Molecular Evolution of the Meiotic Recombination Pathway in Mammals

Investigations

Amy L. Dapper[1,2*] and Bret A. Payseur[1]

[1] Laboratory of Genetics, University of Wisconsin, Madison, WI 53706, USA

[2] Department of Biological Sciences, Mississippi State University, Mississippi State, MS 39762, USA

Running Title: Evolution of the Recombination Pathway

Keywords: pathway evolution, evolutionary rate, adaptive evolution, divergence, crossover

* Corresponding Author : Amy L. Dapper

Address: 295 E. Lee Blvd., P.O. Box GY, Mississippi State, MS 39762

Phone: (662) 325-7575

Email: dapper@biology.msstate.edu

Abstract

Meiotic recombination, the exchange of genetic material between homologous chromosomes during meiosis, is required for successful gametogenesis in most sexually reproducing species. Recombination is also a fundamental evolutionary force, influencing the fate of new mutations and determining the genomic scale over which selection shapes genetic variation. Species recombine at different rates, but the genetic basis of this evolution is poorly understood and the molecular evolution of most recombination genes remains uncharacterized. Using a phylogenetic comparative approach, we measure rates of evolution in 32 recombination pathway genes across 16 mammalian species, spanning primates, murids, and laurasithians. By analyzing a carefully-selected panel of genes involved in key components of recombination – spanning double strand break formation, strand invasion, the crossover/non-crossover decision, and resolution – we generate a comprehensive picture of the evolution of the recombination pathway in mammals. Recombination genes exhibit marked heterogeneity in the rate of protein evolution, both across and within genes. We report signatures of rapid evolution and positive selection that could underlie species differences in recombination rate. In particular, we highlight *TEX11* and associated genes involved in the synaptonemal complex and the early stages of the crossover/non-crossover decision as candidates for the evolution of recombination rate. We also reveal that genes involved in recombination exhibit strong correlations in evolutionary rate, including among genes known to physically interact.

Introduction

The reciprocal exchange of DNA between homologous chromosomes during meiosis – recombination – is required for successful gametogenesis in most species that reproduce sexually (Hassold and Hunt 2001). The rate of recombination is a major determinant of patterns of genetic diversity in populations, influencing the fate of new mutations (Hill and Robertson 1966), the efficacy of selection (Felsenstein 1974; Charlesworth et al. 1993; Comeron et al. 1999; Gonen et al. 2017), and several features of the genomic landscape (Begun and Aquadro 1992; Charlesworth et al. 1994; Duret and Arndt 2008).

Although recombination rate is often treated as a constant, this fundamental parameter evolves over time. Genomic regions ranging in size from short sequences to entire chromosomes vary in recombination rate – both within and between species (Burt and Bell 1987; Broman et al. 1998; Jeffreys et al. 2005; Coop and Przeworski 2007; Kong et al. 2010; Dumont et al. 2011; Smukowski and Noor 2011; Comeron et al. 2012; Segura et al. 2013; Dapper and Payseur 2017; Stapley et al. 2017).

Genome-wide association studies are beginning to reveal the genetic basis of differences in recombination rate within species. Individual recombination rates have been associated with variation in specific genes in populations of *Drosophila melanogaster* (Hunter et al. 2016), humans (Kong et al. 2008; Chowdhury et al. 2009; Fledel-Alon et al. 2011; Kong et al. 2014), domesticated cattle (Sandor et al. 2012; Ma et al. 2015; Kadri et al. 2016; Shen et al. 2018), domesticated sheep (Petit et al. 2017), Soay sheep (Johnston et al. 2016), and red deer (Johnston et al. 2018). Variants in several of these genes correlate with recombination rate in multiple species, including: *RNF212* (Kong et al. 2008; Chowdhury et al. 2009; Fledel-Alon et al. 2011; Sandor et al. 2012; Johnston et al. 2016; Kadri et al. 2016; Petit et al. 2017), *RNF212B* (Johnston et al. 2016; Kadri et al. 2016; Johnston et al. 2018), *REC8* (Sandor et al. 2012; Johnston et al. 2016; Johnston et al. 2018), *HEI10/CCNB1IP1* (Kong et al. 2014; Petit et al. 2017), *MSH4* (Kong et al. 2014; Ma et al. 2015; Kadri et al. 2016; Shen et al. 2018), *CPLX1* (Kong et al. 2014; Ma et al. 2015; Johnston et al. 2016; Shen et al. 2018) and *PRDM9* (Fledel-Alon et al. 2011; Sandor et al. 2012; Kong et al. 2014; Ma et al. 2015; Shen et al. 2018).

In contrast, the genetics of recombination rate variation among species remains poorly understood. Divergence at the di-cistronic gene *mei-217/mei-218* explains much of the disparity in genetic map length between *D. melanogaster* and *D. mauritiana* (Brand et al. 2018). *mei-217/mei-218* is the only gene known to confer a recombination rate difference between species, though quantitative trait loci that contribute to shifts in rate among subspecies of house mice have been identified (Murdoch et al. 2010; Dumont and Payseur 2011; Balcova et al. 2016).

One strategy for understanding how species diverge in recombination rate is to inspect patterns of molecular evolution at genes involved in the recombination pathway. This approach incorporates knowledge of the molecular and cellular determinants of recombination and is motivated by successful examples. *mei-217/mei-218* was targeted for functional analysis based on its profile of rapid evolution between *D. melanogaster* and *D. mauritiana* (Brand et al. 2018). *PRDM9*, a protein that positions recombination hotspots in house mice and humans through histone methylation (Myers et al. 2010; Parvanov et al. 2010; Grey et al. 2011; Paigen2018; Grey et al. 2018), shows accelerated divergence across mammals (Oliver et al. 2009). The rapid evolution of *PRDM9* – which localizes to its zinc-finger DNA binding domain (Oliver et al. 2009) – appears to be driven by selective pressure to recognize new hotpot motifs as old ones are destroyed via biased gene conversion (Myers et al. 2010; Ubeda and Wilkins 2011; Lesecque et al. 2014; Latrille et al. 2017). Although these examples demonstrate the promise of signatures of molecular evolution for illuminating recombination rate differences between species, patterns of divergence have yet to be reported for most genes involved in meiotic recombination.

Mammals provide a useful system for dissecting the molecular evolution of the recombination pathway for several reasons. First, the evolution of recombination rate has been measured along the mammalian phylogeny (Dumont and Payseur 2008; Segura et al. 2013). Second, recombination rate variation has been associated with specific genes in mammalian populations (Kong et al. 2008; Chowdhury et al. 2009; Sandor et al. 2012; Kong et al. 2014; Ma et al. 2015; Johnston et al. 2016; Kadri et al. 2016; Petit et al. 2017; Johnston et al. 2018; Shen et al. 2018). Third, laboratory mice have proven to be instrumental in the identification and functional characterization of recombination genes (Vries et al. 1999; Baudat et al. 2000; Romanienko and Camerini-Otero 2000; Yang et al. 2006; Ward et al. 2007; Schramm et al. 2011; Bisig et al. 2012; Bolcun-Filas and Schimenti 2012; La Salle et al. 2012; Kumar et al. 2015; Finsterbusch et al. 2016; Stanzione et al. 2016).

Work in mice indicates that the mammalian recombination pathway is roughly divided into five major steps, each of which is regulated by a handful of genes. The first step is the formation of hundreds of double strand breaks (DSBs) throughout the genome (Bergerat et al. 1997; Keeney et al. 1997; Baudat et al. 2000; Romanienko and Camerini-Otero 2000; Baudat and Massy 2007; Finsterbusch et al. 2016; Lange et al. 2016). After formation, DSBs are identified, processed, and paired with their corresponding location on the homologous chromosome through the processes of homology search and strand invasion (Keeney 2007; Cloud et al. 2012; Brown and Bishop 2014; Finsterbusch et al. 2016; Kobayashi et al. 2016; Oh et al. 2016; Xu et al. 2017). The pairing of homologous chromosomes is then stabilized by a proteinaceous structure referred to as the synaptonemal complex (SC) (Meuwissen et al. 1992; Schmekel and Daneholt 1995; Costa et

al. 2005; Vries et al. 2005; Hamer et al. 2006; Yang et al. 2006; Schramm et al. 2011; Fraune et al. 2014; Hernández-Hernández et al. 2016). The SC also forms a substrate on which the eventual crossover events will take place (Page and Hawley 2004; Hamer et al. 2008). It is at this point that a small subset of DSBs is designated to mature into crossovers, leaving the majority of DSBs to be resolved as non-crossovers (Snowden et al. 2004; Yang et al. 2008; Reynolds et al. 2013; Finsterbusch et al. 2016; Rao et al. 2017). Finally, this designation is followed, and each DSB is repaired as a crossover or a non-crossover (Baker et al. 1996; Edelmann et al. 1996; Lipkin et al. 2002; Rogacheva et al. 2014; Xu et al. 2017).

In this article, we examine the molecular evolution of 32 key recombination genes, evenly distributed across each major step in the recombination pathway, in 16 mammalian species spanning primates, murids, and laurasiatherians. Our results point to steps of the pathway most likely to contribute to differences in recombination rate between species.

Materials and Methods

Data Acquisition & Processing

We selected a focal panel of 32 recombination genes (See Table1). The panel was constructed to: (1) cover each major step in the recombination pathway as evenly as possible, (2) contain genes that have integral functions in each step, and (3) include genes that have been associated with inter-individual differences in recombination rate within mammalian populations. Reference sequences from 16 species of mammals for each gene were downloaded from both NCBI and Ensembl (Release-89)(Wheeler et al. 2006; Zerbino et al. 2017).

Alternative splicing is widespread and presents a challenge for molecular evolution studies (Pan et al. 2008; Barbosa-Morais et al. 2012). To focus our analyses on coding sequences that are transcribed during meiosis and to validate the computational annotations for each gene in each species, we used available testes expression datasets. We downloaded raw testes expression data for each species from NCBI Gene Expression Omnibus (GEO) (Table S1)(Barrett et al. 2012). We converted the SRA files into FASTQ files using SRAtoolkit (Leinonen et al. 2010). The reads were mapped to an indexed reference genome (Tables S2 and S3)(Bowtie2; Langmead and Salzberg 2012) using TopHat (Trapnell et al. 2009). The resulting bam files were sorted using Samtools (Li et al. 2009) and visualized using IGV 2.4.10 (Thorvaldsdóttir et al. 2013). We used this approach to: (1) identify the transcript expressed in testes, (2) check the reference transcript for errors, and (3) revise the reference transcript based upon the transcript data.

We compared expression data to annotations from both Ensembl and NCBI (Wheeler et al. 2006; Zerbino et

al. 2017). When both transcripts were identical, we selected the NCBI transcript. The Ensembl transcript was used instead when: (1) the NCBI reference sequence was not available, (2) when none of the NCBI transcripts matched the expression data, or (3) when there were sequence differences between the two transcripts and the Ensembl transcript was more parsimonious (i.e. had the fewest differences when compared to the rest of sequences in the alignment). The use of testes expression data was a key quality control step and the inclusion of species in this study was primarily determined by the availability of testes expression data.

Phylogenetic Comparative Approach

For each gene, we used phylogenetic analysis by maximum likelihood (PAML 4.8) to measure the rate of evolution across the mammalian phylogeny and to search for molecular signatures indicative of positive selection (Table 2) (Yang 1997; Yang 2007). This approach requires a sequence alignment and a phylogenetic tree. For each gene, sequences were aligned using Translator X, a codon-based alignment tool, powered by MUSCLE v3.8.31 (Edgar 2004; Abascal et al. 2010). Each alignment was examined by hand and edited as necessary. We used a species tree that reflects current understanding of the phylogenetic relationships of the species included in our study (Figure 1)(Prasad et al. 2008; Perelman et al. 2011; Fan et al. 2013; Chen et al. 2017).

Due to the ambiguity in the relationship between laurasithians and the placement of tree shrews, we also inferred gene trees using MrBayes (Ronquist et al. 2012; Fan et al. 2013; Chen et al. 2017). To infer each gene tree, we selected the General Time Reversible (GTR) substitution model with gamma-distributed rate variation across sites and the Markov chain Monte Carlo (mcmc) sampling was run until the standard deviation frequency was less than 0.01 (Ronquist et al. 2012). We used this approach to account for effects of incomplete lineage sorting (ILS) (Pamilo and Nei 1988; Rosenberg 2002; Scornavacca and Galtier 2017). Using gene trees and using the consensus species tree produced highly similar results (Table S5).

For 19 genes, transcripts from all 16 species were used. For 11 genes in which the chimpanzee and bonobo sequences were identical, we excluded the bonobo sequence. For one gene in which the chimpanzee, bonobo and human sequences were all identical, we excluded the chimpanzee and bonobo sequences. In only two cases, a suitable reference sequence could not be identified for a given species (*RNF212B*: Rat; *TEX11*: Tree Shrew).

We estimated rates of synonymous and non-synonymous substitutions per site using the CODEML program in PAML4.8 (Yang 2007). This program considers multiple substitutions per site, different rates of transitions and transversions, and effects of codon usage (Yang 2007). Rates of substitution were computed for 6 different

models of molecular evolution (Table 2). The fit of each model was compared using a likelihood ratio test. Reported substitution rates assume the best-fit model for each gene.

Identifying Signatures of Selection

To test for positive selection, we compared the fit of models including a class of sites with $\omega > 1$ to the fit of models in which all classes of sites have $\omega \leq 1$. Specifically, we report three comparisons: M1 vs. M2, M7 vs. M8, and M8 vs. M8a (Table 2). The first comparison, M1 vs. M2, compares a model with two classes of sites ($\omega < 1$, $\omega = 1$) to a model with a third class of sites where $\omega > 1$, indicative of positive selection (Yang 2007). More complex models (M7 & M8) were developed to consider variation in $\omega < 1$ among sites within genes by including 10 site classes drawn from a beta distribution ranging from 0 to 1 (Yang 2007). In this case, Model 8 includes one additional class of sites in which $\omega > 1$ (for a total of 11 site classes), allowing for the identification of signatures of positive selection (Yang 2007). In cases in which a large fraction of sites within a gene are evolving neutrally ($\omega = 1$), Model 8 will fit significantly better due to a very poor fit of Model 7 rather than a signature of positive selection. To avoid incorrectly identifying signatures of positive selection in this case, we also compared Model 8 to Model 8a, which contains a larger fraction of neutrally evolving sites than Model 7 (Swanson et al. 2003). We report the number of codons in each gene estimated to have $\omega > 1$ (Bayes-Empiricial-Bayes (BEB); $P > 0.95$).

Multinucleotide Mutations

Multi-nucleotide mutations (MNMs) occur when two mutations happen simultaneously in close proximity (Schrider et al. 2011; Besenbacher et al. 2016). MNMs violate the PAML assumption that the probability of two simultaneous mutations in the same codon is 0 (Yang 2007; Venkat et al. 2018). Recent work has shown that MNMs can lead to the false inference of positive selection when using branch-site tests in PAML (Venkat et al. 2018). Although we did not use branch-site tests, it is possible that MNMs contributed to some of the signatures of positive selection we observed. Although we could not directly identify MNMs in our dataset, we conducted an additional analysis to gauge the potential effects of MNMs on our results. We used PAML to reconstruct the ancestral sequence at each node in the phylogeny (Yang 2007). For the reconstruction, Model 8 was chosen because we specifically re-analyzed genes that showed evidence for positive selection when comparing Model 7 with Model 8. From the ancestrally reconstructed sequences, we identified any codons in which PAML inferred more than 1 substitution on a single branch (codons with multiple differences; CMDs). All identified CMDs were removed from the sequences in which they occurred. For example, if a CMD was

identified in an external branch, that codon was replaced with '—' only in the sequence of that species. If a CMD was inferred on an internal branch, the codon was replaced with '—' in all species descended from that internal branch. For each gene that showed evidence of positive selection using the unedited sequences, we also conducted PAML analyses using sequences from which all CMDs were removed.

Polymorphism & Divergence in the Primate Lineage

To further examine evidence for selection on recombination genes, we compared divergence between humans and macaque to polymorphism within humans in the recombination genes. Human polymorphism data was downloaded from ExAC database (Lek et al. 2016). The ExAC database spans 60,706 unrelated individuals sequenced as part of both disease-specific and population genetic studies (Lek et al. 2016). To avoid biases introduced by population structure, we restricted our analyses to the population with the largest representation in the database: European, non-Finnish, individuals (N = 33,370)(Lek et al. 2016). Polymorphism data for the correct transcript of *RNF212* (based upon expression data) was not available in the ExAC database; this gene was not included in this analysis.

We compared counts of non-synonymous and synonymous polymorphisms to counts of non-synonymous and synonymous substitutions using the McDonald-Kreitman test (McDonald and Kreitman 1991). The neutral expectation is that the ratio of non-synonymous to synonymous substitutions is equal to the ratio of non-synonymous to synonymous polymorphisms (McDonald and Kreitman 1991). Significant deviations provide evidence of natural selection. The neutrality index (NI) measures the direction and degree of departures from the neutral expectation (Charlesworth 1994). An NI < 1 indicates positive selection, and the fraction of adaptive amino acid substitutions can be estimated as 1 - NI (Charlesworth 1994; Fay et al. 2001; Smith and Eyre-Walker 2002). We also measure the direction of selection (DoS) for each gene, an additional statistic that estimates the direction and degree of departures from the neutral expectation and has been shown to be less biased than NI under certain conditions (Stoletzki and Eyre-Walker 2010). A positive DoS is consistent with positive selection; a negative DoS indicated purifying selection (Stoletzki and Eyre-Walker 2010). Additionally, we estimated pairwise divergence (ω) between human and macaque using the *yn00* package in PAML (Table 5 and Table S4) (Yang 2007).

Identifying Evolutionary Patterns

To identify evolutionary patterns among recombination genes, we compared the rate of evolution and the proportion of genes experiencing positive selection among groups of interest. We asked: (1) Do genes that

function in different steps of the pathway exhibit different rates of evolution? (2) Do genes that function post-synapsis evolve more rapidly than genes that function pre-synapsis? and (3) Do genes associated with variation in recombination rate (within species) diverge more rapidly between species? All statistical analyses were performed in R (R Core Team 2015).

To determine whether recombination genes co-evolve, we computed the evolutionary rate covariation (ERC) metric: the correlation coefficient between branch-specific rates among pairs of proteins (Clark et al. 2012). ERC is frequently elevated among interacting proteins (Pazos and Valencia 2001; Hakes et al. 2007; Clark et al. 2009) and is assumed to result from: (1) concordance in fluctuating evolutionary pressures, (2) parallel evolution of expression level, and/or (3) compensatory changes between co-evolving genes (Clark et al. 2012; Clark et al. 2013; Priedigkeit et al. 2015). We used a publicly available ERC dataset (https://csb.pitt.edu/erc_analysis/index.php) to compare the median ERC-value among a subset of the focal recombination genes ($N = 25$) to other genes in the genome, as described in Priedigkeit et al. (2015).

To control for an observed elevation in ERC among recombination genes and test for relationships between specific groups, we also conducted an ERC analysis that was restricted to the focal set of 32 recombination genes. Branch lengths were calculated using the *aaML* package in PAML (Yang 2007) and pairwise ERC values were calculated following the methods of Clark et al. (2012). Using this approach, we specifically compared the ERC values among three of the most rapidly evolving recombination genes (*TEX11*, *SHOC1*, and *SYCP2*) to the other recombination genes.

Data Availability

Sources and accession numbers for all publicly available sequence data used in this study can be found in Supplemental Tables S1-S3. All data, sequence alignments, and code necessary to replicate this study can be found in the following publicly available GitHub Repository (https://github.com/adapper/Evo_RecGenes_MS).

Results

Recombination genes evolve at different rates in mammals

We observed substantial heterogeneity in the rate of evolution of recombination genes, spanning a range of 0.0268 – 0.8483 (mean $\omega = 0.3275$, SD = 0.1971, median = 0.3095) (Figure 2A, Figure 3, Table 3). Four genes exhibit particularly rapid evolution compared to other recombination genes, with evolutionary rates greater than 1 SD above the mean (*IHO1*, *SHOC1*, *SYCP2*, *TEX11*). At the other end of the spectrum, five genes

have evolutionary rates more than 1 SD below the mean and are highly conserved across the mammalian phylogeny (*BRCC3*, *DMC1*, *HEI10*, *RAD50*, *RAD51*). In comparisons between human and macaque, six genes have evolutionary rates more than 1 SD above the mean (*CNTD1*, *IHO1*, *MEI4*, *RAD21L*, *SHOC1*, *TEX11*) and six genes have evolutionary rates more than 1 SD below the mean (*DMC1*, *HORMAD1*, *MLH1*, *MRE11*, *RAD50*, *RAD51*). The genes that show the most rapid and most conserved rates of divergence between humans and macaques are mostly the same genes that show extreme evolutionary rates across the mammalian phylogeny. Notable exceptions include *MEI4* ($\omega_{\text{mammals}} = 0.4332$, $\omega_{\text{human-macaque}} = 0.7252$), *CNTD1* ($\omega_{\text{mammals}} = 0.2496$, $\omega_{\text{human-macaque}} = 0.6803$), *HEI10* ($\omega_{\text{mammals}} = 0.1226$, $\omega_{\text{human-macaque}} = 0.3235$), and *HORMAD1* ($\omega_{\text{mammals}} = 0.3036$, $\omega_{\text{human-macaque}} = 0.0901$. In general, there is very high concordance between evolutionary rate across mammals and pairwise divergence between humans and macaques (mean $\omega = 0.3301$, SD = 0.2370, median = 0.30925)(Spearman's $\rho = 0.833774$, $p = 3.11\text{e-}9$)(Figure 2B, Table 4). It should be noted that these two measures are not independent because divergence between human and macaque sequences was incorporated in the phylogenetic analysis across mammals.

Recombination genes evolve faster than other genes in primates

Gradnigo et al. (2016) measured the rate of divergence between human and macaque for 3,606 genes throughout the genome. We used this dataset to ask whether the rate of evolution of recombination genes as a group is different than expected from the genome-wide distribution. Mean rates for sets of 32 ω values randomly sampled from the 3,606-gene list rarely exceeded the mean rate for recombination genes ($p = 0.0075$, 10,000 random draws) (Figure 4), suggesting that recombination genes evolve faster on average, at least between human and macaque.

Recombination genes display signatures of positive selection across mammals

Comparing polymorphism within humans to divergence between human and macaque revealed that 17 out of 31 genes depart from neutral predictions in the form of significant McDonald-Kreitman tests (Fisher's Exact Test, $p < 0.05$; Table 5). These seventeen genes harbor an excess of non-synonymous polymorphisms (Table 5). This pattern suggests the presence of weakly deleterious mutations at recombination genes in human populations. Contrary to predictions under this model, however, we detected no significant differences in allele frequency between non-synonymous and synonymous polymorphism (Wilcoxon rank sum test; $p > 0.05$**). None of the recombination genes we surveyed displays a significant excess of non-synonymous substitutions, the expected signature of positive selection. Only one gene (*TEX11*) has a higher ratio of

non-synonymous to synonymous substitutions than non-synonymous to synonymous polymorphisms (NI = 0.7879; DoS = 0.0534)(Table 5).

In contrast to conventional McDonald-Kreitman tests, phylogenetic comparative methods enable the identification of signatures of selection acting on a subset of sites within a gene over long evolutionary timescales. We identified signatures of positive selection in 11 of 32 (34.3%) recombination genes using site models in *CODEML*: *IHO1*, *MSH4*, *MRE11*, *NBS1*, *RAD21L*, *REC8*, *RNF212*, *SHOC1*, *SYCP1*, *SYCP2*, and *TEX11* (Table 2). For each of these genes, models that include a fraction of sites where the rate of non-synonymous substitutions is estimated to be greater than the rate of synonymous substitutions ($\omega > 1$, Model 8) fit better than models that did not include such a class of sites (Model 7, 8a). To mitigate the potential for multi-nucleotide mutations to produce false signatures of positive selection, we re-analyzed this subset of genes after removing any codons inferred to have accumulated multiple changes on a single branch (CMDs). After removing all CMDs, 1 gene (*TEX11*) retained a significant signature of positive selection (Table 5).

Recombination genes associated with inter-individual differences do not diverge more rapidly between species

Recombination genes previously associated with inter-individual differences in recombination rate within species do not evolve significantly faster between species of mammals (average ω = 0.3943 vs. average ω = 0.2925, respectively; p = 0.2381, Mann-Whitney U Test), though the difference in evolutionary rates between these two classes of genes is greater when considering only divergence between humans and macaques (average ω = 0.4181 vs. average ω = 0.2839, respectively; p = 0.08816, Mann-Whitney U Test). Likewise, the proportion of recombination genes that exhibit signatures of positive selection is not significantly higher among genes that have been associated with inter-individual differences (5/11 vs. 6/21; p = 0.4424, Fisher's Exact Test).

Recombination gene evolution does not depend strongly on position in the pathway

Comparisons among groups of genes assigned to six major steps in the recombination pathway yielded no significant differences in evolutionary rate (mammals: p = 0.1422, Kruskal-Wallis Test; human vs. macaque: p = 0.2682, Kruskal-Wallis Test)(Figure 6). Similarly, genes acting before and after synapsis show similar evolutionary rates across mammals (average ω_{before} = 0.2723 vs. ω_{after} = 0.3762, p = 0.1425, Mann-Whitney U Test). Post-synapsis genes show modest evidence of evolving faster than pre-synapsis genes in comparisons between human and macaque (average ω_{before} = 0.2514 vs. ω_{after} = 0.3994, p = 0.05827, Mann-Whitney U

Test).

Evolutionary rates are correlated among recombination genes

We used a publicly available database (https://csb.pitt.edu/erc_analysis/index.php) to measure correlations in evolutionary rate among pairs of recombination genes across mammals. Recombination genes show levels of evolutionary rate covariation (mean ERC = 0.134) that are significantly higher than the genome-wide distribution of gene pairs (permutation $p = 0.000358$).

Motivated by the findings that *TEX11*, *SYCP2*, and *SHOC1* are three of the most rapidly evolving recombination genes among mammals (Table 3) and that *TEX11* has direct protein-to-protein interactions with both *SHOC1* and *SYCP2* (Yang et al. 2008; Guiraldelli et al. 2018), we focused on rate correlations between these genes. *TEX11*, *SYCP2*, and *SHOC1* show significantly higher rate correlations (mean ERC = 0.42369) than randomly sampled subsets of recombination genes (permutation $p = 0.025$).

Discussion

Species of mammals recombine at different rates (Burt and Bell 1987 ; Dumont and Payseur 2008; Smukowski and Noor 2011; Segura et al. 2013; Stapley et al. 2017), but the genetic changes responsible for this evolution remain unknown. Patterns of molecular divergence we discovered point to genes and steps in the pathway that are good candidates for the evolution of recombination rate.

Genes involved in meiotic recombination vary substantially in evolutionary rate. Genes that are highly conserved across the mammalian phylogeny (*BRCC3*, *DMC1*, *HEI10*, *RAD50*, and *RAD51*) mostly function in the detection and processing of DSB breaks (except *HEI10*; Dong et al. 2003; Hopfner 2005; Keeney 2007; Ward et al. 2007; Cloud et al. 2012; Qiao et al. 2014; Kobayashi et al. 2016), suggesting that these processes are not primary drivers of recombination rate evolution in mammals.

Four genes exhibit especially rapid evolution across the mammalian phylogeny (compared to other recombination genes): *IHO1*, *SHOC1*, *SYCP2*, and *TEX11*. Three of these genes are known to interact. *TEX11* binds to the synaptonemal complex, including *SYCP2*, and recruits proteins, including *SHOC1*, that regulate the first step of the crossover vs. non-crossover decision (Yang et al. 2008; Guiraldelli et al. 2018). *IHO1* recruits and activates *SPO11*, a topoisomerase-like protein that generates DSBs (Stanzione et al. 2016). Eleven of the 32 recombination genes we examined display signatures of positive selection across the mammalian phylogeny. These genes are predominantly found in two steps of the pathway - formation of the synaptonemal complex

(*REC8*, *RAD21l*, *SYCP1*, and *SYCP2*; Parisi et al. 1999; Vries et al. 2005; Yang et al. 2006; Lee and Hirano 2011) and regulation of the first steps of the crossover vs. non-crossover decision (*TEX11*, *SHOC1*, *RNF212*, and *MSH4*; Snowden et al. 2004; Yang et al. 2008; Qiao et al. 2014; Guiraldelli et al. 2018) - raising the possibility that adaptive evolution of these processes caused divergence in recombination rate among species. Five of these genes have been associated with inter-individual variation in recombination rate within species: *RAD21l* (Kong et al. 2014); *REC8* (Sandor et al. 2012; Johnston et al. 2016; Johnston et al. 2018); *MSH4* (Kong et al. 2014; Ma et al. 2015; Kadri et al. 2016; Shen et al. 2018); *RNF212* (Kong et al. 2008; Chowdhury et al. 2009; Fledel-Alon et al. 2011; Sandor et al. 2012; Johnston et al. 2016; Kadri et al. 2016; Petit et al. 2017); and *TEX11* (Murdoch et al. 2010).

Our results highlight an evolutionary contrast between mammals and Drosophila. *MCMDC2*, the mammalian homolog of the *mei-217*/*mei-218* gene that evolves rapidly and adaptively in Drosophila (Brand et al. 2018), exhibits below average rates of evolution and no evidence of positive selection in mammals. These two homologs occupy different positions in the recombination pathway. In Drosophila, *mei-218* has evolved to replace the function of the missing *MSH4* and *MSH5* (Kohl et al. 2012; Finsterbusch et al. 2016). This shift in both evolutionary rate and pathway function suggests that functional homology is a better predictor of evolutionary rate than sequence homology for this recombination gene.

Deeper consideration of the recombination gene that shows the strongest evidence for positive selection, *TEX11*, provides additional clues about the genesis of recombination differences among mammals. Fourteen amino acid residues in *TEX11* exhibit patterns consistent with adaptive evolution (BEB, $P > 0.95$). In contrast to *MSH4* or *PRDM9* – where targets of selection localize to certain protein domains (Oliver et al. 2009; Thomas et al. 2009; Grey et al. 2011) – the *TEX11* residues of interest are distributed across the length of the gene. This pattern matches aspects of *TEX11* protein function. The gene encompasses three large, ubiquitous protein interaction (TRP) domains (Guiraldelli et al. (2018)). Most of the residues with signatures of selection localize to two of the large TRP domains, one of which is known to bind to *SHOC1* (Guiraldelli et al. 2018).

If mammals have experienced directional selection to increase crossover number (Segura et al. 2013), a key role for *TEX11* suggests that divergence in this trait reflects genetic changes acting early during the crossover/non-crossover decision. Alternatively, the positioning of *TEX11* at this stage in the pathway could suggest that selection favored a consistent recombination rate across the mammalian phylogeny. In this case, the signature of positive selection in *TEX11* could be driven by its role in maintaining crossover homeostasis despite the accumulation of changes in other genes in the pathway. The correlation in evolutionary rates between *TEX11*, *SYCP2*, and *SHOC1* can support either scenario (Clark et al. 2012; Clark et al.

2013). Either all three genes experience concordant selection pressures due to their similar functions in the recombination pathway or the correlation in evolutionary rate is driven by compensatory changes in *TEX11*. Notably, sequence variation in *TEX11* has been associated with substantial difference in recombination rate in mice and in humans (Murdoch et al. 2010; Yang et al. 2015).

Genes in the recombination pathway reveal additional evolutionary patterns when considered as a group. Recombination genes tend to evolve faster than other genes, at least based on comparisons between human and macaque. Several factors could generate this pattern. First, the central role of recombination genes in reproduction could accelerate their divergence. The rapid evolution of reproductive genes is usually attributed to post-copulatory sexual selection or relaxed selection (from sex-specific expression and low female re-mating rates) (Swanson and Vacquier 2002; Dapper and Wade 2016). However, recombination occurs prior to copulation in mammals, and recombination genes are typically expressed in both sexes, two observations that argue against these explanations for elevated divergence. Second, the restriction of expression of some recombination genes to meiotic cells could reduce the pleiotropic consequences of amino acid substitutions (Duret and Mouchiroud 1999; Liao et al. 2006). A third possibility is that recombination itself is frequently subject to positive selection, driving divergence at the underlying genes (Dapper and Payseur 2017; Ritz et al. 2017).

Recombination genes previously associated with intra-specific variation in the genome-wide recombination rate evolve at similar rates to recombination genes lacking such an association. Genes responsible for species differences in recombination rate could be subject to strong directional selection within populations, reducing their contributions to intra-specific variation. Alternatively, genes that confer within-species rate variation could be targets of diversifying or antagonistic selection, limiting their divergence between species. For example, variants of *RNF212*, a gene associated with intra-specific variation in recombination rate in several mammalian species, have contrasting effects in women and men (Kong et al. 2008).

The structure of genetic pathways is expected to influence evolutionary trajectories (Rausher et al. 1999; Lu and Rausher 2003). Matching this prediction, recombination genes show relatively high rate correlations compared to other sets of genes. Nevertheless, our results suggest that the selection pressures targeting a gene are not easily deduced from its position in the recombination pathway. Perhaps rate variation among domains within proteins masks a clearer effect of pathway position. For example, the signal of adaptive evolution in *PRDM9* is restricted to the zinc finger residues, with much of the gene sequence being conserved between species (Oliver et al. 2009; Thomas et al. 2009). Rate heterogeneity between genes within steps of the recombination pathway motivates a more thorough investigation of functional domains in genes of interest.

Within-gene variation in evolutionary rate might explain another apparent discrepancy in our results. Despite evidence for positive selection across the mammalian phylogeny at many genes, comparisons of polymorphism and divergence yielded no significant signatures of adaptive evolution between human and macaque. Instead, many recombination genes display an excess of non-synonymous polymorphisms, consistent with an accumulation of weakly deleterious mutations within humans. However, this approach searches for patterns of selection at the level of the entire gene, whereas positive selection can target certain domains. For example, *MSH4* exhibits evidence for adaptive evolution (though with a lower than average evolutionary rate) along the mammalian phylogeny and shows an excess of non-synonymous polymorphisms within humans. These two seemingly disparate results are unified by the observation that all 6 codons in *MSH4* with significant signatures of positive selection (BEB, $P > 0.95$) are highly localized in the first 100 bp of the gene, in a putative DNA binding domain (Rakshambikai et al. 2013; Piovesan et al. 2017).

One cost of the increased sensitivity of PAML is an inflation of the false-positive rate in the presence of multi-nucleotide substitutions (Venkat et al. 2018). It was not possible to directly identify MNMs in our dataset, so we chose the highly conservative approach of removing all codons inferred to have accumulated multiple mutations on a single branch in the phylogeny. Codons removed using this approach could be MNMs, but they also likely include codons that either have accumulated sequential mutations along the long branches in the mammalian phylogeny or are neither MNMs nor CMDs, due to uncertainty in the inference of ancestral sequences. Despite the conservative nature of this approach, we still found a signature of positive selection in *TEX11*, even when all putative CMDs were removed. Nevertheless, the conservative nature of the filter makes it difficult to draw conclusions about the robustness of signals of selection in the other recombination genes.

Another caveat concerns the interpretation of our findings. Although we would prioritize rapidly evolving genes with evidence of adaptive evolution as candidates, evolution of the recombination rate between species could be caused by only a few amino acid substitutions (especially along particular mammalian lineages) or by regulatory changes located outside protein-coding regions. We hope our results will motivate genetic dissection of between-species differences in recombination rate through functional evaluation of the candidate genes we identified, especially *TEX11* and associated gene involved in the synaptonemal complex and early stages of the crossover/non-crossover decision.

Acknowledgements

We thank Nathan Clark for assistance with evolutionary covariation rate analyses and Francesca Cole for advice on selection of recombination genes. A.L.D. was supported by NHGRI Training Grant to the Genomic

Sciences Training Program 5T32HG002760. B.A.P. was supported by NIH grant NIH grant R01 GM120051.

Table 1 : List of 32 genes surveyed, organized by step in the recombination pathway. Genes in bold have been associated with inter-individual differences in recombination rate in at least one species of mammals.

Pathway Step	**Genes**
DSB Formation	*HORMAD1, IHO1, MEI4, SPO11,* ***REC114***
DSB Processing	*BRCC3, HORMAD2, MRE11, NBS1, RAD50*
Strand Invasion	*DMC1, MEIOB, MCMDC2, SPATA22, RAD51*
Homologous Pairing	*SYCP1, SYCP2,* ***RAD21L****,* ***REC8****, TEX12*
CO vs. NCO Decision	***MSH4****,* ***MSH5****,* ***RNF212****,* ***RNF212B****,* ***TEX11****, SHOC1*
Resolution	*CNTD1,* ***HEI10****,* ***MER3****, MLH1,* ***MLH3****, MUS81*

Table 2: Six PAML site models used to measure evolutionary rate and test for positive selection. Models varied in the number of ω classes, the range of ω for each of these classes, and whether a class of sites subject to positive selection was included.

Model	**# Site Classes**	**ω Range**	**Pos. Selection?**
0	1	<1	No
1	2	<1, =1	No
2	3	<1, =1, >1	Yes
7	10	0-1	No
8	11	0-1, >1	Yes
8a	6	0-1, =1	No

Figure 1: Species tree assumed in analyses of molecular evolution.

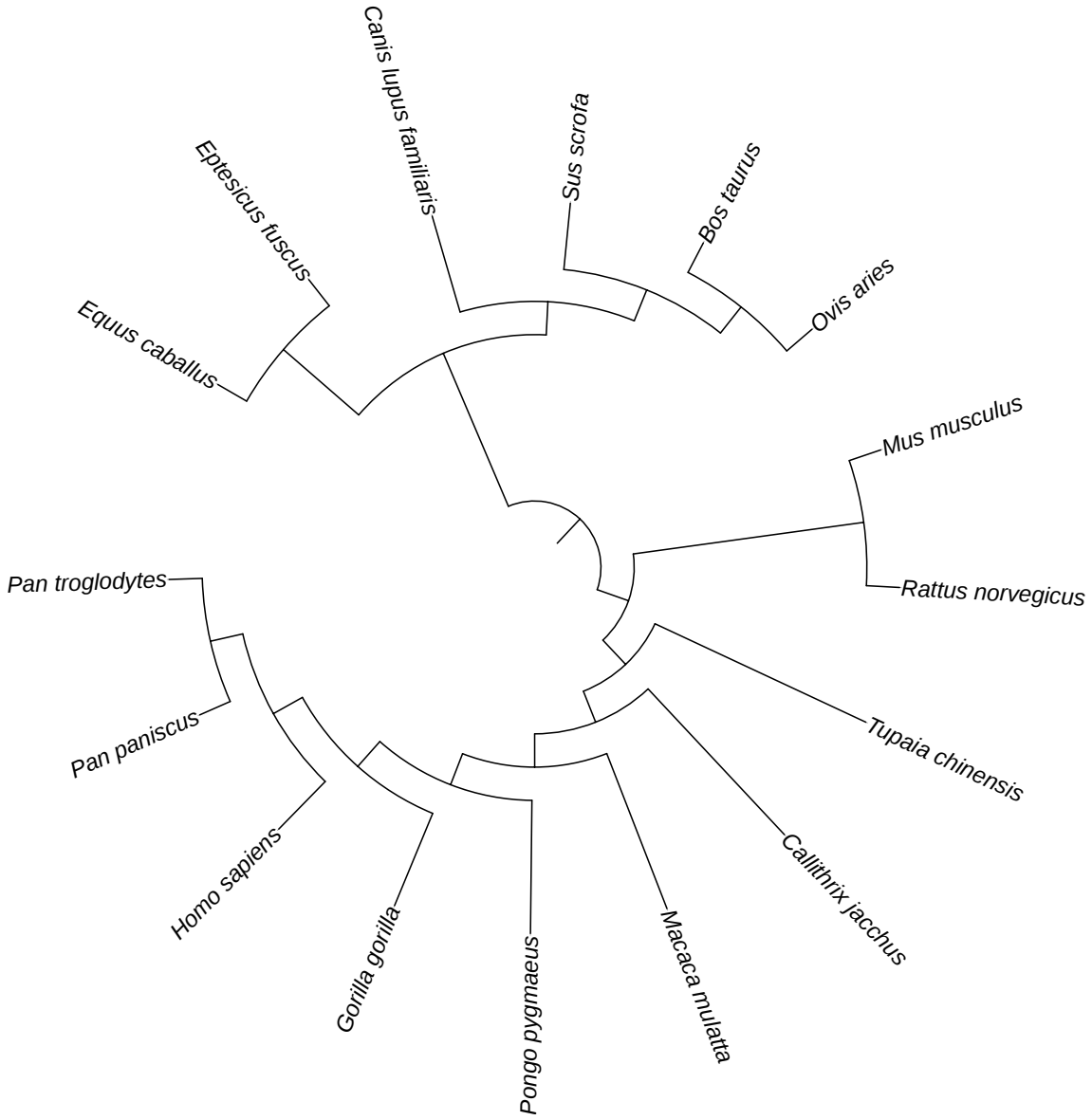
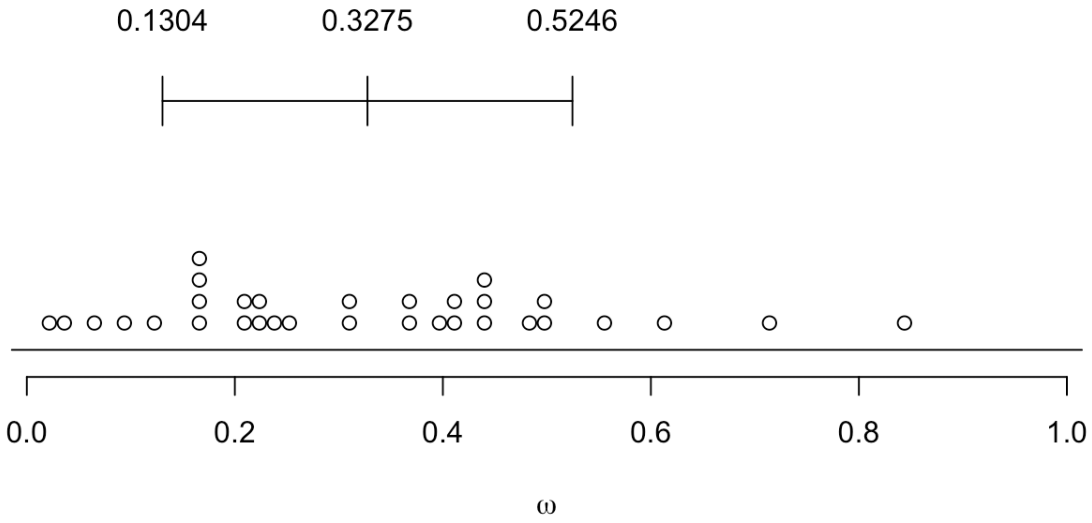


Figure 2:Distribution of ω for 32 recombination genes. Bar shows the mean +/- 1 standard deviation. (A) Divergence estimated across the mammalian phylogeny. (B) Pairwise divergence between human and macaque.

(A)



(B)

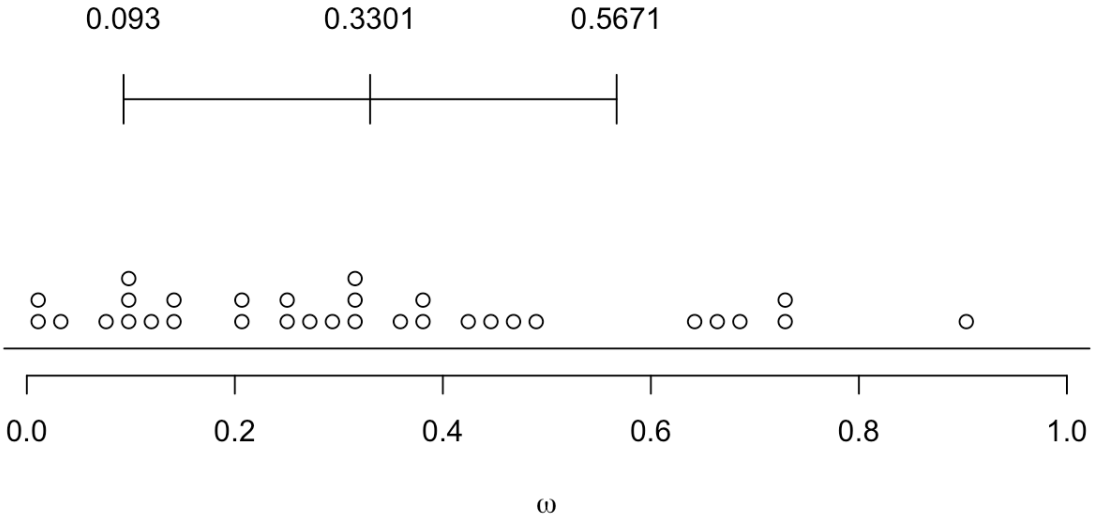


Figure 3: Evolutionary rate of key recombination genes in the context of the recombination pathway. The color of each gene represents its evolutionary rate relative to the average for recombination genes ($\omega = 0.3275$): more rapidly evolving genes are depicted in darker shades of red and more conserved genes are depicted in darker shades of blue. Genes that exhibit a signature of positive selection are in bold.

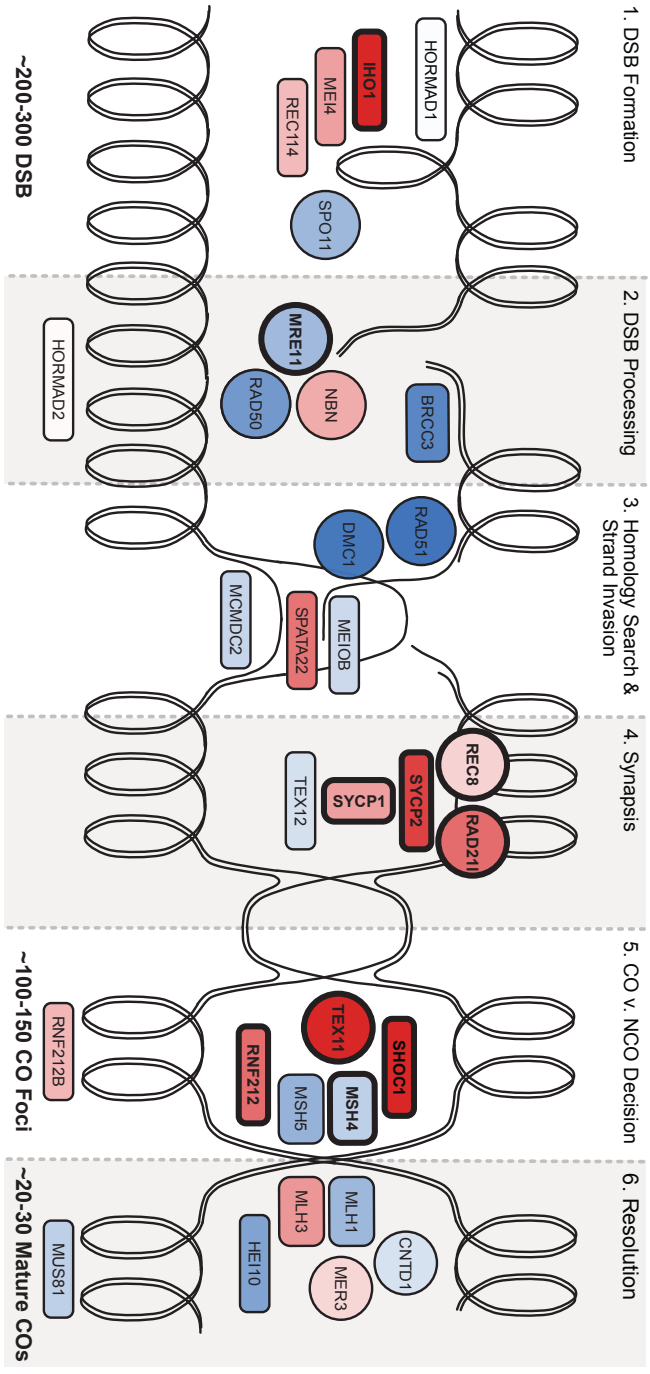


Figure 4: Distribution of average divergence (ω) between human and macaque of 10,000 gene sets randomly drawn from the entire genome. Average ω among these random draws was observed to be equal to or greater than that observed among recombination genes less than 1% of the time ($p = 0.0075$).

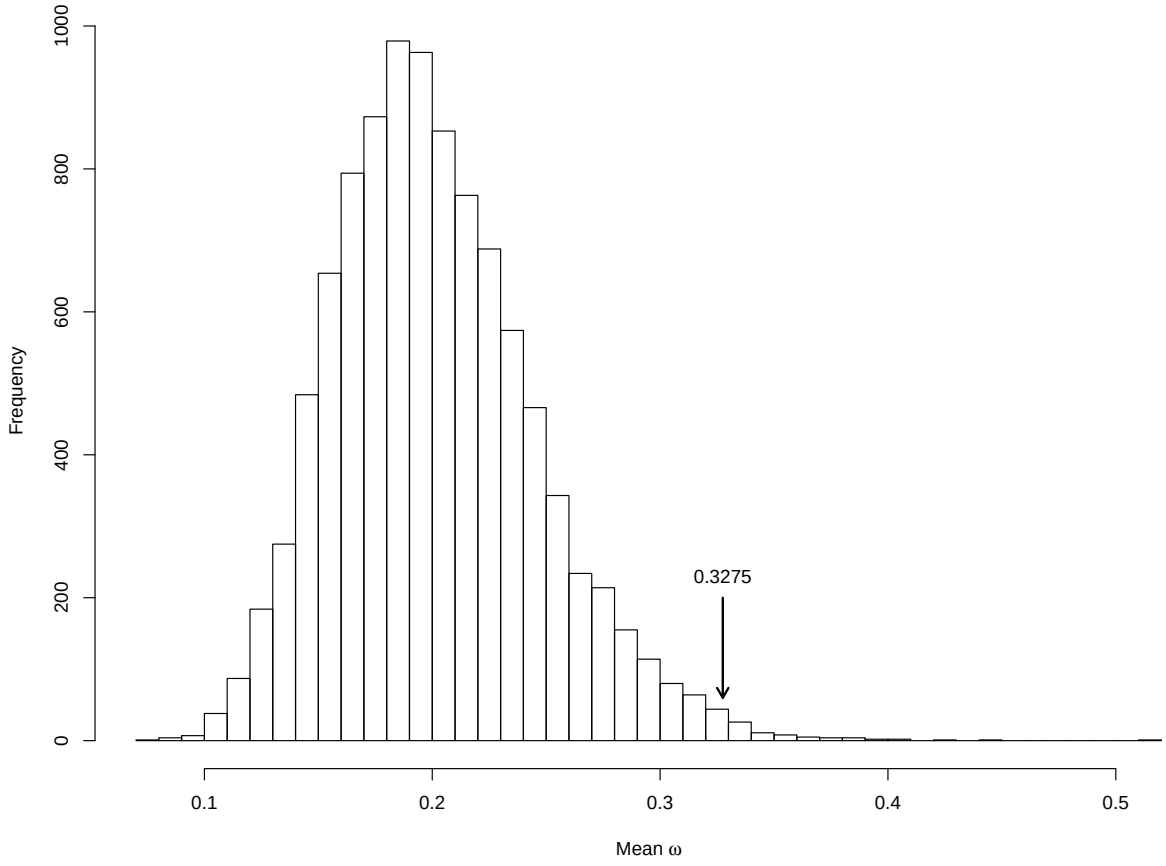


Figure 5: High concordance in evolutionary rate computed across mammals and between humans and macaque. The best-fit regression line is shown in red and the 1:1 line is shown as a dashed line.

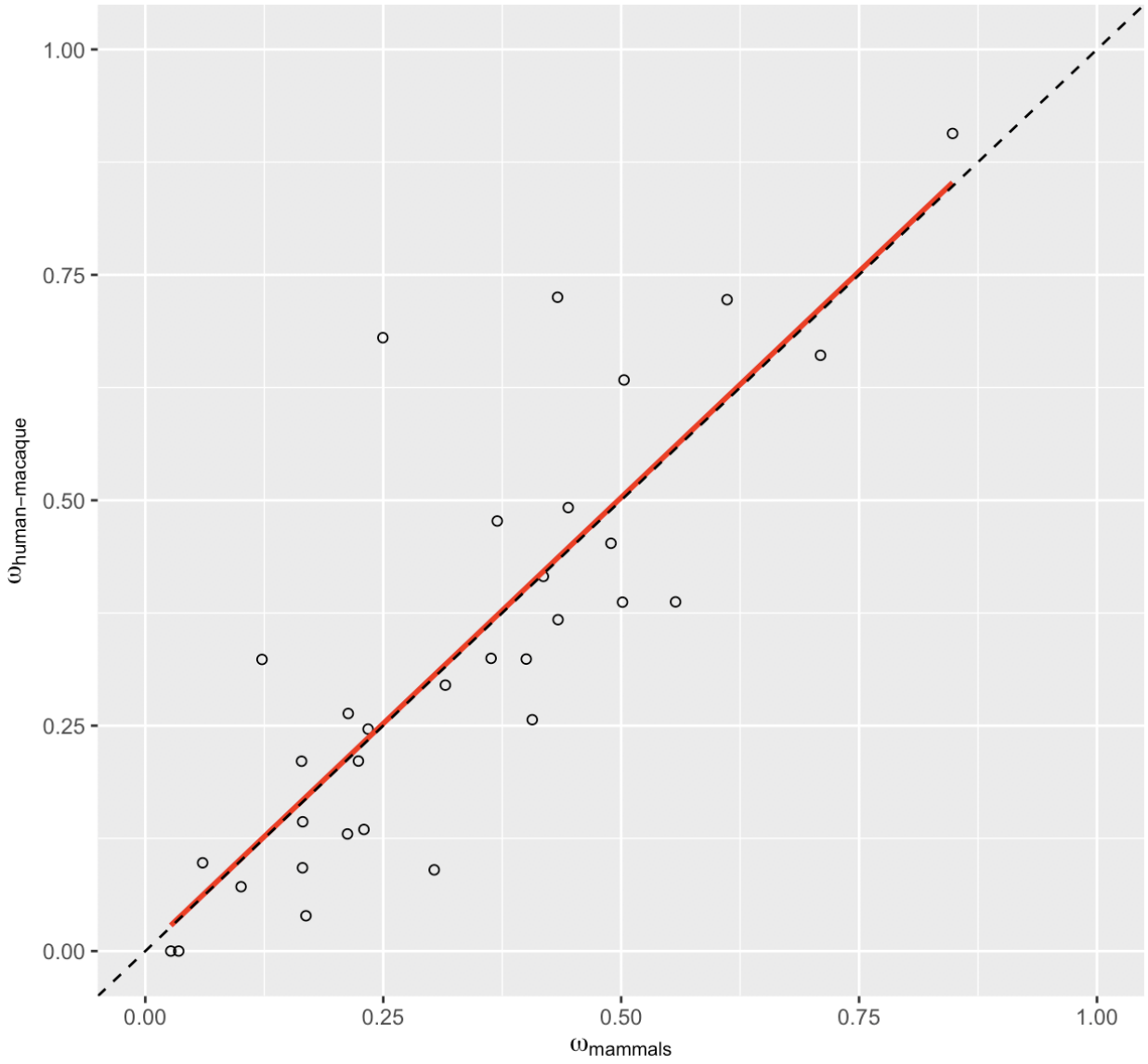
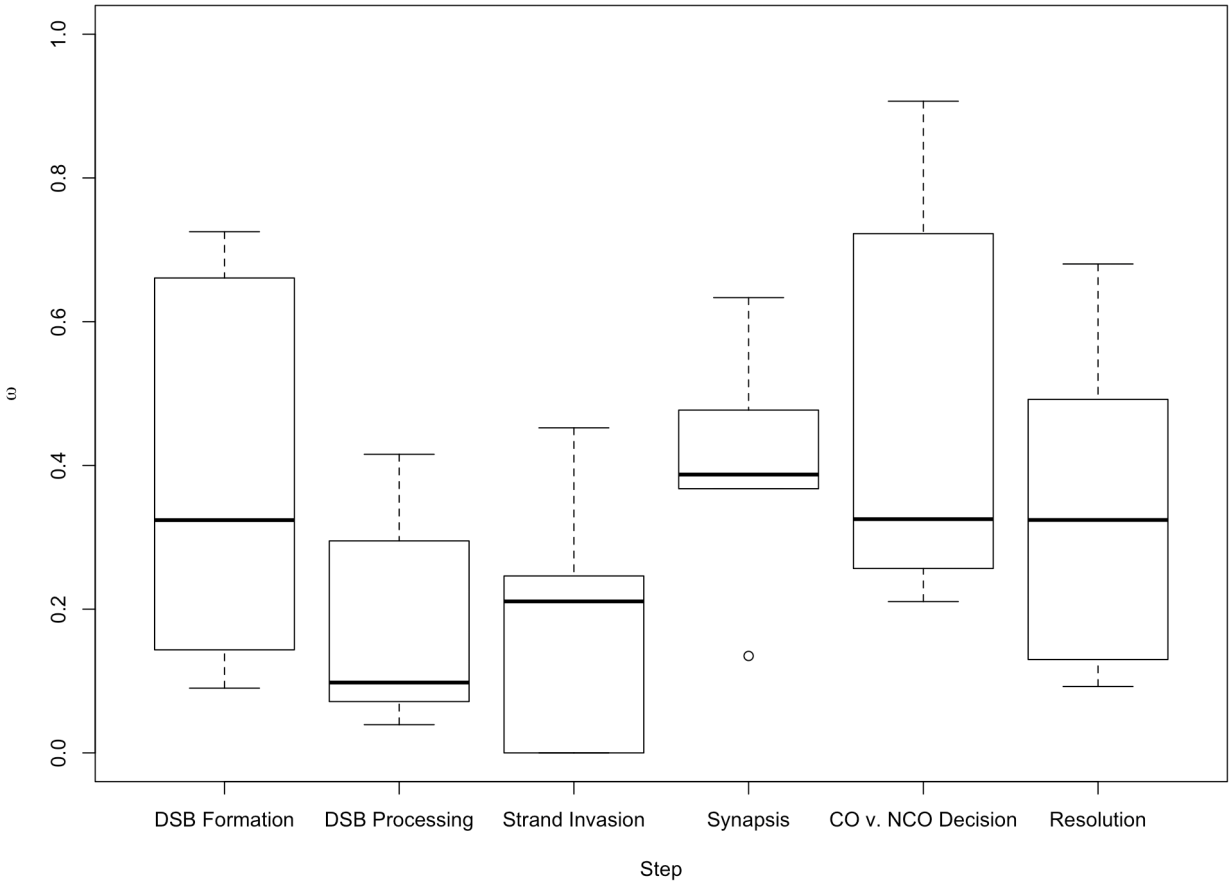


Figure 6: Boxplot of ω by step in recombination pathway.



449 **Table 3**: Evolutionary rates and tests for positive selection across mammals at 32 recombination genes.

450 Genes are organized by step in the pathway, as labeled in Figure 3.

Gene	***bp***	***N***	ω	***M***	***M1-M2***	***p-value***	***M7-M8***	***p-value***	***M8a-M8***	***p-value***	***BEB***
A)											
HORMAD1	1212	16	0.3036	7	0	*1.000*	1.795	*0.4076*	—	—	0
MEI4	1170	16	0.4332	7	0	*1.000*	0.005	*0.9976*	—	—	0
REC114	870	15	0.4003	7	0	*1.000*	5.384	*0.0677*	—	—	0
IHO1	1824	16	0.7095	**8**	13.061	***0.0015***	17.571	***0.0002***	14.527	***0.0001***	**1**
SPO11	1188	15	0.1654	7	0	*1.000*	4.648	*0.0980*	—	—	0
B)											
HORMAD2	981	15	0.3153	7	0	*1.000*	3.650	*0.1612*	—	—	0
MRE11	2136	16	0.1688	**8**	0.363	*0.8342*	11.931	***0.0026***	4.706	***0.0301***	0
NBS1	2289	15	0.4183	**8**	0	*1.000*	12.763	***0.0017***	4.087	***0.0432***	0
RAD50	3936	16	0.1006	7	0	*1.000*	0.301	*0.8605*	—	—-	0
BRCC3	954	15	0.0602	7	0	*1.000*	0.250	*0.8826*	—	—	0
C)											
DMC1	1020	15	0.0351	1	0.488	*0.7835*	5.000	*0.0821*	—	—	**1**
RAD51	1017	16	0.0268	7	0	*1.000*	0	*1.000*	—	—	0
SPATA22	1101	16	0.4893	7	0	*1.000*	0.429	*0.8070*	—	—	0
MEIOB	1425	16	0.2341	7	0	*1.000*	0.665	*0.7172*	—	—	0
MCMDC2	2052	16	0.2239	7	0	*1.000*	0.628	*0.7307*	—	—	0
D)											
REC8	1833	16	0.3698	**8**	0	*1.000*	14.690	***0.0006***	5.927	***0.0149***	0
RAD21L	1686	15	0.503	**8**	12.124	***0.0023***	32.050	***>0.0001***	12.049	***0.0005***	**4**
SYCP1	3015	16	0.4337	**8**	8.711	***0.0128***	26.860	***>0.0001***	9.243	***0.0024***	**3**
SYCP2	4650	16	0.5572	**8**	11.584	***0.0031***	37.200	***>0.0001***	15.838	***0.0001***	0
TEX12	369	14	0.2297	7	0.0565	*0.9721*	1.549	*0.4610*	—	—	0
E)											
TEX11	2844	15	0.8483	**8**	60.872	***>0.0001***	82.665	***>0.0001***	61.141	***>0.0001***	**14**
SHOC1	4644	16	0.6113	**8**	12.447	***0.0020***	30.561	***>0.0001***	15.645	***0.0001***	0
RNF212	948	16	0.5014	**8**	0	*1.000*	16.366	***0.0003***	5.202	***0.0226***	**1**
RNF212B	906	14	0.4066	7	0	*1.000*	0.500	*0.7788*	—	—	0

Gene	***bp***	***N***	ω	***M***	***M1-M2***	***p-value***	***M7-M8***	***p-value***	***M8a-M8***	***p-value***	***BEB***
MSH4	2814	16	0.2132	**8**	16.608	***0.0002***	39.447	***>0.0001***	23.238	***>0.0001***	**6**
MSH5	2565	15	0.1642	7	0	*1.000*	4.214	*0.1216*	—	—	0
F)											
MER3	4458	16	0.3633	8a	0	*1.000*	12.838	***0.0016***	3.109	*0.0779*	0
CNTD1	1026	15	0.2496	7	0	*1.000*	0.936	*0.6263*	—	—	0
HEI10	831	15	0.1226	7	0	*1.000*	0.250	*0.8826*	—	—	0
MLH1	2313	15	0.1652	8a	0	*1.000*	12.221	***0.0022***	0.280	*0.5970*	0
MLH3	4419	16	0.4444	7	0	*1.000*	3.757	*0.1528*	—	—	0
MUS81	1665	16	0.2124	7	0	*1.000*	0.628	*0.7304*	—	—	0

Table 4: Evolutionary rates and tests for positive selection across mammals at recombination genes after removal of potential MNMs.

Gene	***bp***	***N***	ω	***M***	***M1-M2***	***p-value***	***M7-M8***	***p-value***	***M8a-M8***	***p-value***
IHO1	1824	16	0.6104	7	0	*1.000*	0.258	*0.8789*	—	—
MRE11	2136	16	0.1330	7	0.226	*0.8930*	3.056	*0.2169*	—	—
NBS1	2289	15	0.3413	7	0	*1.000*	1.956	*0.3761*	—	—
REC8	1833	16	0.2905	7	0	*1.000*	5.321	*0.0699*	—	—
RAD21L	1686	15	0.4271	8a	2.329	*0.3121*	9.497	***0.0087***	1.620	*0.2031*
SYCP1	3015	16	0.3731	8a	3.328	*0.1893*	13.440	***0.0012***	2.122	*0.1452*
SYCP2	4650	16	0.4752	7	0	*1.000*	1.758	*0.4151*	—	—
TEX11	2844	15	0.7287	**8**	9.989	***0.0068***	18.776	***0.0001***	10.656	***0.0011***
SHOC1	4644	16	0.5519	8a	0	*1.000*	7.439	***0.0242***	0.292	*0.5887*
RNF212	948	16	0.3685	7	0	*1.000*	0	*1.000*	—	—
MSH4	2814	16	0.1509	7	0	*1.000*	2.079	*0.3536*	—	—

Table 5: Comparisons of polymorphism within humans to divergence between human and macaque at recombination genes. Genes are organized by step in the pathway, as labeled in Figure 3.

Gene	ω	***Pn***	***Ps***	***Pn/Ps***	***Dn***	***Ds***	***Dn/Ds***	***MK Test***	***NI***	***DoS***	
A)											
HORMAD1	0.0901	43	10	4.3	5	12	0.4167	***0.0002***	10.32	-0.5172	Neg.
MEI4	0.7252	9	2	4.5	24	9	2.6667	*0.7013*	1.6875	-0.0909	—
REC114	0.3239	49	21	2.3333	11	14	0.7857	***0.02949***	2.9700	-0.2600	Neg.
IHO1	0.6608	72	28	2.5714	36	19	1.8947	*0.4658*	1.3571	-0.0645	—
SPO11	0.1434	62	28	2.2143	11	22	0.5000	***0.0008***	4.4286	-0.3556	Neg.
B)											
HORMAD2	0.295	50	16	3.125	7	9	0.7778	***0.0177***	4.0179	-0.3201	Neg.
MRE11	0.0392	139	48	2.8958	5	35	0.1429	***>0.0001***	20.2708	-0.6183	Neg.
NBS1	0.4155	119	58	2.0517	34	25	1.3600	*0.2086*	1.5086	-0.0960	—
RAD50	0.0714	168	55	3.0517	8	43	0.1860	***>0.0001***	16.4182	-0.5965	Neg.
BRCC3	0.0979	7	12	0.5833	2	6	0.3333	*0.6758*	1.7500	-0.1184	—
C)											
DMC1	0.000	43	25	1.72	0	11	0.0000	***<0.0001***	—	-0.6324	Neg.
RAD51	0.000	27	29	0.9310	0	13	0.0000	***0.0010***	—	-0.4821	Neg.
SPATA22	0.4523	67	26	2.5769	21	10	2.1000	*0.6535*	1.2271	-0.0430	—
MEIOB	0.2462	45	17	2.6471	20	22	0.9091	***0.0094***	2.9118	-0.2496	Neg.
MCMDC2	0.2108	90	24	3.7500	16	26	0.6154	***<0.0001***	6.0938	-0.4085	Neg.
D)											
REC8	0.477	90	45	2.000	38	31	1.2258	*0.1264*	1.6316	-0.1159	—
RAD21L	0.6334	21	6	3.500	27	13	2.0769	*0.4176*	1.6852	-0.1028	—
SYCP1	0.3676	122	60	2.033	33	37	1.2222	*0.1204*	1.6636	-0.1203	—
SYCP2	0.3676	246	87	2.8276	74	53	1.3962	***0.0015***	2.0252	-0.1561	Neg.
TEX12	0.1349	15	9	1.6667	2	4	0.5000	*0.3598*	3.3333	-0.2917	—
E)											
TEX11	0.9068	78	45	1.7333	55	25	2.200	*0.4541*	0.7879	0.05335	—
SHOC1	0.7225	227	72	3.1528	85	37	2.2973	*0.2199*	1.3724	-0.0625	—
RNF212	0.387	—	—	—	17	18	0.9444	—	—	—	—
RNF212B	0.2566	9	3	3.000	8	12	0.6667	*0.0759*	4.5000	-0.3500	—

Gene	ω	***Pn***	***Ps***	***Pn/Ps***	***Dn***	***Ds***	***Dn/Ds***	***MK Test***	***NI***	***DoS***	
MSH4	0.2635	149	50	2.9800	24	29	0.8276	***<0.0001***	3.6008	-0.2959	Neg.
MSH5	0.2106	129	64	2.0156	19	33	0.5758	***0.0001***	3.5008	-0.3030	Neg.
F)											
MER3	0.3247	236	92	2.5652	54	44	1.2273	***0.0029***	2.0902	-0.1685	Neg.
CNTD1	0.6803	56	29	1.9310	13	8	1.6250	*0.8001*	1.1883	-0.0398	—
HEI10	0.3235	50	21	2.3810	4	5	0.8000	*0.1417*	2.9762	-0.2598	—
MLH1	0.0924	161	48	3.3542	9	29	0.3103	***>0.0001***	10.8079	-0.5335	Neg.
MLH3	0.4919	252	90	2.8	77	57	1.3509	***0.0009***	2.0727	-0.1622	Neg.
MUS81	0.1299	129	49	2.6327	17	40	0.4250	***>0.0001***	6.1945	-0.4265	Neg.

Supplemental Tables

Table S1: Testes Expression Datasets (Barrett et al. 2012)

Species	**GEO Accession**	**Reference**
Bos taurus	GSM1020728 & GSM1020746	Merkin et al. (2012)
Callithrix jacchus	GSM1227961, GSM1227962 & GSM1227963	Cortez et al. (2014)
Canis lupus familiaris	GSM747469 & GSM1359286	Derti et al. (2012), Vandewege et al. (2016)
Eptesicus fuscus	GSM1359287	Vandewege et al. (2016)
Equus caballus	GSM1139276 & GSM1359288	Coleman et al. (2013), Vandewege et al. (2016)
Gorilla gorilla	GSM752663	Brawand et al. (2011)
Homo sapiens	GSM752707 & GSM752708	Brawand et al. (2011)
Macaca mulatta	GSM752642 & GSM752643	Brawand et al. (2011)
Mus musculus	GSM752629 & GSM752630	Brawand et al. (2011)
Ovis aries	GSM1666944 & GSM1666936	Guan et al. (2017)
Pan paniscus	GSM752690	Brawand et al. (2011)
Pan troglodytes	GSM752678	Brawand et al. (2011)
Pongo pygmaeus	GSM1858310 & GSM1858311	Carelli et al. (2016)
Rattus norvegicus	GSM1278058	Cortez et al. (2014)
Sus scrofa	GSM1902350, GSM2033157 & GSM2033163	Li et al. (2016), Yang et al. (2017)
Tupaia chinensis	GSM957062	Fan et al. (2013)

457 **Table S2**: NCBI Reference Genomes (O'Leary et al. 2015)

Species	**Assembly**	**RefSeq Accession**	**WGS Project Reference**
Bos taurus	Bos_taurus_UMD_3.1.1	GCF_000003055.6	Zimin et al. (2009)
Callithrix jacchus	Callithrix_jacchus-3.2	GCF_000004665.1	-
Canis lupus familiaris	CanFam3.1	GCF_000002285.3	Lindblad-Toh et al. (2005)
Eptesicus fuscus	EptFus1.0	GCF_000308155.1	-
Equus caballus	EquCab2.0	GCF_000002305.2	Wade et al. (2009)
Gorilla gorilla	gorGor4	GCF_000151905.2	Scally et al. (2012)
Homo sapiens	GRCh38.p10	GCF_000001405.36	-
Macaca mulatta	Mmul_8.0.1	GCF_000772875.2	Zimin et al. (2014)
Mus musculus	GRCm38.p5	GCF_000001635.25	-
Ovis aries	Oar_v4.0	GCF_000298735.2	International Sheep Genomics Consortium et al.
Pan paniscus	panpan1.1	GCF_000258655.2	Prüfer et al. (2012)
Pan troglodytes	Pan_tro_3.0	GCF_000001515.7	The Chimpanzee Sequencing Analysis Consortiu
Pongo abelii	P_pygmaeus_2.0.2	GCF_000001545.4	Locke et al. (2011)
Rattus norvegicus	Rnor_6.0	GCF_000001895.5	Rat Genome Sequencing Project Consortium an
Sus scrofa	Sscrofa11.1	GCF_000003025.6	-
Tupaia chinensis	TupChi_1.0	GCF_000334495.1	Fan et al. (2013)

458 **Table S3**: Ensembl Reference Genomes (Zerbino et al. 2017)

Species	**Assembly**	**RefSeq Accession**	**WGS Project Reference**
Bos taurus	Bos_taurus_UMD_3.1	GCF_000003055.3	Zimin et al. (2009)
Callithrix jacchus	Callithrix_jacchus-3.2	GCF_000004665.1	-
Canis lupus familiaris	CanFam3.1	GCF_000002285.3	Lindblad-Toh et al. (2005)
Eptesicus fuscus	-	-	-
Equus caballus	EquCab2.0	GCF_000002305.2	Wade et al. (2009)
Gorilla gorilla	gorGor3.1	GCF_000151905.1	-
Homo sapiens	GRCh38.p10	GCF_000001405.36	-
Macaca mulatta	Mmul_8.0.1	GCF_000772875.2	Zimin et al. (2014)
Mus musculus	GRCm38.p5	GCF_000001635.25	-
Ovis aries	Oar_v3.1	GCF_000298735.1	International Sheep Genomics Consortium et al. (
Pan paniscus	panpan1.1	GCF_000258655.2	Prüfer et al. (2012)
Pan troglodytes	CHIMP2.1.4	GCF_000001515.6	The Chimpanzee Sequencing Analysis Consortium
Pongo abelii	PPYG2	GCF_000001545.4	Locke et al. (2011)
Rattus norvegicus	Rnor_6.0	GCF_000001895.5	Rat Genome Sequencing Project Consortium and
Sus scrofa	Sscrofa11.1	GCF_000003025.6	-
Tupaia chinensis	-	-	-

459 **Table S4**: Sequence divergence between human (*Homo sapiens*) and macaque (*Macaca mulatta*) for 32
460 recombination genes (Yang and Nielsen 2000; Yang 2007).

Gene	***bp***	ω	***S***	***N***	***t***	κ	***dN***	***dS***
A)								
HORMAD1	1182	**0.0901**	273.9	908.1	0.0443	3.8819	0.0044 +- 0.0022	0.0490 +- 0.0137
MEI4	1167	**0.7252**	331	824	0.0822	4.6295	0.0247 +/- 0.0056	0.0341 +/- 0.0104
REC114	864	**0.3239**	237.2	557.8	0.0974	2.9455	0.0200 +/- 0.0061	0.0618 +/- 0.0168
IHO1	1797	**0.6608**	509	1273	0.0951	3.6035	0.0276 +- 0.0047	0.0418 +- 0.0094
SPO11	1188	**0.1434**	291.2	896.8	0.0872	2.5317	0.0118 +/- 0.0036	0.0823 +/- 0.0178
B)								
HORMAD2	921	**0.295**	256.7	664.3	0.0531	4.2164	0.0106 +- 0.0040	0.0360 +- 0.0121
MRE11	2124	**0.0392**	479.4	1644.6	0.0597	2.6154	0.0030 +- 0.0014	0.0778 +- 0.0135
NBS1	2265	**0.4155**	553.7	1705.3	0.0804	5.0955	0.0199 +- 0.0035	0.0480 +- 0.0097
RAD50	3969	**0.0714**	1118.7	2817.3	0.0401	5.0903	0.0028 +- 0.0010	0.0399 +- 0.0062
BRCC3	951	**0.0979**	264	609	0.028	4.6	0.0025 +- 0.0020	0.0252 +- 0.0100
C)								
DMC1	1020	**0.0000**	273.7	746.3	0.0335	5.1279	0.0000 +- 0.0000	0.0416 +- 0.0127
RAD51	1017	**0.0000**	306.5	710.5	0.0398	6.7467	0.0000 +- 0.0000	0.0441 +- 0.0124
SPATA22	1089	**0.4523**	247.8	841.2	0.0879	3.6505	0.0230 +- 0.0053	0.0508 +- 0.0150
MEIOB	1413	**0.2462**	348.9	1064.1	0.0927	4.3887	0.0176 +- 0.0041	0.0715 +- 0.0151
MCMDC2	2043	**0.2108**	534	1509	0.0635	7.8547	0.0107 +- 0.0027	0.0507 +- 0.0101
D)								
REC8	1701	**0.477**	497	1138	0.1293	2.8869	0.0323 +- 0.0054	0.0678 +- 0.0122
RAD21L	1680	**0.6334**	427.5	1237.5	0.0735	5.6876	0.0213 +- 0.0042	0.0337 +- 0.0091
SYCP1	2928	**0.3676**	761.6	2166.4	0.0628	4.8307	0.0145 +- 0.0026	0.0393 +- 0.0074
SYCP2	4590	**0.3873**	1070.7	3519.3	0.0854	5.994	0.0208 +- 0.0025	0.0537 +- 0.0074
TEX12	369	**0.1349**	80.2	288.8	0.05	1.9678	0.0070 +- 0.0049	0.0516 +- 0.0260
E)								
TEX11	2775	**0.9068**	805.9	1933.1	0.0897	7.8022	0.0290 +- 0.0040	0.0320 +- 0.0064
SHOC1	4332	**0.7225**	1203	3129	0.0865	9.5737	0.0261 +- 0.0029	0.0361 +- 0.0057
RNF212	816	**0.387**	243.2	572.8	0.1342	4.996	0.0304 +- 0.0074	0.0785 +- 0.0189
RNF212B	900	**0.2566**	255.6	644.4	0.0685	3.4122	0.0125 +- 0.0044	0.0488 +- 0.0143

Gene	***bp***	ω	***S***	***N***	***t***	κ	***dN***	***dS***
MSH4	2808	**0.2635**	731.3	2073.7	0.058	7.5194	0.0112 +- 0.0023	0.0425 +- 0.0079
MSH5	2502	**0.2106**	728.7	1770.3	0.0643	3.9993	0.0102 +- 0.0024	0.0486 +- 0.0085
F)								
MER3	4305	**0.3247**	987.6	3317.4	0.0703	7.0099	0.0159 +- 0.0022	0.0488 +- 0.0074
CNTD1	990	**0.6803**	335.3	651.7	0.065	8.0721	0.0187 +- 0.0054	0.0274 +- 0.0092
HEI10	1059	**0.3235**	241.5	589.5	0.0329	5.9591	0.0068 +- 0.0034	0.0211 +- 0.0095
MLH1	2268	**0.0924**	602.3	1665.7	0.0522	2.4752	0.0048 +- 0.0017	0.0521 +- 0.0097
MLH3	4368	**0.4919**	1209.8	3149.2	0.0949	6.4296	0.0246 +- 0.0028	0.0500 +- 0.0067
MUS81	1653	**0.1299**	465.8	1187.2	0.1106	5.7915	0.0128 +- 0.0033	0.0983 +- 0.0158

Table S5: Evolutionary rates and tests for positive selection across mammals at 32 recombination genes using the gene tree. Genes are organized by step in the pathway, as labeled in Figure 3. (Yang 2007).

Gene	***bp***	***N***	ω	***M***	***M1-M2***	***p-value***	***M7-M8***	***p-value***	***M8a-M8***	***p-value***
A)										
HORMAD1	1212	16	0.3037	7	0	*1.000*	3.135	*0.2086*	—	—
MEI4	1170	16	0.4310	7	0	*1.000*	0.058	*0.9715*	—	—
REC114	870	15	0.4237	7	0	*1.000*	4.1874	*0.1232*	—	—
IHO1	1824	16	0.7099	**8**	13.384	***0.0012***	17.714	***0.0001***	14.707	***0.0001***
SPO11	1188	15	0.1701	7	0	*1.000*	4.697	*0.0955*	—	—
B)										
HORMAD2	981	15	0.3290	1	0	*1.000*	3.881	*0.1436*	—	—
MRE11	2136	16	0.1686	**8**	0.636	*0.7277*	12.014	***0.0025***	4.822	***0.0281***
NBS1	2289	15	0.4185	**8**	0	*1.000*	12.899	***0.0016***	4.298	***0.0382***
RAD50	3936	16	0.0322	1	0	*1.000*	0.5615	*0.7552*	—	—-
BRCC3	954	15	0.0601	7	0	*1.000*	0.573	*0.7509*	—	—
C)										
DMC1	1020	15	0.0365	7	0	*1.000*	4.288	*0.1172*	—	—
RAD51	1017	16	0.0322	1	0	*1.000*	0.562	*0.7552*	—	—
SPATA22	1101	16	0.4932	7	0	*1.000*	0.200	*0.9049*	—	—
MEIOB	1425	16	0.2340	7	0	*1.000*	0.221	*0.8955*	—	—
MCMDC2	2052	16	0.2242	7	0	*1.000*	0.610	*0.7370*	—	—
D)										
REC8	1833	16	0.3698	**8**	0	*1.000*	14.690	***0.0006***	5.927	***0.0149***
RAD21L	1686	15	0.503	**8**	12.124	***0.0023***	32.050	***>0.0001***	12.049	***0.0005***
SYCP1	3015	16	0.4337	**8**	8.711	***0.0128***	26.860	***>0.0001***	9.243	***0.0024***
SYCP2	4650	16	0.5572	**8**	11.584	***0.0031***	37.200	***>0.0001***	15.838	***0.0001***
TEX12	369	14	0.2297	7	0.0565	*0.9721*	1.549	*0.4610*	—	—
E)										
TEX11	2844	15	0.8483	**8**	60.872	***>0.0001***	82.665	***>0.0001***	61.141	***>0.0001***
SHOC1	4644	16	0.6113	**8**	12.447	***0.0020***	30.561	***>0.0001***	15.645	***0.0001***
RNF212	948	16	0.5014	**8**	0	*1.000*	16.366	***0.0003***	5.202	***0.0226***
RNF212B	906	14	0.4066	7	0	*1.000*	0.500	*0.7788*	—	—

Gene	***bp***	***N***	ω	***M***	***M1-M2***	***p-value***	***M7-M8***	***p-value***	***M8a-M8***	***p-value***
MSH4	2814	16	0.2132	**8**	16.608	***0.0002***	39.447	***>0.0001***	23.238	***>0.0001***
MSH5	2565	15	0.1642	7	0	*1.000*	4.214	*0.1216*	—	—
F)										
MER3	4458	16	0.3633	8a	0	*1.000*	12.838	***0.0016***	3.109	*0.0779*
CNTD1	1026	15	0.2496	7	0	*1.000*	0.936	*0.6263*	—	—
HEI10	831	15	0.1226	7	0	*1.000*	0.250	*0.8826*	—	—
MLH1	2313	15	0.1652	8a	0	*1.000*	12.221	***0.0022***	0.280	*0.5970*
MLH3	4419	16	0.4444	7	0	*1.000*	3.757	*0.1528*	—	—
MUS81	1665	16	0.2124	7	0	*1.000*	0.628	*0.7304*	—	—

References

Abascal F, Zardoya R, Telford MJ. 2010. TranslatorX: Multiple alignment of nucleotide sequences guided by amino acid translations. Nucleic acids research 38:W7–W13.

Baker SM, Plug AW, Prolla TA, Bronner CE, Harris AC, Yao X, Christie D-M, Monell C, Arnheim N, Bradley A, et al. 1996. Involvement of mouse Mlh1 in DNA mismatch repair and meiotic crossing over. Nature genetics 13:336.

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

Barbosa-Morais NL, Irimia M, Pan Q, Xiong HY, Gueroussov S, Lee LJ, Slobodeniuc V, Kutter C, Watt S, Çolak R, et al. 2012. The evolutionary landscape of alternative splicing in vertebrate species. Science 338:1587–1593.

Barrett T, Wilhite SE, Ledoux P, Evangelista C, Kim IF, Tomashevsky M, Marshall KA, Phillippy KH, Sherman PM, Holko M, et al. 2012. NCBI geo: Archive for functional genomics data sets—update. Nucleic acids research 41:D991–D995.

Baudat F, Manova K, Yuen JP, Jasin M, Keeney S. 2000. Chromosome synapsis defects and sexually dimorphic meiotic progression in mice lacking Spo11. Molecular cell 6:989–998.

Baudat F, Massy B de. 2007. Regulating double-stranded dna break repair towards crossover or non-crossover during mammalian meiosis. Chromosome research 15:565–577.

Begun DJ, Aquadro CF. 1992. Levels of naturally occurring DNA polymorphism correlate with recombination rates in D. Melanogaster. Nature 356:519.

Bergerat A, Massy B de, Gadelle D, Varoutas P-C, Nicolas A, Forterre P. 1997. An atypical topoisomerase II from Archaea with implications for meiotic recombination. Nature 386:414.

Besenbacher S, Sulem P, Helgason A, Helgason H, Kristjansson H, Jonasdottir A, Jonasdottir A, Magnusson OT, Thorsteinsdottir U, Masson G, et al. 2016. Multi-nucleotide de novo mutations in humans. PLoS genetics 12:e1006315.

Bisig CG, Guiraldelli MF, Kouznetsova A, Scherthan H, Höög C, Dawson DS, Pezza RJ. 2012. Synaptonemal complex components persist at centromeres and are required for homologous centromere pairing in mouse spermatocytes. PLoS genetics 8:e1002701.

Bolcun-Filas E, Schimenti JC. 2012. Genetics of meiosis and recombination in mice. International review of cell and molecular biology 298:179–227.

Brand CL, Cattani MV, Kingan SB, Landeen EL, Presgraves DC. 2018. Molecular evolution at a meiosis gene mediates species differences in the rate and patterning of recombination. Current Biology 28:1289–1295.

Brawand D, Soumillon M, Necsulea A, Julien P, Csárdi G, Harrigan P, Weier M, Liechti A, Aximu-Petri A, Kircher M, et al. 2011. The evolution of gene expression levels in mammalian organs. Nature 478:343.

Broman KW, Murray JC, Sheffield VC, White RL, Weber JL. 1998. Comprehensive human genetic maps: Individual and sex-specific variation in recombination. The American Journal of Human Genetics 63:861–869.

Brown MS, Bishop DK. 2014. DNA strand exchange and RecA homologs in meiosis. Cold Spring Harbor perspectives in biology:a016659.

Burt A, Bell G. 1987. Red queen versus tangled bank models. Nature 330:118.

Carelli FN, Hayakawa T, Go Y, Imai H, Warnefors M, Kaessmann H. 2016. The life history of retrocopies illuminates the evolution of new mammalian genes. Genome research:gr–198473.

Charlesworth B. 1994. The effect of background selection against deleterious mutations on weakly selected, linked variants. Genetics Research 63:213–227.

Charlesworth B, Jarne P, Assimacopoulos S. 1994. The distribution of transposable elements within and between chromosomes in a population of drosophila melanogaster. III. Element abundances in heterochromatin. Genetics Research 64:183–197.

Charlesworth B, Morgan M, Charlesworth D. 1993. The effect of deleterious mutations on neutral molecular variation. Genetics 134:1289–1303.

Chen M-Y, Liang D, Zhang P. 2017. Phylogenomic resolution of the phylogeny of laurasiatherian mammals: Exploring phylogenetic signals within coding and noncoding sequences. Genome biology and evolution 9:1998–2012.

Chowdhury R, Bois PR, Feingold E, Sherman SL, Cheung VG. 2009. Genetic analysis of variation in human meiotic recombination. PLoS genetics 5:e1000648.

Clark NL, Alani E, Aquadro CF. 2012. Evolutionary rate covariation reveals shared functionality and coexpression of genes. Genome research.

Clark NL, Alani E, Aquadro CF. 2013. Evolutionary rate covariation in meiotic proteins results from fluctuating evolutionary pressure in yeasts and mammals. Genetics 193:529–538.

Clark NL, Gasper J, Sekino M, Springer SA, Aquadro CF, Swanson WJ. 2009. Coevolution of interacting fertilization proteins. PLoS genetics 5:e1000570.

Cloud V, Chan Y-L, Grubb J, Budke B, Bishop DK. 2012. Rad51 is an accessory factor for Dmc1-mediated joint molecule formation during meiosis. Science 337:1222–1225.

Coleman SJ, Zeng Z, Hestand MS, Liu J, Macleod JN. 2013. Analysis of unannotated equine transcripts identified by mRNA sequencing. PLoS One 8:e70125.

Comeron JM, Kreitman M, Aguadé M. 1999. Natural selection on synonymous sites is correlated with gene length and recombination in drosophila. Genetics 151:239–249.

Comeron JM, Ratnappan R, Bailin S. 2012. The many landscapes of recombination in drosophila melanogaster. PLoS genetics 8:e1002905.

Coop G, Przeworski M. 2007. An evolutionary view of human recombination. Nature Reviews Genetics 8:23.

Cortez D, Marin R, Toledo-Flores D, Froidevaux L, Liechti A, Waters PD, Gruetzner F, Kaessmann H. 2014. Origins and functional evolution of Y chromosomes across mammals. Nature 508:488.

Costa Y, Speed R, Öllinger R, Alsheimer M, Semple CA, Gautier P, Maratou K, Novak I, Höög C, Benavente R, et al. 2005. Two novel proteins recruited by synaptonemal complex protein 1 (SYCP1) are at the centre of meiosis. Journal of cell science 118:2755–2762.

Dapper AL, Payseur BA. 2017. Connecting theory and data to understand recombination rate evolution. Phil. Trans. R. Soc. B 372:20160469.

Dapper AL, Wade MJ. 2016. The evolution of sperm competition genes: The effect of mating system on levels of genetic variation within and between species. Evolution 70:502–511.

Derti A, Garrett-Engele P, MacIsaac KD, Stevens RC, Sriram S, Chen R, Rohl CA, Johnson JM, Babak T. 2012. A quantitative atlas of polyadenylation in five mammals. Genome research:gr–132563.

Dong Y, Hakimi M-A, Chen X, Kumaraswamy E, Cooch NS, Godwin AK, Shiekhattar R. 2003. Regulation of BRCC, a holoenzyme complex containing BRCA1 and BRCA2, by a signalosome-like subunit and its role in dna repair. Molecular cell 12:1087–1099.

Dumont BL, Payseur BA. 2008. Evolution of the genomic rate of recombination in mammals. Evolution: International Journal of Organic Evolution 62:276–294.

Dumont BL, Payseur BA. 2011. Genetic analysis of genome-scale recombination rate evolution in house mice. PLoS genetics 7:e1002116.

Dumont BL, White MA, Steffy B, Wiltshire T, Payseur BA. 2011. Extensive recombination rate variation in the house mouse species complex inferred from genetic linkage maps. Genome research 21:114–125.

Duret L, Arndt PF. 2008. The impact of recombination on nucleotide substitutions in the human genome. PLoS genetics 4:e1000071.

Duret L, Mouchiroud D. 1999. Expression pattern and, surprisingly, gene length shape codon usage in Caenorhabditis, Drosophila, and Arabidopsis. Proceedings of the National Academy of Sciences 96:4482–4487.

Edelmann W, Cohen PE, Kane M, Lau K, Morrow B, Bennett S, Umar A, Kunkel T, Cattoretti G, Chaganti R, et al. 1996. Meiotic pachytene arrest in MLH1-deficient mice. Cell 85:1125–1134.

Edgar RC. 2004. MUSCLE: Multiple sequence alignment with high accuracy and high throughput. Nucleic acids research 32:1792–1797.

Fan Y, Huang Z-Y, Cao C-C, Chen C-S, Chen Y-X, Fan D-D, He J, Hou H-L, Hu L, Hu X-T, et al. 2013. Genome of the Chinese tree shrew. Nature communications 4:1426.

Fay JC, Wyckoff GJ, Wu C-I. 2001. Positive and negative selection on the human genome. Genetics 158:1227–1234.

Felsenstein J. 1974. The evolutionary advantage of recombination. Genetics 78:737–756.

Finsterbusch F, Ravindranathan R, Dereli I, Stanzione M, Tränkner D, Tóth A. 2016. Alignment of homologous chromosomes and effective repair of programmed dna double-strand breaks during mouse meiosis require the minichromosome maintenance domain containing 2 (MCMDC2) protein. PLoS genetics 12:e1006393.

Fledel-Alon A, Leffler EM, Guan Y, Stephens M, Coop G, Przeworski M. 2011. Variation in human recombination rates and its genetic determinants. PloS one 6:e20321.

Fraune J, Alsheimer M, Redolfi J, Brochier-Armanet C, Benavente R. 2014. Protein SYCP2 is an ancient component of the metazoan synaptonemal complex. Cytogenetic and genome research 144:299–305.

Gonen S, Battagin M, Johnston SE, Gorjanc G, Hickey JM. 2017. The potential of shifting recombination hotspots to increase genetic gain in livestock breeding. Genetics Selection Evolution 49:55.

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

Grey C, Baudat F, Massy B de. 2018. PRDM9, a driver of the genetic map. PLoS genetics 14:e1007479.

578 Guan Y, Liang G, Martin GB, others. 2017. Functional changes in mRNA expression and alternative
579 pre-mRNA splicing associated with the effects of nutrition on apoptosis and spermatogenesis in the adult
580 testis. BMC genomics 18:64.

581 Guiraldelli MF, Felberg A, Almeida LP, Parikh A, Castro RO de, Pezza RJ. 2018. SHOC1 is a ERCC4-(HhH)
582 2-like protein, integral to the formation of crossover recombination intermediates during mammalian meiosis.
583 PLoS genetics 14:e1007381.

584 Hakes L, Lovell SC, Oliver SG, Robertson DL. 2007. Specificity in protein interactions and its relationship
585 with sequence diversity and coevolution. Proceedings of the National Academy of Sciences 104:7999–8004.

586 Hamer G, Gell K, Kouznetsova A, Novak I, Benavente R, Höög C. 2006. Characterization of a novel
587 meiosis-specific protein within the central element of the synaptonemal complex. Journal of cell science
588 119:4025–4032.

589 Hamer G, Wang H, Bolcun-Filas E, Cooke HJ, Benavente R, Höög C. 2008. Progression of meiotic recombi-
590 nation requires structural maturation of the central element of the synaptonemal complex. Journal of cell
591 science 121:2445–2451.

592 Hassold T, Hunt P. 2001. To err (meiotically) is human: The genesis of human aneuploidy. Nature Reviews
593 Genetics 2:280.

594 Hernández-Hernández A, Masich S, Fukuda T, Kouznetsova A, Sandin S, Daneholt B, Höög C. 2016. The
595 central element of the synaptonemal complex in mice is organized as a bilayered junction structure. J Cell
596 Sci 129:2239–2249.

597 Hill WG, Robertson A. 1966. The effect of linkage on limits to artificial selection. Genetics Research
598 8:269–294.

599 Hopfner K-P. 2005. Structure and function of Rad50/SMC protein complexes in chromosome biology. In:
600 Genome integrity. Springer. pp. 201–218.

601 Hunter CM, Huang W, Mackay TF, Singh ND. 2016. The genetic architecture of natural variation in
602 recombination rate in drosophila melanogaster. PLoS genetics 12:e1005951.

603 International Sheep Genomics Consortium, Archibald A, Cockett N, Dalrymple B, Faraut T, Kijas J, Maddox
604 J, McEwan J, Hutton Oddy V, Raadsma H, et al. 2010. The sheep genome reference sequence: A work in
605 progress. Animal genetics 41:449–453.

606 Jeffreys AJ, Neumann R, Panayi M, Myers S, Donnelly P. 2005. Human recombination hot spots hidden in

607 regions of strong marker association. Nature genetics 37:601.

608 Johnston SE, Bérénos C, Slate J, Pemberton JM. 2016. Conserved genetic architecture underlying individual
609 recombination rate variation in a wild population of Soay sheep (Ovis aries). Genetics:genetics–115.

610 Johnston SE, Huisman J, Pemberton JM. 2018. A genomic region containing REC8 and RNF212B is
611 associated with individual recombination rate variation in a wild population of red deer (Cervus elaphus).
612 G3: Genes, Genomes, Genetics:g3–200063.

613 Kadri NK, Harland C, Faux P, Cambisano N, Karim L, Coppieters W, Fritz S, Mullaart E, Baurain D,
614 Boichard D, et al. 2016. Coding and noncoding variants in HFM1, MLH3, MSH4, MSH5, RNF212, and
615 RNF212B affect recombination rate in cattle. Genome research.

616 Keeney S. 2007. Spo11 and the formation of DNA double-strand breaks in meiosis. In: Recombination and
617 meiosis. Springer. pp. 81–123.

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

620 Kobayashi W, Takaku M, Machida S, Tachiwana H, Maehara K, Ohkawa Y, Kurumizaka H. 2016. Chromatin
621 architecture may dictate the target site for DMC1, but not for RAD51, during homologous pairing. Scientific
622 reports 6:24228.

623 Kohl KP, Jones CD, Sekelsky J. 2012. Evolution of an MCM complex in flies that promotes meiotic crossovers
624 by blocking BLM helicase. Science 338:1363–1365.

625 Kong A, Thorleifsson G, Frigge ML, Masson G, Gudbjartsson DF, Villemoes R, Magnusdottir E, Olafsdottir
626 SB, Thorsteinsdottir U, Stefansson K. 2014. Common and low-frequency variants associated with genome-wide
627 recombination rate. Nature genetics 46:11.

628 Kong A, Thorleifsson G, Gudbjartsson DF, Masson G, Sigurdsson A, Jonasdottir A, Walters GB, Jonasdottir
629 A, Gylfason A, Kristinsson KT, et al. 2010. Fine-scale recombination rate differences between sexes,
630 populations and individuals. Nature 467:1099.

631 Kong A, Thorleifsson G, Stefansson H, Masson G, Helgason A, Gudbjartsson DF, Jonsdottir GM, Gudjonsson
632 SA, Sverrisson S, Thorlacius T, et al. 2008. Sequence variants in the RNF212 gene associate with genome-wide
633 recombination rate. Science 319:1398–1401.

634 Kumar R, Ghyselinck N, Ishiguro K-i, Watanabe Y, Kouznetsova A, Höög C, Strong E, Schimenti J, Daniel K,
635 Toth A, et al. 2015. MEI4: A central player in the regulation of meiotic dna double strand break formation

in the mouse. J Cell Sci:jcs–165464.

Lange J, Yamada S, Tischfield SE, Pan J, Kim S, Zhu X, Socci ND, Jasin M, Keeney S. 2016. The landscape of mouse meiotic double-strand break formation, processing, and repair. Cell 167:695–708.

Langmead B, Salzberg SL. 2012. Fast gapped-read alignment with Bowtie 2. Nature methods 9:357.

La Salle S, Palmer K, O'Brien M, Schimenti JC, Eppig J, Handel MA. 2012. Spata22, a novel vertebrate-specific gene, is required for meiotic progress in mouse germ cells. Biology of reproduction 86:45–41.

Latrille T, Duret L, Lartillot N. 2017. The red queen model of recombination hot-spot evolution: A theoretical investigation. Phil. Trans. R. Soc. B 372:20160463.

Lee J, Hirano T. 2011. RAD21L, a novel cohesin subunit implicated in linking homologous chromosomes in mammalian meiosis. The Journal of cell biology 192:263–276.

Leinonen R, Sugawara H, Shumway M, Collaboration INSD. 2010. The sequence read archive. Nucleic acids research 39:D19–D21.

Lek M, Karczewski KJ, Minikel EV, Samocha KE, Banks E, Fennell T, O'Donnell-Luria AH, Ware JS, Hill AJ, Cummings BB, et al. 2016. Analysis of protein-coding genetic variation in 60,706 humans. Nature 536:285.

Lesecque Y, Glémin S, Lartillot N, Mouchiroud D, Duret L. 2014. The red queen model of recombination hotspots evolution in the light of archaic and modern human genomes. PLoS genetics 10:e1004790.

Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G, Durbin R. 2009. The sequence alignment/map format and samtools. Bioinformatics 25:2078–2079.

Li Y, Li J, Fang C, Shi L, Tan J, Xiong Y, Fan B, Li C. 2016. Genome-wide differential expression of genes and small rnas in testis of two different porcine breeds and at two different ages. Scientific reports 6:26852.

Liao B-Y, Scott NM, Zhang J. 2006. Impacts of gene essentiality, expression pattern, and gene compactness on the evolutionary rate of mammalian proteins. Molecular biology and evolution 23:2072–2080.

Lindblad-Toh K, Wade CM, Mikkelsen TS, Karlsson EK, Jaffe DB, Kamal M, Clamp M, Chang JL, Kulbokas III EJ, Zody MC, et al. 2005. Genome sequence, comparative analysis and haplotype structure of the domestic dog. Nature 438:803.

Lipkin SM, Moens PB, Wang V, Lenzi M, Shanmugarajah D, Gilgeous A, Thomas J, Cheng J, Touchman JW, Green ED, et al. 2002. Meiotic arrest and aneuploidy in MLH3-deficient mice. Nature genetics 31:385.

Locke DP, Hillier LW, Warren WC, Worley KC, Nazareth LV, Muzny DM, Yang S-P, Wang Z, Chinwalla AT, Minx P, et al. 2011. Comparative and demographic analysis of orang-utan genomes. Nature 469:529.

Lu Y, Rausher MD. 2003. Evolutionary rate variation in anthocyanin pathway genes. Molecular biology and evolution 20:1844–1853.

Ma L, O'Connell JR, VanRaden PM, Shen B, Padhi A, Sun C, Bickhart DM, Cole JB, Null DJ, Liu GE, et al. 2015. Cattle sex-specific recombination and genetic control from a large pedigree analysis. PLoS genetics 11:e1005387.

McDonald JH, Kreitman M. 1991. Adaptive protein evolution at the Adh locus in Drosophila. Nature 351:652.

Merkin J, Russell C, Chen P, Burge CB. 2012. Evolutionary dynamics of gene and isoform regulation in mammalian tissues. Science 338:1593–1599.

Meuwissen R, Offenberg HH, Dietrich A, Riesewijk A, Iersel M van, Heyting C. 1992. A coiled-coil related protein specific for synapsed regions of meiotic prophase chromosomes. The EMBO Journal 11:5091.

Murdoch B, Owen N, Shirley S, Crumb S, Broman KW, Hassold T. 2010. Multiple loci contribute to genome-wide recombination levels in male mice. Mammalian genome 21:550–555.

Myers S, Bowden R, Tumian A, Bontrop RE, Freeman C, MacFie TS, McVean G, Donnelly P. 2010. Drive against hotspot motifs in primates implicates the prdm9 gene in meiotic recombination. Science 327:876–879.

Oh J, Al-Zain A, Cannavo E, Cejka P, Symington LS. 2016. Xrs2 dependent and independent functions of the Mre11-Rad50 complex. Molecular cell 64:405–415.

O'Leary NA, Wright MW, Brister JR, Ciufo S, Haddad D, McVeigh R, Rajput B, Robbertse B, Smith-White B, Ako-Adjei D, et al. 2015. Reference sequence (RefSeq) database at NCBI: Current status, taxonomic expansion, and functional annotation. Nucleic acids research 44:D733–D745.

Oliver PL, Goodstadt L, Bayes JJ, Birtle Z, Roach KC, Phadnis N, Beatson SA, Lunter G, Malik HS, Ponting CP. 2009. Accelerated evolution of the Prdm9 speciation gene across diverse metazoan taxa. PLoS genetics 5:e1000753.

Page SL, Hawley RS. 2004. The genetics and molecular biology of the synaptonemal complex. Annu. Rev. Cell Dev. Biol. 20:525–558.

Pamilo P, Nei M. 1988. Relationships between gene trees and species trees. Molecular biology and evolution 5:568–583.

693 Pan Q, Shai O, Lee LJ, Frey BJ, Blencowe BJ. 2008. Deep surveying of alternative splicing complexity in the
694 human transcriptome by high-throughput sequencing. Nature genetics 40:1413.

695 Parisi S, McKay MJ, Molnar M, Thompson MA, Van Der Spek PJ, Drunen-Schoenmaker E van, Kanaar R,
696 Lehmann E, Hoeijmakers JH, Kohli J. 1999. Rec8p, a meiotic recombination and sister chromatid cohesion
697 phosphoprotein of the Rad21p family conserved from fission yeast to humans. Molecular and cellular biology
698 19:3515–3528.

699 Parvanov ED, Petkov PM, Paigen K. 2010. Prdm9 controls activation of mammalian recombination hotspots.
700 Science 327:835–835.

701 Pazos F, Valencia A. 2001. Similarity of phylogenetic trees as indicator of protein–protein interaction. Protein
702 engineering 14:609–614.

703 Perelman P, Johnson WE, Roos C, Seuánez HN, Horvath JE, Moreira MA, Kessing B, Pontius J, Roelke M,
704 Rumpler Y, et al. 2011. A molecular phylogeny of living primates. PLoS genetics 7:e1001342.

705 Petit M, Astruc J-M, Sarry J, Drouilhet L, Fabre S, Moreno C, Servin B. 2017. Variation in recombination
706 rate and its genetic determinism in sheep populations. Genetics:genetics–300123.

707 Piovesan D, Tabaro F, Paladin L, Necci M, Mičetić I, Camilloni C, Davey N, Dosztányi Z, Mészáros B,
708 Monzon AM, et al. 2017. MobiDB 3.0: More annotations for intrinsic disorder, conformational diversity and
709 interactions in proteins. Nucleic acids research 46:D471–D476.

710 Prasad AB, Allard MW, Program NCS, Green ED. 2008. Confirming the phylogeny of mammals by use of
711 large comparative sequence data sets. Molecular Biology and Evolution 25:1795–1808.

712 Priedigkeit N, Wolfe N, Clark NL. 2015. Evolutionary signatures amongst disease genes permit novel methods
713 for gene prioritization and construction of informative gene-based networks. PLoS genetics 11:e1004967.

714 Prüfer K, Munch K, Hellmann I, Akagi K, Miller JR, Walenz B, Koren S, Sutton G, Kodira C, Winer R, et
715 al. 2012. The bonobo genome compared with the chimpanzee and human genomes. Nature 486:527.

716 Qiao H, Rao HP, Yang Y, Fong JH, Cloutier JM, Deacon DC, Nagel KE, Swartz RK, Strong E, Holloway
717 JK, et al. 2014. Antagonistic roles of ubiquitin ligase HEI10 and SUMO ligase RNF212 regulate meiotic
718 recombination. Nature genetics 46:194.

719 Rakshambikai R, Srinivasan N, Nishant KT. 2013. Structural insights into saccharomyces cerevisiae Msh4–
720 Msh5 complex function using homology modeling. PLoS One 8:e78753.

721 Rao HP, Qiao H, Bhatt SK, Bailey LR, Tran HD, Bourne SL, Qiu W, Deshpande A, Sharma AN, Beebout

CJ, et al. 2017. A sumo-ubiquitin relay recruits proteasomes to chromosome axes to regulate meiotic recombination. Science 355:403–407.

Rat Genome Sequencing Project Consortium, others. 2004. Genome sequence of the brown norway rat yields insights into mammalian evolution. Nature 428:493.

Rausher MD, Miller RE, Tiffin P. 1999. Patterns of evolutionary rate variation among genes of the anthocyanin biosynthetic pathway. Molecular biology and evolution 16:266–274.

R Core Team. 2015. R: A language and environment for statistical computing.

Reynolds A, Qiao H, Yang Y, Chen JK, Jackson N, Biswas K, Holloway JK, Baudat F, De Massy B, Wang J, et al. 2013. RNF212 is a dosage-sensitive regulator of crossing-over during mammalian meiosis. Nature genetics 45:269.

Ritz KR, Noor MA, Singh ND. 2017. Variation in recombination rate: Adaptive or not? Trends in Genetics 33:364–374.

Rogacheva MV, Manhart CM, Chen C, Guarne A, Surtees J, Alani E. 2014. Mlh1-Mlh3, a meiotic crossover and DNA mismatch repair factor, is a Msh2-Msh3-stimulated endonuclease. Journal of Biological Chemistry:jbc–M113.

Romanienko PJ, Camerini-Otero RD. 2000. The mouse Spo11 gene is required for meiotic chromosome synapsis. Molecular cell 6:975–987.

Ronquist F, Teslenko M, Van Der Mark P, Ayres DL, Darling A, Höhna S, Larget B, Liu L, Suchard MA, Huelsenbeck JP. 2012. MrBayes 3.2: Efficient Bayesian phylogenetic inference and model choice across a large model space. Systematic biology 61:539–542.

Rosenberg NA. 2002. The probability of topological concordance of gene trees and species trees. Theoretical population biology 61:225–247.

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

Scally A, Dutheil JY, Hillier LW, Jordan GE, Goodhead I, Herrero J, Hobolth A, Lappalainen T, Mailund T, Marques-Bonet T, et al. 2012. Insights into hominid evolution from the gorilla genome sequence. Nature 483:169.

Schmekel K, Daneholt B. 1995. The central region of the synaptonemal complex revealed in three dimensions. Trends in cell biology 5:239–242.

751 Schramm S, Fraune J, Naumann R, Hernandez-Hernandez A, Höög C, Cooke HJ, Alsheimer M, Benavente
752 R. 2011. A novel mouse synaptonemal complex protein is essential for loading of central element proteins,
753 recombination, and fertility. PLoS genetics 7:e1002088.

754 Schrider DR, Hourmozdi JN, Hahn MW. 2011. Pervasive multinucleotide mutational events in eukaryotes.
755 Current Biology 21:1051–1054.

756 Scornavacca C, Galtier N. 2017. Incomplete lineage sorting in mammalian phylogenomics. Systematic biology
757 66:112–120.

758 Segura J, Ferretti L, Ramos-Onsins S, Capilla L, Farré M, Reis F, Oliver-Bonet M, Fernández-Bellón H,
759 Garcia F, Garcia-Caldés M, et al. 2013. Evolution of recombination in eutherian mammals: Insights into
760 mechanisms that affect recombination rates and crossover interference. Proceedings of the Royal Society of
761 London B: Biological Sciences 280:20131945.

762 Shen B, Jiang J, Seroussi E, Liu GE, Ma L. 2018. Characterization of recombination features and the genetic
763 basis in multiple cattle breeds. BMC genomics 19:304.

764 Smith NG, Eyre-Walker A. 2002. Adaptive protein evolution in Drosophila. Nature 415:1022.

765 Smukowski C, Noor M. 2011. Recombination rate variation in closely related species. Heredity 107:496.

766 Snowden T, Acharya S, Butz C, Berardini M, Fishel R. 2004. hMSH4-hMSH5 recognizes Holliday junctions
767 and forms a meiosis-specific sliding clamp that embraces homologous chromosomes. Molecular cell 15:437–451.

768 Stanzione M, Baumann M, Papanikos F, Dereli I, Lange J, Ramlal A, Tränkner D, Shibuya H, Massy B de,
769 Watanabe Y, et al. 2016. Meiotic DNA break formation requires the unsynapsed chromosome axis-binding
770 protein IHO1 (CCDC36) in mice. Nature cell biology 18:1208.

771 Stapley J, Feulner PG, Johnston SE, Santure AW, Smadja CM. 2017. Variation in recombination frequency
772 and distribution across eukaryotes: Patterns and processes. Phil. Trans. R. Soc. B 372:20160455.

773 Stoletzki N, Eyre-Walker A. 2010. Estimation of the neutrality index. Molecular biology and evolution
774 28:63–70.

775 Swanson WJ, Nielsen R, Yang Q. 2003. Pervasive adaptive evolution in mammalian fertilization proteins.
776 Molecular biology and evolution 20:18–20.

777 Swanson WJ, Vacquier VD. 2002. The rapid evolution of reproductive proteins. Nature reviews genetics
778 3:137.

779 The Chimpanzee Sequencing Analysis Consortium, Waterson RH, Lander ES, Wilson RK. 2005. Initial
780 sequence of the chimpanzee genome and comparison with the human genome. Nature 437:69.

781 Thomas JH, Emerson RO, Shendure J. 2009. Extraordinary molecular evolution in the PRDM9 fertility gene.
782 PloS one 4:e8505.

783 Thorvaldsdóttir H, Robinson JT, Mesirov JP. 2013. Integrative Genomics Viewer (IGV): High-performance
784 genomics data visualization and exploration. Briefings in bioinformatics 14:178–192.

785 Trapnell C, Pachter L, Salzberg SL. 2009. TopHat: Discovering splice junctions with RNA-Seq. Bioinformatics
786 25:1105–1111.

787 Ubeda F, Wilkins J. 2011. The red queen theory of recombination hotspots. Journal of evolutionary biology
788 24:541–553.

789 Vandewege MW, Platt RN, Ray DA, Hoffmann FG. 2016. Transposable element targeting by piRNAs in
790 Laurasiatherians with distinct transposable element histories. Genome biology and evolution 8:1327–1337.

791 Venkat A, Hahn MW, Thornton JW. 2018. Multinucleotide mutations cause false inferences of lineage-specific
792 positive selection. Nature ecology & evolution 2:1280.

793 Vries FA de, Boer E de, Bosch M van den, Baarends WM, Ooms M, Yuan L, Liu J-G, Zeeland AA van, Heyting
794 C, Pastink A. 2005. Mouse sycp1 functions in synaptonemal complex assembly, meiotic recombination, and
795 xy body formation. Genes & development 19:1376–1389.

796 Vries SS de, Baart EB, Dekker M, Siezen A, Rooij DG de, Boer P de, Riele H te. 1999. Mouse MutS-like
797 protein Msh5 is required for proper chromosome synapsis in male and female meiosis. Genes & Development
798 13:523–531.

799 Wade C, Giulotto E, Sigurdsson S, Zoli M, Gnerre S, Imsland F, Lear T, Adelson D, Bailey E, Bellone R, et
800 al. 2009. Genome sequence, comparative analysis, and population genetics of the domestic horse. Science
801 326:865–867.

802 Ward JO, Reinholdt LG, Motley WW, Niswander LM, Deacon DC, Griffin LB, Langlais KK, Backus VL,
803 Schimenti KJ, O'Brien MJ, et al. 2007. Mutation in mouse hei10, an e3 ubiquitin ligase, disrupts meiotic
804 crossing over. PLoS genetics 3:e139.

805 Wheeler DL, Barrett T, Benson DA, Bryant SH, Canese K, Chetvernin V, Church DM, DiCuccio M, Edgar
806 R, Federhen S, et al. 2006. Database resources of the national center for biotechnology information. Nucleic
807 acids research 35:D5–D12.

Xu Y, Greenberg RA, Schonbrunn E, Wang PJ. 2017. Meiosis-specific proteins MEIOB and SPATA22 cooperatively associate with the single-stranded DNA-binding replication protein A complex and DNA double-strand breaks. Biology of reproduction 96:1096–1104.

Yang F, De La Fuente R, Leu NA, Baumann C, McLaughlin KJ, Wang PJ. 2006. Mouse SYCP2 is required for synaptonemal complex assembly and chromosomal synapsis during male meiosis. The Journal of Cell Biology 173:497–507.

Yang F, Gell K, Van Der Heijden GW, Eckardt S, Leu NA, Page DC, Benavente R, Her C, Höög C, McLaughlin KJ, et al. 2008. Meiotic failure in male mice lacking an X-linked factor. Genes & development 22:682–691.

Yang F, Silber S, Leu NA, Oates RD, Marszalek JD, Skaletsky H, Brown LG, Rozen S, Page DC, Wang PJ. 2015. TEX11 is mutated in infertile men with azoospermia and regulates genome-wide recombination rates in mouse. EMBO molecular medicine 7:1198–1210.

Yang Y, Liang G, Niu G, Zhang Y, Zhou R, Wang Y, Mu Y, Tang Z, Li K. 2017. Comparative analysis of dna methylome and transcriptome of skeletal muscle in lean-, obese-, and mini-type pigs. Scientific reports 7:39883.

Yang Z. 1997. PAML: A program package for phylogenetic analysis by maximum likelihood. Bioinformatics 13:555–556.

Yang Z. 2007. PAML 4: Phylogenetic analysis by maximum likelihood. Molecular Biology and Evolution 24:1586–1591.

Yang Z, Nielsen R. 2000. Estimating synonymous and nonsynonymous substitution rates under realistic evolutionary models. Molecular Biology and Evolution 17:32–43.

Zerbino DR, Achuthan P, Akanni W, Amode MR, Barrell D, Bhai J, Billis K, Cummins C, Gall A, Girón CG, et al. 2017. Ensembl 2018. Nucleic acids research 46:D754–D761.

Zimin AV, Cornish AS, Maudhoo MD, Gibbs RM, Zhang X, Pandey S, Meehan DT, Wipfler K, Bosinger SE, Johnson ZP, et al. 2014. A new rhesus macaque assembly and annotation for next-generation sequencing analyses. Biology direct 9:20.

Zimin AV, Delcher AL, Florea L, Kelley DR, Schatz MC, Puiu D, Hanrahan F, Pertea G, Van Tassell CP, Sonstegard TS, et al. 2009. A whole-genome assembly of the domestic cow, Bos taurus. Genome biology 10:R42.