these analyses (i.e., expected resolutions of 0.5–2 cM in humans; Clark et al. [2010]). To overcome these initial limitations, LD-based genetic maps were developed by applying population genetics (taking into account parameters such as effective population size, \( N_e \), and demography) to relate local levels of LD to local recombination rates using large-scale surveys [reviewed in Clark et al., 2010]. That is, regions with extensive LD would correspond to regions with low recombination rates. The implementation of such approaches results in indirect estimations of recombination maps at a greater resolution, normally including larger samples. The parameter considered is \( \rho = 4N_e r \) (fig. 2), which provides estimates of population recombination rates. LD analyses are therefore based on the detection of SNPs to estimate historical recombination events, and differ from traditional linkage mapping methods in that marker associations are analyzed in populations of unrelated individuals. Such associations reflect the long evolutionary history of the chromosomal region (or the entire genome, in the case of genome-wide screens) and have been used in mammals in several studies [Ptak et al., 2005; Clark et al., 2010; Kong et al., 2010; Johnston et al., 2016; Stevison et al., 2016]. Sperm typing analysis, on the other hand, is based on the use of spermatozoa from an individual whose somatic genotype is known [Hubert et al., 1994; Jeffreys et al., 2001; Cullen et al., 2002; Jeffreys and Neumann, 2002; Yauk et al., 2003]. This technique has been used in humans and mice and permits the analysis of large amounts of cells from a single individual although the resolution can be a limitation, depending on marker density [Wu et al., 2010].

Genetic and Epigenetic Marks of DSBs and Recombination Hotspots

The fact that germ cells induce genome-wide DSBs in each meiotic cycle begs for the existence of a programmed pathway that permits the faithful formation and repair of DNA breaks (i.e., between 200–300 DSBs in mice and around 150 DSBs in humans [Barlow et al., 1997; Cole et al., 2012]). Since failure in recombination repair can result in a prevalence of aneuploidy, infertile phenotypes, and genetic disorders [Sasaki et al., 2010; Nagaoka et al., 2012], the spatial distribution of DSBs, and therefore of COs, is highly regulated. Depending on the species (either yeast, mouse, or human), local determinants of DSB formation can include specific DNA motifs, transcription factors, epigenetic modifications, the presence/absence of nucleosomes, and the general organization of the meiotic chromosome [Baudat et al., 2013; de Massy, 2013; Lam and Keeney, 2014; Hunter, 2015].

The initial description of the PRDM9 (PR domain 9) protein unveiled the importance of specific DNA motifs and epigenetic modifications in determining recombination hotspots [Baudat et al., 2010; Myers et al., 2010; Parvanov et al., 2010; Brick et al., 2012]. The mouse Prdm9 gene (also known as Hst1), was first defined as a hybrid sterility locus responsible for hybrid sterility in males between the house mouse subspecies Mus musculus and M. m. domesticus [Forejt, 1996]. Initially mapped in mouse chromosome 17 [Forejt et al., 1991; Gregorová and Forejt, 2000], the gene corresponding to this locus, also known as Meisetz [meiosis induced factor containing a PR/SET domain and zinc finger (ZnF) motif], was found to code for PRDM9 [Mihola et al., 2009]. The PRDM9
protein presents a KRAB (Krueppel-associated box) domain in the N-terminal region which behaves as a transcriptional repressor domain [see Urrutia, 2003 and references therein], a SET domain with histone 3 lysine 4 (H3K4) methyltransferase activity, and an array of C2H2 ZnF repeats in the C-terminal domain that recognizes a DNA-specific repeat sequence (fig. 3). The PRDM9 sequence and allelic diversity has been described in several mammalian species, such as humans [Baudat et al., 2010; Jeffreys et al., 2013], non-human primates [Oliver et al., 2009; Schwartz et al., 2014], rodents [Oliver et al., 2009; Baudat et al., 2010; Capilla et al., 2014; Kono et al., 2014], and equids [Steiner and Ryder, 2013], among others.

PRDM9 is expressed only during early stages of meiosis in testes and ovaries [Hayashi et al., 2005], and it was first described as a marker of transcription recognition sites [Brick et al., 2012]. The mechanism by which PRDM9 controls the distribution pattern of meiotic recombination is still not fully understood, but it has been proposed that it occurs due to the recognition of specific DNA motifs by its C2H2 ZnF domain (fig. 3), specific sequences that vary among species [Baudat et al., 2010].
humans, for example, the degenerated 13-bp motif (CCnCCnTnnCCnC) was found to be associated with nearly 40% of the described hotspots [Baudat et al., 2010; Myers et al., 2010]. Once the DNA motif is recognized, the SET domain leaves an epigenetic signature (H3K4me3) and recruits SPO11 by a mechanism yet to be discovered (fig. 3). Therefore, PRDM9 would target the DSB repair machinery to specific sites, although it is not required for the formation of DSBs themselves since, in the absence of PRDM9, most recombination is initiated at promoter and other genomic regions with PRDM9-independent H3K4 trimethylation [Brick et al., 2012].

What is evolutionarily relevant about the PRDM9 protein is that the ZnF array recognizes specific DNA motifs characterized by a high variability in both number and sequence [Oliver et al., 2009; Baudat et al., 2010]. These differences in size and sequence are translated into a variety of different alleles across mammalian species [Oliver et al., 2009; Buard et al., 2014; Capilla et al., 2014; Kono et al., 2014; Pratto et al., 2014; Schwartz et al., 2014]. In primates (humans and great apes), the number of ZnF repeats can range from 6 to 19 [Berg et al., 2011; Jeffreys et al., 2013; Pratto et al., 2014; Schwartz et al., 2014], whereas in different Mus species and subspecies the variability ranges from 7 to 17 ZnF repeats [Buard et al., 2014; Capilla et al., 2014; Kono et al., 2014]. Importantly, modifications of the ZnF domain can be translated into a redistribution of recombination sites [Baudat et al., 2010]. This effect was initially described by Brick et al. [2012] by comparing mice with 2 different Prdm9 alleles, which shared only 1.1% of recombination hotspot positions; this reduced overlapping was attributed to the differences of the Prdm9 sequence. It was proposed that single amino acid changes within the ZnF domain could lead to new Prdm9 variants with novel DNA-binding specificity that, in turn, could potentially create a new genome-wide distribution of hotspots [Brick et al., 2012]. Along with this view, studies in humans [Pratto et al., 2014] have determined that different alleles have affinities to different DNA motifs, resulting in different distributions of DSBs. However, the implication of additional factors other than PRDM9 in modulating the frequencies of recombination initiation cannot be ruled out at this stage. Likewise, different PRDM9 alleles (associated with different DNA motifs) have been identified in chimpanzees [Auten et al., 2012; Schwartz et al., 2014] and gorillas [Schwartz et al., 2014].

More recently, the generation of 26 DSB high-resolution maps (derived from DMC1 ChIP-seq experiments) in 6 inbred mice strains and their corresponding F1 hybrids has provided new clues on the evolutionary turnover of recombination initiation hotspots in different Prdm9 allelic backgrounds [Smagulova et al., 2016]. First, the overlap between DSB initiation sites in the 6 different mice strains ranged from 1.1 to 34%, based on ZnF domain sequence similarity [Smagulova et al., 2016]. More importantly, novel DSB initiation sites are generated in hybrids which are heterozygous for different Prdm9 alleles in a frequency that ranges from 2 to 35%, most probably attributed to the appearance of novel polymorphisms in the parental genomes. That is, novel DSB initiation sites in hybrids are generated at genomic regions not previously used in either parental genome [Smagulova et al., 2016].

As we have seen, the genomic distribution of recombination hotspots can be affected by a complex interplay between the cis-acting DNA sequence at hotspots (or the gene conversion bias of CO repairing mechanisms) and the trans-acting factor that binds to that DNA sequence (i.e., PRDM9). Different studies have identified that 1 of the 2 chromatids presents a higher probability of undergoing gene conversion, where the donor chromatid is used as the template to repair the DNA sequence lost from the active partner in the course of creating the DSB [Baudat and de Massy, 2007; Paigen et al., 2008; Berg et al., 2011]. Therefore, this biased gene conversion predicts that hotspots should undergo evolutionary erosion due to the mutagenic effect of meiotic recombination. Thus, if hotspots drive themselves to extinction, it is unlikely that recombination persists in the same region, leading to a phenomenon known as the “hotspot paradox” [Boulton et al., 1997; Coop and Myers, 2007; Baker et al., 2015]. However, the PRDM9 function brings a solution to this paradox, as its rapid evolutionary change can overcome hotspot loss by undergoing mutations altering its ZnF array and, thus, changing the genome-wide distribution of hotspots. However, the question how Prdm9 is evolving so fast is still under discussion. One of the current explanations is that Prdm9 is under positive selection, associated with the rapid evolution of its binding sites [Coop and Myers, 2007; Oliver et al., 2009; Thomas et al., 2009; Myers et al., 2010]. In fact, both SNPs and short indels can induce the appearance of novel hotspots, at an estimated rate of 0.7–1.4 DSBs every 1,000 generations, as recent studies have shown in mice [Smagulova et al., 2016]. However, and despite the importance of the Prdm9 gene in modulating recombination, the existence of species such as dogs and finches [Axelsson et al., 2012; Muñoz-Fuentes et al., 2015; Singhal et al., 2015] lacking either the gene or a functional PRDM9 protein suggests the exis-
ence of additional, yet to be discovered, recombination effectors [Dzur-Gejdosova et al., 2012; Bhattacharyya et al., 2013; Turner and Harr, 2014; Balcoa et al., 2016].

**Variation of Recombination Rates within Genomes and among Species**

Since meiotic recombination strongly influences genome evolution, mammalian recombination landscapes can be considered as a reflection of the selective forces that affect the DNA sequence itself (determined by population genetics and the evolutionary history of each taxon), the chromosomal/genome distribution of COs, and how the DNA is packaged into chromosomes during meiosis. Therefore, it is important to take into consideration the working level of resolution when analyzing variation of recombination within and among species, which can span the whole chromosome up to the finest scale (i.e., single base pairs).

**Variability at the Chromosomal Level**

At the chromosomal level, the presence of 1 single CO per chromosome ensures proper disjunction of homologous chromosomes in the first meiotic division. Early studies [Dutrillaux, 1986] reported a correlation between the number of chiasmata and the haploid number of chromosome arms followed by subsequent studies in a wide range of mammalian species [Pardo-Manuel de Villena and Sapienza, 2001; Segura et al., 2013]. In this context, substantial progress has been made in elucidating the mechanisms that control both the formation and genome-wide distribution of COs. MLH1 recombination maps have been constructed for a wide variety of mammalian species, including humans and non-human primates [Sun et al., 2005; Codina-Pascual et al., 2006; Has-sold et al., 2009; García-Cruz et al., 2011; Gruhn et al., 2013, 2016; Baier et al., 2014], rodents [Froenicke et al., 2002; Dumont and Payne, 2011; Baier et al., 2014; Bacheva et al., 2014; Capilla et al., 2014], pigs [Muñoz et al., 2012; Mary et al., 2014, 2016], bovids [Vozdova et al., 2013, 2014, 2016; Sebestova et al., 2016], and other eutherian groups such as afrotherian species, carnivorans, and insectivorans [Borodin et al., 2008; Segura et al., 2013; Muñoz-Fuentes et al., 2015].

Despite the minimum requirement of 1 single CO per chromosome, the total number and distribution of COs in a specific chromosome can be influenced by several factors, such as chromosomal size and gender [Lynn et al., 2004; Paigen and Petkov, 2010]. Regarding the former, studies in eutherian mammals have shown that larger chromosomes tend to accumulate larger numbers of COs, and a chromosome generally presents at least 1 CO to ensure faithful chromosomal segregation [Sun et al., 2005; Farré et al., 2013; Segura et al., 2013]. However, the reasons behind the generation of a massive amount of self-induced DNA breaks (i.e., between 200–300 DSBs in mice) to produce a final outcome of at least 1 CO per chromosomal arm are not fully understood. Recent studies have suggested that homolog pairing and not only the minimal requirement of COs influences the formation of DSBs [Kauppi et al., 2013]. As discussed earlier, meiotic chromosomes are organized in large DNA loops attached to the lateral elements of the SC (fig. 1), the spacing of which is evolutionarily conserved [Kleckner, 2006; Zickler and Kleckner, 2015]. Thus, the repair of DSBs occurs in the context of the axial chromosome structure [reviewed in de Massy, 2013 and Lam and Keeney, 2014]. This fact has important implications, since it suggests that DSBs are necessary not only for the formation of COs but also for homolog pairing, at least in mice [Kauppi et al., 2013]. In fact, when focusing on the pseudoautosomal region (PAR) of the sex chromosomes in mice, Kauppi et al. [2011] reported 2 interesting observations. First, the PAR axes were disproportionately long, relative to DNA length, and showed short chromatin loops. And second, the DNA arranged into smaller loops in the PAR experiences more than 10 times more DSBs than do autosomes [Kauppi et al., 2011]. Therefore, large amounts of DSBs are necessary to ensure that small chromosomes, as in the case of the small pairing portion in sex chromosomes, find and synapse with their homologs in early stages of meiosis. In fact, it has been proposed that variations in SC length can determine the total number of COs under conditions of constant total DNA length [Kleckner et al., 2003]. This is well exemplified by the differences found in the recombination landscape in human males and females [Lynn et al., 2002]. Under the same conditions of DNA content, human females have longer SC lengths (approximately a 1.6-fold increase) and a higher number of COs per chromosome than do males [Lynn et al., 2002; Baier et al., 2014]. This correlation between COs and SC length in mammals has also been shown in mice [Anderson et al., 1999] and other eutherian mammals [Segura et al., 2013]. Further evidence in mice sex chromosomes indicates that not only the chromosomal axis, but also the number and length of DNA loops anchored to the SC might influence the number of DSBs induced [Kauppi et al., 2011; Gruhn et al., 2016]. Chromosomes with longer DNA loops (derived from shorter SC axes) would show a...
reduced (low) number of DSBs in the early stages of meiosis. Conversely, shorter DNA loops (and therefore larger SCs) would offer substrate for the formation of a large number of DSBs.

Moreover, initial cytological studies already recognized a remarkable variation in recombination rates within and among individuals in human and mouse [Koehler et al., 2002b; Lynn et al., 2002]. In humans with normal spermatogenesis, significant inter-individual variation from 40 to 60 MLH1 foci per cell has been described [Lynn et al., 2002; Gruhn et al., 2013], with differences of nearly 15%. The same trend has been reported for inbred laboratory mice, where strains with low (CAST/Ei), medium (A/J), and high (C57/BL6, SPRET/Ei) levels of recombination were identified [Koehler et al., 2002b; Lynn et al., 2002; Baier et al., 2014]. In fact, in wild-type C57/BL6 mice, a coefficient of variation (at 95% confidence interval) of 11.5% for MLH1 foci has been reported [Cole et al., 2012]. The same pattern of variability has been described for primates [Hassold et al., 2009], equids [Al-Jaru et al., 2014], ruminants [Sebestova et al., 2016], and other eutherian mammals [Segura et al., 2013]. Importantly, this variability is also observed when analyzing the numbers of DSBs initiated in early stages of meiosis (exemplified as RAD51 and DMC1 foci [Baier et al., 2014]). Therefore, there is a cell-to-cell variation in recombination rates assumable for the individual to ensure fertility, resulting in a unique genetic puzzle that is inherited to warrant genetic diversity. The determinants of this variation are not fully understood, but chromatin conformation (i.e., the organization of DNA loops) might play a major role [Gruhn et al., 2013; Baier et al., 2014; Wang et al., 2015].

Variation of Fine-Scale Recombination Maps

Despite the cell-to-cell and intraindividual variability observed, closely related species tend to show similar average rates of recombination in homologous chromosomes when comparing recombination rates on a broad scale (Mb) [Hassold et al., 2009; Garcia-Cruz et al., 2011; Auton et al., 2012]. However, similarities disappear on the finest scale (kb) [reviewed in Smukowski and Noor, 2011], highlighting the importance of considering the level of resolution when analyzing variation of recombination within and among species (fig. 2). The analysis of recombination rates on the finest scale has revealed that, in mammals, both the DSBs induced in early stages of meiosis (i.e., ChIP-seq of recombination proteins) and the eventual sites of exchange of genetic information (i.e., LD estimates) are not randomly distributed genome-wide. In fact, there are regions of the genome considered to be hotspots and coldspots of recombination and DSB formation, hotspots being defined as genomic sites where the increase in the recombination rate is significant as compared to the average whole-genome recombination rate.

In humans, several recombination maps have been generated based on different populations and methodological approaches. These include pedigree-based maps [Kong et al., 2002; Coop et al., 2008] and genome-wide LD-based maps using either HapMap data [Myers et al., 2005] or the 1000-genome project. Using these approaches, about 23,000 recombination hotspots have been described across the human genome [Myers et al., 2005, 2010]. From these studies, it is known that human hotspots are normally 1–2 kb in size and spaced 50–100 kb apart, accounting for no more than 20% of the genome. Centromeres are normally recombination deserts, whereas in (sub)telomeres recombination rates increase. Mirroring what has been described in mice [Smagulova et al., 2011; Khil et al., 2012], hotspots are localized in genic and intergenic regions likewise. In the case of other primate species such as chimpanzee, initial recombination maps were inferred from genome-wide SNP data in 10 non-related individuals [Auton et al., 2012]. These maps estimate the location of recombination events in the progeny [for a review, see Lynn et al., 2004] and reflect the integration of population-level processes over several generations. Thus, this approach provides a historical view of recombination events, incorporating data on population growth and natural selection, among others [Clark et al., 2010]. Using this approach, it has been found that recombination increases around transcription start and end (about 20%, on average) and decreases within the transcribed region (about 30%, on average) [Auton et al., 2012]. Moreover, it is known that recombination hotspots are associated with GC content, high recombination rates being associated with GC-rich regions [Kauppi et al., 2004; Myers et al., 2005; Buard and de Massy, 2007; Coop and Przeworski, 2007].

More recent studies taking advantage from the Great Ape Genome Project [Prado-Martinez et al., 2013] described new LD-based recombination maps from 10 chimpanzees, 13 bonobos, and 15 gorillas defining a total of 8,037, 10,704, and 22,012 hotspots, respectively [Stevenson et al., 2016]. When comparing the degree of conservation between great apes and human recombination maps, a low degree of hotspots has been conserved on the finest scale, even between chimpanzees and bonobos, accounting for a rapid hotspot turnover during evolution.
Evolvability of Recombination Rates

A large body of both theoretical and empirical work has focused on uncovering the selective forces that are instrumental in driving recombination [for a review, see Butlin, 2005; Smukowski and Noor, 2011; Ortiz-Barrientos et al., 2016]. As discussed earlier, fine-scale recombination maps change rapidly within species, and this can be greatly influenced by population genetics and the evolutionary history of each taxon. In fact, recent studies have suggested the existence of a phylogenetic effect in recombination rates by indicating that closely related species tend to have similar average rates of recombination [Dumont and Payseur, 2008, 2011; Segura et al., 2013]. Moreover, this phylogenetic component in recombination rates appears to be directional, strongly punctuated, and subject to selection [Segura et al., 2013].

Homologs exchange genetic information through recombination, a process that generates novel gene combinations and increases genetic diversity. But, at the same time, species need to find a balance between the mutagenic effect of recombination and its positive effects in increasing genetic diversity (i.e., by the action of selection). Genetic diversity has been considered a good predictor of recombination rates; that is, levels of DNA sequence variation are normally reduced in genomic regions with low recombination rates [Begun and Aquadro, 1992; Nachman, 2001; Stevison et al., 2016]. Four different causes have been considered to explain this correlation [Begun and Aquadro, 1992; Aquadro, 1997]: (1) the mutagenic effect of the recombination process itself, (2) functional constraints, (3) adaptive evolution (such as selective sweeps – the reduction of nucleotide variation due to positive selection), and (4) background selection (i.e., loss of genetic diversity at a non-deleterious locus due to negative selection). Among all of these possibilities, the correlation between recombination and genetic divergence (scored as the ratio of rates of substitution at non-synonymous and synonymous nucleotide sites, $d_N/d_S$) is, however, more controversial, given that conflicting results have been reported in different organisms [for a review, see Smukowski and Noor, 2011]. According to the Hill-Robertson effect [Hill and Robertson, 1966], natural selection can be less effective in regions of low recombination rates, affecting, as a result, rates of adaptation. Whereas this correlation has been detected neither in Drosophila nor in mice [Begun and Aquadro, 1992], contrasting results have been obtained in great apes [Nachman, 2001; Bussell et al., 2006; Stevison et al., 2016]. We should bear in mind that both selective sweeps and background selection could cause the variation in recombination correlation by reducing the effective population size ($N_e$) in chromosomal regions experiencing low rates of recombination [Aquadro, 1997]. Therefore, the demographic history of species can have an impact on average recombination rate evolution, as recently shown in great apes [Stevison et al., 2016]. But not only that: $N_e$ varies across a genome and this can also have a relevant impact on recombination variability within genomes. In this context, chromosomal reorganizations play a major role by influencing recombination rates.

Chromosomal Rearrangements as Recombination Modifiers

Recombination is not only modulated by genetic factors, but can also be altered by mechanistic factors such as chromosomal reorganizations [Dumas and Britton-Davidian, 2002; Farré et al., 2013; Capilla et al., 2014]. In fact, the hypothesis that chromosomal reorganizations are associated with underdominant fitness due to their associated effects with meiotic abnormalities, and the creation of unbalanced gametes in heterozygotes, has long been discussed [White, 1973]. In this way, inversions and fusions can alter meiotic recombination by inducing the formation of inversion loops [Hale, 1986], delaying pairing and synapsis [Manterola et al., 2009; Torgasheva et al., 2013; Capilla et al., 2014], or by altering the epigenetic signatures for heterochromatinization in the pericentric regions of metacentric chromosomes [Capilla et al., 2014]. As a consequence, gene flow across reorganized genomic regions is reduced in the heterokaryotype (hybrid) since COs within these regions are selected against, permitting sympatric divergent evolution [Rieseberg, 2001; Faria and Navarro, 2010]. And this is due to the fact that chromosomal reorganization can capture combinations of alleles or polymorphisms that confer certain degrees of selective advantage.

Direct and indirect evidence of suppressed recombination within rearranged segments has been reported in the literature. In the case of inversions, direct evidence has been provided by cytogenetic studies in mammals [Ashley et al., 1981; Greenbaum and Reed, 1984; Ullastres et al., 2014] and Drosophila [Navarro et al., 1997]. Boro-
abilities that, in the long term, will contribute to speciation. Thus facilitating the accumulation of genetic incompatibilities, chromosomal rearrangements as barriers to gene flow, productive isolation and to determine the role of the factors involved in recombination. Additional efforts are in need to identify new genes involved in recombination variation across genomes. The one hand, genome reshuffling influences recombination landscapes. On the other hand, gene incompatibilities, reduced introgression, and higher differentiation are often associated with genomic regions with reduced recombination. Speciation genes have been described mostly in Drosophila, including, for instance, Odysseus-site homeobox (OdsH) [Ting et al., 1998], JYalpha [Masly et al., 2006], and Overdrive (Ovd) genes [Phadnis and Orr, 2009]. In mammals, the Prdm9 gene is the single speciation gene described so far for the group [Mihola et al., 2009]. However, recent studies using quantitative trait locus mapping have suggested the role of additional genomic regions (such as the proximal region of mouse chromosome X) containing genes related to mouse sterility phenotypes [Dzur-Gejdosova et al., 2012; Bhattacharryya et al., 2013; Turner and Harr, 2014; Balcova et al., 2016]. Additional efforts are in need to identify new genes involved in reproductive isolation and to determine the role of the chromosomal rearrangements as barriers to gene flow, thus facilitating the accumulation of genetic incompatibilities that, in the long term, will contribute to speciation.

Conclusions

Decades of research have resulted in a general appreciation of meiotic recombination being essential for genome evolution. As we have seen, both genetic and mechanistic factors influence the genomic makeup of recombination landscapes in mammals, reflecting several superimposed layers of regulation. These include: (1) the selective forces that affect the DNA sequence itself, which can be modulated by genome reshuffling and the evolutionary history of each taxon (population genetics) and (2) the mechanistic forces that control how the DNA is packaged into chromosomes during meiosis (i.e., DNA loops size, SC length, and chromosomal reorganizations) that, in turn, would determine CO chromosomal distribution. Only with the help of a multidisciplinary approach and by studying different phylogroups, we will be able to unveil the factors that govern recombination. This will allow researchers to fully understand the impact of recombination in genome structure and evolution.

Acknowledgements

We wish to thank Michael Schmid for the invitation to contribute this article. A.R.H.’s laboratory is supported by the Spanish ‘Ministerio de Economia y Competitividad’ (CGL-2014-54317-P and BFU2015-71786-REDT) and ‘Agencia de Gestio d’Ajuts Universitari i de Recerca’ (2015 DI 003). L.C. was the beneficiary of a FPI predoctoral fellowship (BES-2011-04772).
Mammalian Meiotic Recombination


