

Why does the target-recognition site in PRDM9 have unusually high mutation rates?

Frédéric Fyon¹ and Francisco Úbeda¹

Department of Biology, Royal Holloway University of London, Egham, United Kingdom.

Handling editor: Rich Meisel, Guest editor: Denis Roze

Corresponding authors: Frédéric Fyon, Department of Biology, Royal Holloway University of London, Egham TW20 0EX, United Kingdom.

Email: frederic.fyon@rhul.ac.uk; Francisco Úbeda, Department of Biology, Royal Holloway University of London, Egham TW20 0EX, United Kingdom.

Email: francisco.ubeda@rhul.ac.uk

Abstract

The gene PRDM9, responsible for the location of recombination hotspots in mammals, exhibits one of the highest mutation rates in the human genome. This has a direct impact on the life expectancy of recombination hotspots and the variability of the recombination landscape. Why would PRDM9 evolve such a high mutation rate? Here we explore this question by formulating and analysing population genetics models for modifiers of mutation rates at PRDM9 and modifiers of mutation rates at PRDM9's target sites. By letting mutation rates evolve, we find that natural selection indeed favours the evolution of high mutation rates at PRDM9 and low mutation rates at PRDM9's target sites. However, considerations about linkage disequilibrium and genetic drift suggest that only high mutation rates at PRDM9 are likely to evolve. Our modifiers of mutation rates at PRDM9 are most likely to evolve in close linkage with the PRDM9 locus. This is consistent with the binding domain in PRDM9 adopting a zinc-finger structure—molecular structure with high intrinsic mutation rates. Interestingly, our results are an exception to the reduction principle in modifier theory. We draw parallels between our results and another exception to the reduction principle: when genes are evolving in fluctuating environments. We suggest that the red-queen dynamics that characterize the coevolution between PRDM9 and its target can generate a fluctuating genomic environment that drives the evolution of our modifier of mutation rates at PRDM9.

Keywords: recombination hotspot, population genetics, evolution, zinc fingers

Introduction

Meiotic recombination—the process by which genetic associations get reshuffled—occupies a central place in evolutionary genetics. It is involved in questions as diverse as the efficacy of adaptation (Burt, 2000; Otto & Barton, 2001), genetic diversity (Burt, 2000; Hudson & Kaplan, 1995), speciation (Nachman & Payseur, 2012), the evolution of sex (Barton & Charlesworth, 1998; Otto & Lenormand, 2002) and sex chromosomes (Bachtrog, 2013; Lenormand et al., 2020; Úbeda et al., 2015), hitch-hiking selection (Fyon et al., 2015), or the evolution of selfish genetic elements (Haig & Grafen, 1991; Úbeda & Haig, 2005). Its ubiquity, complex mechanism, controversial benefits, and diverse evolutionary consequences make it one of the most relevant topics in modern genetics.

From an evolutionary perspective, one of the most intriguing aspects of recombination is its spatial distribution within the genome (Lenormand et al., 2016). In most eukaryotic species, recombination is not uniformly distributed within the genome but concentrated in small regions of high recombinatorial activity: *recombination hotspots* (Baudat et al., 2013; Paigen & Petkov, 2010; Petes, 2001). In many vertebrates—including humans, most mammals, many fishes, many amphibians, and many reptilians (Cavassim et al., 2022)—the localization of these hotspots is directed by a protein called PRDM9 (Baudat et al., 2010; Myers et al., 2008, 2010).

PRDM9 protein consists of various functional domains (Paigen & Petkov, 2018). PRDM9's zinc-finger domain is

formed by a number of repeated sequences, and its function is to recognize and bind DNA motifs (Paigen & Petkov, 2018). Once PRDM9 is attached to a specific DNA sequence, its PR/SET domain methylates nearby histones, regulating chromatin decondensation (Powers et al., 2016; Wu et al., 2013). Finally, PRDM9's KRAB domain helps recruiting the double-strand break (DSB) machinery (Imai et al., 2017). It is the DSB that initiates recombination. All these domains have evolved under very constrained selection with the exception of the zinc-finger domain (Baker et al., 2017; Cavassim et al., 2022). The zinc finger has evolved very rapidly and show signs of strong positive selection to the point that it is considered one of the fastest-evolving sequences in the genome (Baker et al., 2017; Ponting, 2011).

Population genetic models show that the rapid evolution of the PRDM9's zinc-finger domain is likely to be the result of the antagonistic coevolution of gene conversion and fertility selection (Úbeda & Wilkins, 2011; Úbeda et al., 2023, 2019). By inducing a DSB in the sequence that the zinc finger recognizes, PRDM9 converts the motif that is recognized into the motif that is not (Baudat et al., 2013; Jeffreys & Neumann, 2002, 2005; Petes, 2001). Conversion of the allele that allows recombination (*hot allele*) into the allele that prevents recombination (*cold allele*) leads to the loss of hot alleles and their associated recombination hotspots (Boulton et al., 1997). However, recombination hotspots are abundant (Arnheim et al., 2007; Baudat et al., 2013; Myers et al., 2005; Paigen & Petkov, 2010). The mismatch between predicted rarity and observed

Received June 19, 2024; revised March 17, 2025; accepted May 20, 2025

© The Author(s) 2025. Published by Oxford University Press on behalf of the European Society of Evolutionary Biology. All rights reserved. For commercial re-use, please contact reprints@oup.com for reprints and translation rights for reprints. All other permissions can be obtained through our RightsLink service via the Permissions link on the article page on our site—for further information please contact journals.permissions@oup.com

abundance of recombination hotspots is referred to as the *Recombination Hotspot Paradox* (Boulton et al., 1997).

What saves recombination hotspots from their extinction is that crossover events impact meiotic synapsis and proper chromosome segregation into gametes (Fledel-Alon et al., 2009; Ren et al., 2016). Lack of crossover events results in aneuploid gametes, which translates into reduced fertility (Fledel-Alon et al., 2009; Ren et al., 2016). Population genetic models show that the strength of selection for hotspots is low when they are abundant but increases as they become rare (Úbeda et al., 2019, 2023). As the number of recombination hotspots fades away, fertility selection drives the birth of new hotspots—thus preventing their extinction (Úbeda et al., 2019, 2023). This alternation of the relative strengths of conversion and selection results in an evolutionary treadmill that can explain the rapid change of the recombinatorial landscape observed in species with a functional PRDM9 (Coop et al., 2008; Ptak et al., 2005; Stevison et al., 2015; Winckler et al., 2005; Úbeda et al., 2019, 2023).

It is only recently that the role of mutations in the coevolution between PRDM9 and its target sites has been investigated (Úbeda et al., 2023). Population genetic models show that mutations are not only necessary to maintain the number of hotspots but also responsible for the speed with which recombinatorial landscapes change (Úbeda & Wilkins, 2011; Úbeda et al., 2023). In addition, these models show that larger mutation rates reduce the variance of genome-wide crossover rate (Úbeda et al., 2023). Because mutation rates affect genome-wide crossover rates—that are under selection—they are likely to be under selection. The evolution of mutation rates in PRDM9 and its targets has not been investigated in previous work. Here, we fill this gap by exploring how selection acts on modifiers of the mutation rate on the target recognition domain of gene PRDM9 or on PRDM9's target sequences. We attempt to shed light on (i) why PRDM9 has an unusually high mutation rate, (ii) why the DNA-binding domain of PRDM9 consists of repeated sequences, a highly mutable genetic architecture, and (iii) how mutation rates may impact the rapid evolution of the zinc-finger domain of PRDM9.

Materials and methods

Model

We consider a three-locus two-allele Wright–Fisher model that builds upon a two-locus two-allele model of the coevolution between PRDM9 and one of its targets (Úbeda et al., 2023).

The model for the coevolution between PRDM9 and one its targets considers two loci: locus *A* representing PRDM9 locus and locus *B* representing one of its target sites. The PRDM9 locus carries alleles coding for proteins that recognize and bind matching sequences at its target locus (Úbeda et al., 2019, 2023). In particular, we assume that allele A_1 in locus *A* binds target sequence B_1 in locus *B*, but does not bind sequence B_2 . Similarly, A_2 binds sequence B_2 , but does not bind sequence B_1 (Úbeda et al., 2019, 2023). Binding induces a DSB that initiates recombination (Úbeda et al., 2019, 2023). Recombination may result in biased gene conversion of the sequence that breaks (hot allele) into the sequence that does not break (cold allele) with probability c (Úbeda et al., 2019, 2023). Recombination may also result in crossover. Lack of crossover results in a loss of fertility due to the production of aneuploid gametes with probability f (Úbeda et al., 2019, 2023). Coevolution between PRDM9 and its target results in cyclic (red-queen)

dynamics, where biased gene conversion drives the fixation of mismatching alleles at target loci, while fertility selection drives the fixation of matching alleles at the PRDM9 locus as long as the fertility coefficient is less than half the conversion coefficient, that is $f < \frac{1}{2}c$ (Úbeda et al., 2019, 2023) (see Figure 1A for a summary).

Although we model a single target, the insight gained applies to multiple targets. For simplicity, the coevolutionary model we build upon assumes the existence of two alleles in each locus and forward and backward mutations between the alleles in each locus (Úbeda et al., 2023) (Figure 1A). This simplification allows the derivation of analytical results and an in-depth understanding of the dynamics. Assuming forward and backward mutations (μ^A and μ^B denote the probability of mutation $A_1 \leftrightarrow A_2$ and $B_1 \leftrightarrow B_2$) results in cycles where the same combination of targeting and target alleles repeats, that is $A_1B_2 \rightarrow A_2B_2 \rightarrow A_2B_1 \rightarrow A_1B_1 \rightarrow A_1B_2 \rightarrow \dots$ (Úbeda et al., 2023) (Figure 1A.i). From a phenotypic perspective, the recurrence of targeting and target alleles translates into a recombination hotspot in the same genomic location, oscillating between hot and cold phases over evolutionary time (Úbeda et al., 2023) (Figure 1A.ii). In nature, however, the zinc finger domain in PRDM9—the domain that determines the matched sequence—consists of a sequence of 10 nucleotides with 4 nucleotides to choose from in each position (Baudat et al., 2010; Oliver et al., 2009). Although the number of possible combinations is finite, the probability of mutating to the same combination of nucleotides is smaller than the probability of mutating to a different one. In that sense, assuming forward mutations only would be more realistic. However, the model we use still applies to forward mutations only.

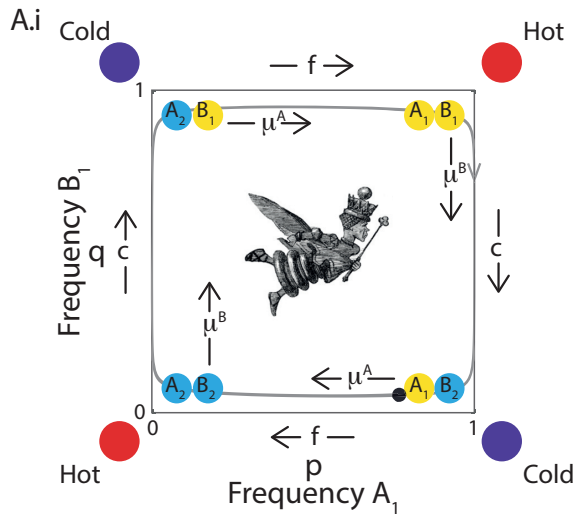
The model we use assumes that subscripts of alleles *A* and *B* represent specific nucleotide sequences; however, they are only labels, and the results of the model apply when replacing old labels with new labels. Assuming that mutations are no longer backward but only forward (μ_A and μ_B denote the probability of mutation $A_1 \rightarrow A_2 \rightarrow A_3 \dots$ and $B_1 \rightarrow B_2 \rightarrow B_3 \dots$) results in cycles where new combinations of targeting and target alleles succeed old ones, that is $A_1B_2 \rightarrow A_2B_2 \rightarrow A_2B_3 \rightarrow A_3B_3 \rightarrow A_3B_4 \rightarrow \dots$ (Figure 1B.i). From a phenotypic perspective, the nonrecurrence of targeting and target alleles translates into recombination hotspots in new genomic locations becoming hot as the ones in old locations become cold (Figure 1B.ii). What is relevant about these subindices is whether the sequences match or do not match. In a population of mismatching targeting and target sequences, fertility selection favours fixation of a matching targeting sequence. When the previous allele is fixed, biased gene conversion drives fixation of a mismatching target sequence. Coevolution results in an alternation between mismatching and matching haplotypes that is maintained over evolutionary time. This is true independently of the labels used to denote matching or mismatching alleles. We thus contend that the model we choose to build upon captures the key features of the coevolution between target and targeting loci without sacrificing realism in terms of nonrecurrence of sequences and location of hotspots. This conclusion has been validated by repeated extensions of the one target model to multiple targets (Úbeda & Wilkins, 2011; Úbeda et al., 2019, 2023).

The model we use for the interaction between PRDM9 and its target applies to multiple selection models. Recently, it has

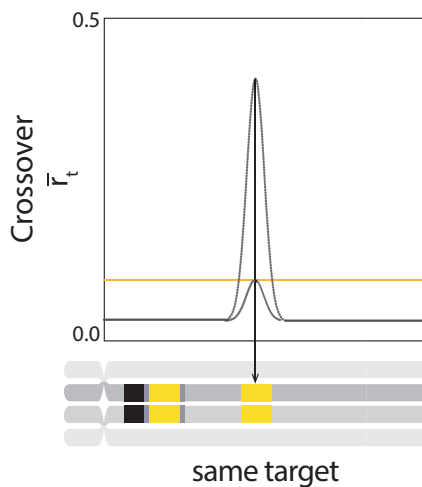
A. Forward and backward mutations

$$A_1 \xrightleftharpoons{\mu^A} A_2$$

$$B_1 \xrightleftharpoons{\mu^B} B_2$$



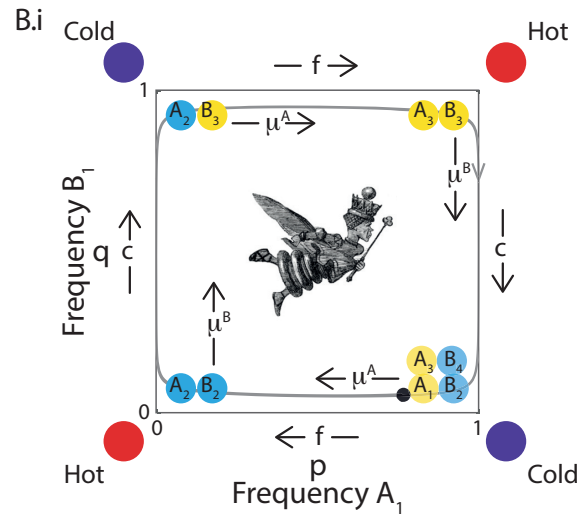
A.ii



B. Forward mutations

$$A_1 \xrightarrow{\mu^A} A_2 \xrightarrow{\mu^A} A_3$$

$$B_1 \xrightarrow{\mu^B} B_2 \xrightarrow{\mu^B} B_3$$



B.ii

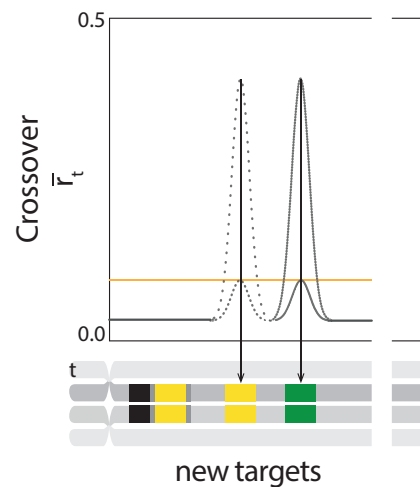


Figure 1. Summary of red-queen dynamics between PRDM9 (A) and target (B) loci resulting in the alternation of hot and cold states of one or multiple targets. Panel (i) shows the coevolutionary dynamics and illustrates the different forces driving each transition. The x-axis corresponds to the frequency of allele A_1 at PRDM9 locus p , while the y-axis corresponds to the frequency of allele B_1 at the target locus. When subscripts of alleles at PRDM9 and target loci match, there may be a crossover event at the target locus. We use the colours yellow and light blue to represent the two types of alleles in each locus. We use the colour red to represent the that a crossover may take place and corresponds to a hot target (recombination hotspots). We use the colour blue to represent that a crossover does not take place and corresponds to a cold target. If there is a crossover, there may be conversion of the target allele that matches the allele at PRDM9, into the target allele that does not match the allele at PRDM9 with probability c . If there is no crossover, there may be a loss in fertility f . Grey line shows an example of oscillatory dynamics in the state space (p, q) . Arrows along the x-axis show the direction in which fertility, f , and mutation in A drive the coevolutionary transition cold-hot. Arrows along the y-axis show the direction in which conversion c , and mutation in B drive the coevolutionary transition hot-cold. Panel (ii) shows the phenotype of the recombination hotspot corresponding to the dynamics above. Panel A shows the dynamics when backward mutations are allowed, resulting in the birth of hotspots always in the same target. Panel B shows the same dynamics when backward mutations are not allowed, resulting in the birth of hotspots always in new targets.

been argued that crossover—and thus fertility selection—is not determined by single binding events of PRDM9 proteins at a target locus (asymmetric binding), but rather by the double binding of proteins at homologous target loci (symmetric binding). The model we use assumes that one binding at a target site is enough for proper chromosomal segregation. The assumption that two bindings in homologous targets are needed does make the dynamics more complex. However, re-

cent computational models where fertility is determined by two bindings in homologous targets find oscillatory dynamics (Baker et al., 2023; Genestier et al., 2024). The exact model used for the coevolution between PRDM9 and its target sites is not key to this research. As long as the model shows red-queen dynamics between targeting and target sites, it will be a good background for the evolution of a modifier of the mutation rate at PRDM9 or its target sites.

The novelty of our model relies on adding a modifier locus M that determines the mutation rate at locus A or locus B . By adding this modifier locus, we allow the mutation rate at PRDM9 and its target loci to evolve, which is something that has not been explored previously.

Throughout this work, we assume an infinite, sexual, diploid population with discrete generations. Let variable $x_{i,k,m}$ denote the frequency of haplotype $M_i A_k B_m$ in the initial pool of gametes, with $\sum_{i,k,m} x_{i,k,m} = 1$. We model the change in frequency of each haplotype through a life cycle consisting of four steps:

- (1) **Random mating.** Zygotes, and the diploid individuals they develop into, come from the random fusion of haploid gametes in a pool produced by the previous adult generation.
- (2) **Conversion and selection.** The protein produced by allele k at the PRDM9 locus binds the corresponding allele m at the target locus, producing a DSB. A DSB may result in biased gene conversion of the sequence that breaks into the sequence that does not break with probability c . Finally, the DSB may result in crossover between flanking regions at the target site with probability r_t . The absence of DSB or crossover results in a loss of fertility due to the production of nonviable gametes with probability f (see online supplementary material Appendix 1 and Úbeda et al., 2023 for a detailed description of this model and its dynamics).
- (3) **Recombination.** For simplicity, we assume free recombination between PRDM9 and its target locus: A and B . We assume free recombination between modifier and target loci $r_{MB} = \frac{1}{2}$ because mutations of many target sequences within the genome will be determined by a modifier acting on the entire genome. We allow for any level of recombination between modifier and PRDM9 locus r_{MA} , that is $0 \leq r_{MA} \leq \frac{1}{2}$. This assumption allows us to consider two alternative genetic architectures: (i) when modifier and PRDM9 are not tightly linked $0 < r_{MA} \leq \frac{1}{2}$, the mutation rate at the sequence recognition domain in PRDM9 is determined by a locus somewhere else in the genome; (ii) when modifier and PRDM9 are tightly linked $r_{MA} \approx 0$, modifier and PRDM9 behave as alleles at a single locus, and the mutation rate is an intrinsic property of the sequence recognition domain of PRDM9. For example, consider a modifier that underpins the number of repeats of the sequence recognition domain of PRDM9. It is the repetitive structure of this domain that causes its high mutation rate. A modifier that is tightly linked to PRDM9 would be equivalent to alleles with a different number of repetitions in their sequence recognition domain.
- (4) **Mutations.** Alleles may experience mutations following the segregation of haplotypes into viable gametes. We assume that allele M_0 codes for baseline mutation rate μ^A when acting on locus A and μ^B when acting on B . Allele M_1 codes for a change of magnitude δ_μ^A in the baseline mutation rate at locus A , or a change δ_μ^B in the mutation rate at locus B . Notice that δ_μ can take positive or negative values, thus allowing for both increments and reductions of the baseline mutation rate

(see online supplementary material Appendix 1). This last step brings us back to the beginning of the life cycle modelled.

Let $x'_{i,k,m}$ denote the frequency of haplotype $M_i A_k B_m$ in the pool of gametes of the next generation. See online supplementary material Appendix 1 for the expression of the haplotype frequency in the next generation given parameters $f, c, r_t, \mu, \delta_\mu^A$, and δ_μ^B and the haplotype frequencies in the initial generation $x_{i,k,m}$. When needed, we use “dots” to represent combinations of haplotype frequencies, $x_{i,\cdot,\cdot} = \sum_k x_{i,k,m}$ in the current generation and $x'_{i,\cdot,\cdot} = \sum_k x'_{i,k,m}$ in the next generation.

Instead of using haplotype frequencies, we find more insightful to describe the system in terms of allele frequencies and linkage disequilibria. Let m, p , and q to denote the frequencies of alleles M_1, A_1 , and B_1 , respectively, that is

$$m = x_{1,\cdot,\cdot} \quad (1a)$$

$$p = x_{\cdot,1,\cdot} \quad (1b)$$

$$q = x_{\cdot,\cdot,1} \quad (1c)$$

We use variables D_{MA}, D_{AB}, D_{MB} , and D_{MAB} to denote the set of linkage disequilibria needed to fully describe the system, that is

$$D_{MA} = x_{0,1,\cdot} x_{1,2,\cdot} - x_{0,2,\cdot} x_{1,1,\cdot} \quad (2a)$$

$$D_{MB} = x_{0,\cdot,1} x_{1,\cdot,2} - x_{0,\cdot,2} x_{1,\cdot,1} \quad (2b)$$

$$D_{AB} = x_{\cdot,1,1} x_{\cdot,2,2} - x_{\cdot,1,2} x_{\cdot,2,1} \quad (2c)$$

$$D_{MAB} = x_{0,1,1} - (1-m)D_{AB} - pD_{MB} - qD_{MA} - (1-m)pq. \quad (2d)$$

Note that linkage disequilibria are defined such that D_{MA} is positive when M_1 is associated with A_2 , D_{MB} is positive when M_1 is associated with B_2 , D_{AB} is positive when A_1 is associated with B_1 , and finally, D_{MAB} is positive when M_1 is associated with $A_1 B_2$ and/or $A_2 B_1$.

The allele frequencies in the next generation are $m' = x'_{1,\cdot,\cdot}$, $p' = x'_{\cdot,1,\cdot}$, and $q' = x'_{\cdot,\cdot,1}$. The corresponding linkage disequilibria in the next generation are $D'_{MA} = x'_{0,1,\cdot} x'_{1,2,\cdot} - x'_{0,2,\cdot} x'_{1,1,\cdot}$, $D'_{MB} = x'_{0,\cdot,1} x'_{1,\cdot,2} - x'_{0,\cdot,2} x'_{1,\cdot,1}$, $D'_{AB} = x'_{\cdot,1,1} x'_{\cdot,2,2} - x'_{\cdot,1,2} x'_{\cdot,2,1}$, and $D'_{MAB} = x'_{0,1,1} - (1-m')D'_{AB} - p'D'_{MB} - q'D'_{MA} - (1-m')p'q'$.

Numerical analysis

We are interested in estimating the selection coefficient s acting on a mutant modifier allele M_1 .

Assume that the modifier in our model can be approximated as a simple 1-locus population genetics model with alleles M_0 and M_1 . In such a simplified model, we would assume that the fitness of genotypes $M_0 M_0$, $M_0 M_1$, and $M_1 M_1$ are 1, $1 + \frac{1}{2}s$, and $1 + s$, respectively. The frequency of allele M_1 in the next generation is here given by

$$m_{t+1} = m_t \frac{1 + \frac{1}{2}s(1 + m_t)}{1 + s m_t}, \quad (3)$$

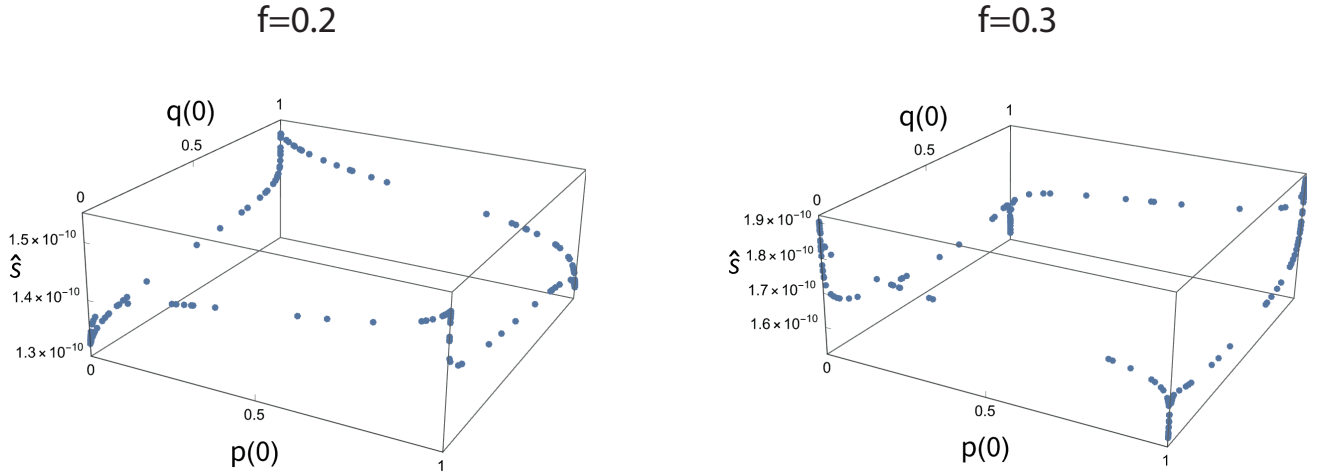


Figure 2. Selection coefficients for different initial conditions along the limit cycle. In each panel, the x and y-axis represent the initial conditions for p and q , that is $[p(0), q(0)]$. The z-axis represents our estimate of the selection coefficient, $\hat{s}$, acting on modifiers of mutation rates at PRDM9 for each initial conditions along the cycle. We present two examples, one corresponding to $f = 0.2$ and another for $f = 0.3$. In all cases, we assume all linkage disequilibria to be initially equal to zero. Other parameter values are $b = c = 1$; free recombination between PRDM9 and its target, $r_{AB} = \frac{1}{2}$; initial mutation rate at PRDM9 and target sites, $\mu_A = \mu_B = 10^{-12}$; and the modifier's effect, $\delta\mu^A = 10^{-8}$.

and its change in frequency from one generation to the next $\Delta m = m_{t+1} - m_t$ is

$$\Delta m = \frac{\frac{1}{2}sm_t(1 - m_t)}{1 + sm_t}. \quad (4)$$

Solving for the selection coefficient, we obtain

$$s = \frac{2\Delta m}{m_t(1 - m_t) - 2m_t\Delta m}. \quad (5)$$

To estimate s , we will now assume that our modifier behaves similarly as the locus in the simplified model, whose basic population genetics equations have just been given.

In our modifier model, we carry out deterministic forward simulations. Given a set of parameter values, we iterate our recursion equations, obtaining the frequency m in 50,000 generations. To estimate the selection coefficient s applying to our modifier, we calculate Equation 5 at every generation, excluding the first 10,000 generations to avoid any transient effect caused by our choice of initial conditions. The selection coefficients obtained alternates between positive and negative values due to the oscillatory nature of the system considered. We calculate the average of all selection coefficients after the burn-in phase to eliminate the short-term effect of the oscillatory dynamics, leaving the long-term trend. The estimate of the selection coefficient we use is

$$\hat{s} = \frac{1}{40,000} \sum_{t=10,000}^{t=50,000} \frac{2(m_{t+1} - m_t)}{m_t(1 - m_t) - 2m_t(m_{t+1} - m_t)}. \quad (6)$$

A positive selection coefficient means that, in the long term, modifier M_1 replaces M_0 , while negative selection coefficients mean that, in the long term, modifier M_1 does not replace M_0 .

Following the burn-in phase, PRDM9 and its target locus should be oscillating between cold and hot phases, as this is the behaviour corresponding to empirical observations (Úbeda et al., 2023). For that reason, we choose parameter values where limit cycles are observed (which requires the fertility cost to be lower than half the conversion probability $f < \frac{1}{2}c$) and initial conditions that are in the basin of attraction of that limit cycle (which requires a high initial frequency of any pair

of alleles in A and B). When initial conditions are not in that basin of attraction, we have shown elsewhere that the system converges to a stable equilibrium point not compatible with empirical observations (Úbeda et al., 2023). Regarding the choice of parameter values in all our simulations, we assume that binding probability is $b = 1$ and conversion probability is $c = 1$. Both parameters are scaling factors: Choosing lower values would only slow down the cycle without altering any qualitative results. We assume $f < \frac{1}{2}c$, as we have also shown, elsewhere that if it is not the case, the system converges to either one of two stable equilibrium points not compatible with empirical observations (Úbeda et al., 2019). For simplicity, we assume that PRDM9 and its target recombine freely, $r_{AB} = \frac{1}{2}$, and that double strand breaks lead to crossover events in all cases $r_t = 1$. Regarding the choice of initial conditions, in all our simulations, we assume that initially the mutant modifier is rare, in particular $m(0) = 0.01$, the frequency of any pair of alleles segregating at PRDM9, and its target are high, in particular $p(0), q(0) = 0.05$ and $D_{AB}(0) = -0.045$ (to favour convergence to the limit cycle), there is no linkage disequilibrium between locus M any other loci, that is $D_{MB}(0) = D_{MA}(0) = D_{MAB}(0) = 0$. This applies to all results and figures presented in the main text unless stated otherwise.

Notice that selection coefficients may depend on the initial conditions regarding the pair of alleles segregating at locus A and B , that is $p(0)$ and $q(0)$. These initial conditions correspond to points in the limit cycle underlying the oscillation between cold and hot states (Úbeda et al., 2023). For simplicity, in our explorations, we chose a very specific initial condition, $p(0), q(0) = 0.05$ with $D_{AB} = -0.045$, corresponding to a point in the hot phase of the limit cycle. However, this choice does not condition our results. We have explored how changes in initial conditions along the limit cycle may affect our estimate of selection coefficients for selected values of f . We find that the change in selection coefficients due to variation in initial conditions is very small and, most importantly, that initial conditions do not change the sign of the selection coefficient. In Figure 2, we provide two examples of how initial conditions may affect the estimate of selection

coefficients along the limit cycle. When $f = 0.2$, the estimate of the selection coefficient ranges between $[1.32 \times 10^{-10}, 1.52 \times 10^{-10}]$ (Figure 2). Similarly, when $f = 0.3$, the estimate of the selection coefficient ranges between $[1.60 \times 10^{-10}, 1.90 \times 10^{-10}]$ (Figure 2).

Frequency change analysis

We explore the change in frequency of the mutant allele from one generation to the next. The equation describing this change can be obtained from $\Delta m = m' - m = \sum_{k,m} x'_{1,k,m} - \sum_{k,m} x_{1,k,m}$, and its expression is

$$\bar{w}\Delta m = f \left[D_{MA} \left(\frac{1}{2} - q \right) + D_{MB} \left(\frac{1}{2} - p \right) - \frac{1}{2} D_{MAB} \right], \quad (7)$$

where

$$\bar{w} = 1 - f [1 - (1 - p(1 - q) - (1 - p)q + D_{AB})]. \quad (8)$$

Equation 7 shows that the change in frequency of the mutant allele is directed by its genetic associations with PRDM9 and target alleles. Terms D_{MA} , D_{MB} , and D_{MAB} are not equally strong. D_{MA} and D_{MB} alternate in terms of which one is stronger. Each term dominates in different parts of the limit cycle that describes oscillations between p and q . D_{MAB} is much smaller than the other two in many parts of the cycles, but it dominates when the other two terms cancel each other out. In online supplementary material Appendix 2, we provide analytical approximations for D_{MA} , D_{MB} , and D_{MAB} assuming weak selection, weak mutation, and quasi-linkage equilibrium. This allows us to see that the linkage disequilibria between modifier and the other two loci are determined by three processes, namely conversion, selection, and mutation. However, from these expressions, it is difficult to obtain a formal conclusion on whether a given modifier is selected for or against. Part of the problem resides in disentangling the noise introduced by the oscillatory, short-term behaviour of m from its long-term behaviour. The analytical expressions obtained here allow us to gain an understanding of the forces at play, while our numerical simulations provide us with a quantitative overall result of the selection coefficient applying to any given modifier.

Population mean fitness

Here we investigate the effect of changes in mutation rate at PRDM9 relative to its target site, μ^A/μ^B , on population mean fitness, $\bar{w}$. We are interested in determining whether increments in mutation rates in A relative to B increase or decrease population mean fitness. For that purpose, we calculate the population mean fitness in the absence of a modifier allele changing the baseline mutation rate, that is $m(0) = 0$. Notice that in the absence of M_1 , δ_μ^A , and δ_μ^B are not present in our equations, and there is a single mutation rate at each locus μ^A and μ^B .

We carry out deterministic forward simulations for values of the fertility parameter f comprised between 0 and 0.5. For each value of f , we consider the pair of mutation rates 10^{-12} at A and 10^{-10} at B , that is $\mu^A = 10^{-12}$ and $\mu^B = 10^{-10}$. We let the system evolve for 10,000 generations, calculating the average population mean fitness across the last 7,000 generations. We exclude the first 3,000 generations to avoid any temporary effect of initial conditions. Notice that we

need to calculate the average across generations due to the oscillatory nature of the coevolutionary dynamics between PRDM9 and its target. We repeat the same procedure, just changing the mutation rate at A to 10^{-8} . For each value of f , we calculate the difference in average population mean fitness between these two values of the mutation rate at A , that is $\bar{w}(\mu^A = 10^{-8}) - \bar{w}(\mu^A = 10^{-12})$.

Results

Selection for higher mutation rates at PRDM9 and lower mutation rates at target sites

Numerical results (Figure 3) show that when there is free recombination between modifier and PRDM9 ($r_{MA} = \frac{1}{2}$), a rare modifier increasing the mutation rate at PRDM9 is selected for ($\hat{s} > 0$) when it increases mutation rates ($\delta_\mu^A > 0$). On the contrary, it is selected against ($\hat{s} < 0$) when it decreases mutation rates ($\delta_\mu^A < 0$). While modelling free recombination is the most general approach, it could be argued that in nature modifiers of mutation rates at PRDM9 are most likely to be linked to PRDM9. This is because the binding domain of PRDM9 is formed by a zinc finger, which exhibits a high recombination rate on account of its repetitive molecular structure: The very same nucleotide sequence can be seen as a modifier increasing the mutation rate of itself. For this reason, we also explore cases with partial or no recombination between the modifier and PRDM9 locus.

When there is partial linkage between modifier and PRDM9 ($r_{MA} = \frac{1}{4}$), a rare modifier of the mutation rate at PRDM9 is selected for ($\hat{s} > 0$) when it increases mutation rates ($\delta_\mu^A > 0$). On the contrary, it is selected against ($\hat{s} < 0$) when it decreases mutation rates ($\delta_\mu^A < 0$) (Figure 4). When recombination between modifier and PRDM9 is $r_{MA} = \frac{1}{4}$, the selection coefficient is almost one order of magnitude greater than the selection coefficient when there is free recombination (Figure 4). When there is no recombination between modifier and PRDM9, $r_{MA} = 0$, a rare modifier of the mutation rate at PRDM9 is quickly selected for ($\hat{s} > 0$) when it increases mutation rates ($\delta_\mu^A > 0$), and it is selected against ($\hat{s} < 0$) when it decreases mutation rates ($\delta_\mu^A < 0$) (Figure 4). When there is no recombination between modifier and PRDM9, the selection coefficient is almost seven orders of magnitude greater than the selection coefficient when there is free recombination (Figure 4).

In general, selection favours both unlinked and linked rare modifiers that increase mutation rates at PRDM9. The greater the linkage between modifier and PRDM9, the stronger the selection favouring or disfavours modifiers (Figures 3 and 4). The greater the change in mutation rates underpinned by modifiers, the stronger the selection acting on modifiers (Figures 3 and 4). The greater the fertility costs of no crossover, the stronger the selection acting on modifiers that increase mutation rates (Figures 3 and 4). Notice, however, that these increments in selection with fertility costs are not monotonous (Figures 3 and 4).

Finally, numerical results (Figure 3) show that when there is free recombination between modifier and target loci ($r_{MB} = \frac{1}{2}$), rare modifiers of the mutation rates at target loci are selected for ($\hat{s} > 0$) when they decrease mutation rates ($\delta_\mu^B < 0$). However, they are selected against ($\hat{s} < 0$) when they increase mutation rates ($\delta_\mu^B > 0$). In this case, there is no reason to argue that modifiers of mutation rates and the target

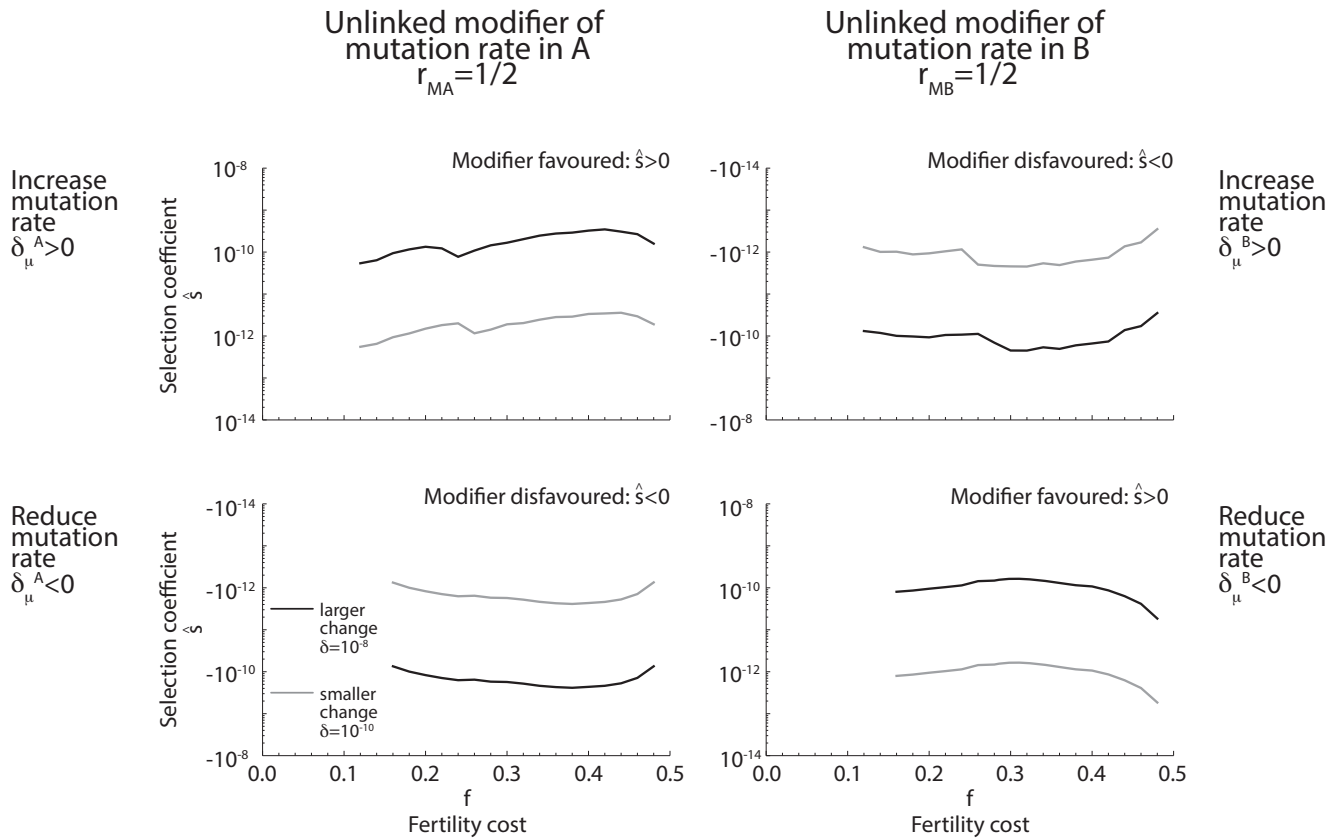


Figure 3. Selection on unlinked modifiers of the mutation rate at PRDM9 and target loci. Within each panel, the x-axis corresponds to the fertility cost of no recombination due to lack of crossovers f . The range of values explored $0 < f < \frac{1}{2}c$ corresponds to the case when recombination hotspots die due to gene conversion and are born due to fertility selection (red-queen dynamics). The y-axis corresponds to an estimate of the selection coefficient s acting on a rare unlinked ($r_{MA} = r_{MB} = \frac{1}{2}$) modifier of mutation rate (see the “Materials and methods” section for its calculation). This modifier is favoured if $s > 0$ and disfavoured if $s < 0$. Columns corresponds to unlinked modifiers acting on PRDM9 (first column) and target (second column) loci. Rows correspond to modifiers increasing (first row) or reducing (second row) the mutation rate. Natural selection favours unlinked modifiers increasing the mutation rate at PRDM9 but decreasing the mutation rate at target loci. Initial conditions are a rare modifier $m = 0.99$, a target in hot state $p = q = 0.05$, no linkage disequilibrium at the modifier locus $D_{MA} = D_{MB} = D_{MAB} = 0$, and small linkage between PRDM9 and its target $D_{AB} = -0.045$. Other parameter values are $b = c = 1$; free recombination between PRDM9 and its target, $r_{AB} = \frac{1}{2}$; and initial mutation rate at PRDM9 and target sites, $\mu_A = \mu_B = 10^{-12}$.

locus may be linked, as there are potentially tens of thousands of targets distributed in a genome. Because the modifier cannot physically be linked to all at the same time, we only study here the general case, relevant for the vast majority of targets, of an unlinked modifier.

Selection on the modifier of mutations is driven by its association with favourable genetic backgrounds

In this section we provide some insight into why natural selection favours increments in mutation probabilities at PRDM9 locus but reductions at target loci. To do so, we look at expression (7) for the change in frequency of a rare modifier from one generation to the next, Δm .

The sign of expression 7 depends on the sign of linkage disequilibria between the modifier and all other loci, D_{MA} , D_{MB} , D_{MAB} , and on which combination of alleles is segregating most often in locus A and B. For example, if alleles M_1 and A_1 are found together more often than expected at random, that is $D_{MA} < 0$, there is positive selection for allele M_1 when allele B_1 is segregating in the population more often than B_2 , that is $\frac{1}{2} - q < 0$. In other words, M_1 is more often associated with alleles A_1 and B_1 . Similarly, if M_1 and B_1 are found together more often, that is $D_{MB} < 0$, M_1 is selected

only when A_1 is segregating more often in locus A, that is $\frac{1}{2} - p < 0$. In other words, M_1 is more often associated with alleles A_1 and B_1 . Finally, M_1 is also favoured when found more often with haplotype A_1B_1 , $D_{MAB} < 0$. Adding up these effects, we get

$$\Delta m = \frac{f}{w} \left[\underbrace{D_{MA} \left(\frac{1}{2} - q \right)}_{-} + \underbrace{D_{MB} \left(\frac{1}{2} - p \right)}_{-} - \frac{1}{2} \underbrace{D_{MAB}}_{-} \right] > 0. \quad (9)$$

The same logic can be applied to positive linkage between the modifier and all other loci and the pair of alleles A_2 and B_2 . In general, we can conclude that mutant modifiers are selected when preferentially associated with combinations of matching alleles at A and B loci, haplotypes A_1B_1 and A_2B_2 . These *hot* combinations of matching alleles result in greater fertility for the individual than *cold* combinations of mismatching alleles, haplotypes A_1B_2 and A_2B_1 . We conclude that rare modifiers of mutation rates are thus selected not because mutations have any intrinsic benefit but because they are preferentially associated with fitter genetic backgrounds corresponding to matching haplotypes.

Figure 1 provides an intuitive explanation of why a modifier that increases mutation rates at locus A relative to locus B tends to become associated with *hot* allelic combinations. The

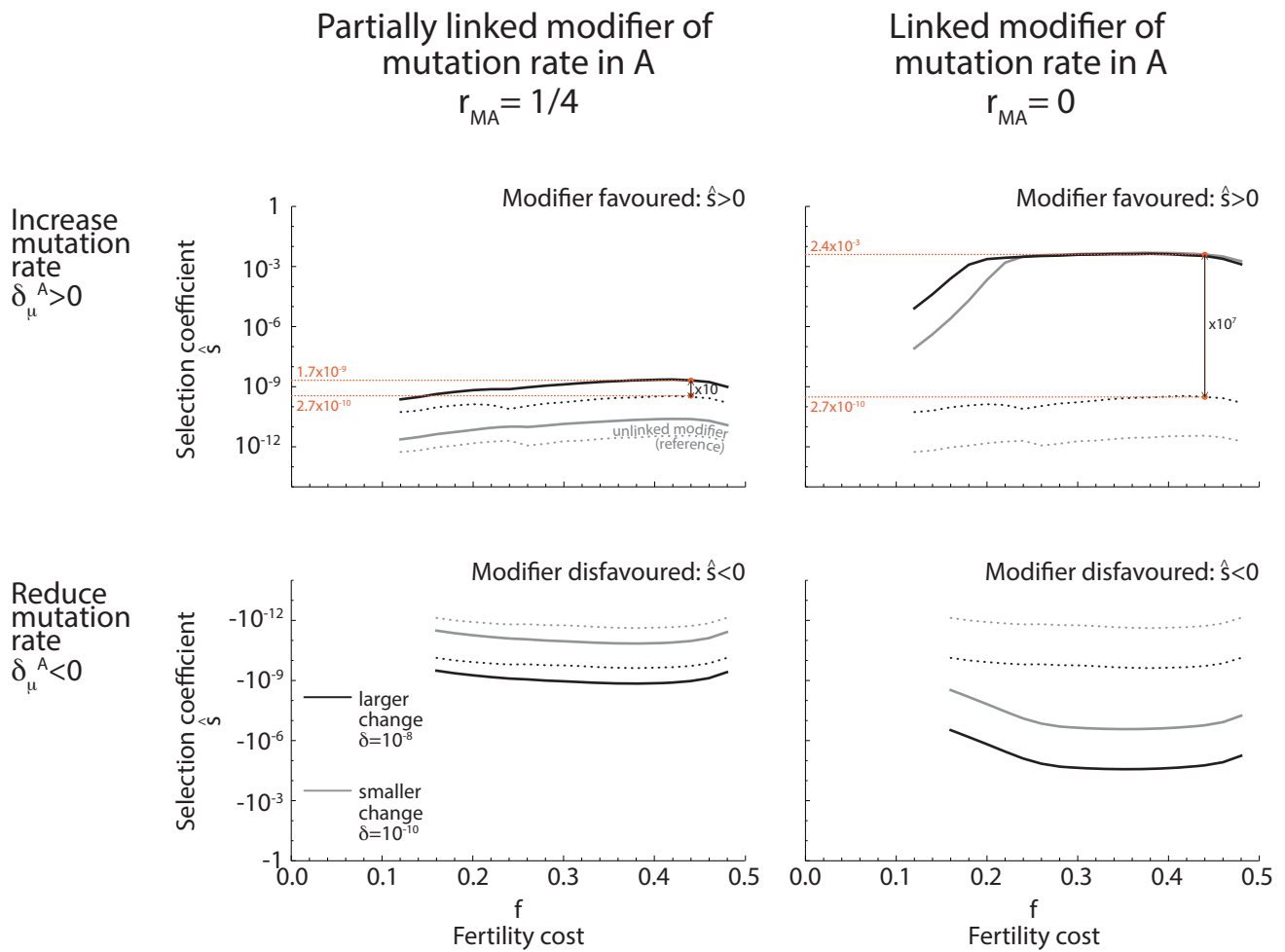


Figure 4. Selection on partially linked modifiers of the mutation rate at PRDM9. Within each panel, the x-axis corresponds to the fertility cost of no recombination due to lack of crossovers f . The range of parameter values explored $0 < f < \frac{1}{2}c$ corresponds to the case when recombination hotspots die due to gene conversion and are born due to fertility selection (red-queen dynamics). The y-axis corresponds to an estimate of the selection coefficient s acting on a rare partially linked ($0 < r_{MA} < \frac{1}{2}$) modifier of mutation rate (see the “Materials and methods” section for its calculation). This modifier is favoured if $s > 0$ and disfavoured if $s < 0$. Columns correspond to modifiers acting on PRDM9 with recombination rates $r_{MA} = \frac{1}{4}$ (first column) and $r_{MA} = 0$ (second column). Rows correspond to modifiers increasing (first row) or reducing (second row) the mutation rate at PRDM9. Dashed lines correspond to values when the modifier is unlinked and are provided as a reference. The lower the recombination rate between modifier and PRDM9 loci, the stronger the selection acting on the modifier. Halving the recombination rate increments the selection acting on a modifier of the mutation rate at PRDM9 roughly an order of magnitude. Setting the recombination rate to 0 increments the selection acting on a modifier of the mutation rate at PRDM9 roughly seven orders of magnitude. Initial conditions are $m = 0.99$, $p = q = 0.05$, $D_{MA} = D_{MB} = D_{MAB} = 0$ and $D_{AB} = -0.045$. Other parameter values are $b = c = 1$, $r_{AB} = \frac{1}{2}$, $\mu_A = \mu_B = 10^{-12}$.

cycle that characterizes the evolutionary treadmill created by conversion and selection is represented clockwise: Haplotype A_1B_1 is converted into A_1B_2 . The reduction of crossing-over frequency creates a selection for A_2B_2 (as this haplotype restores crossing-over and ensures proper meiotic division). In turn, it gets converted into A_2B_1 , and again from there selection favours haplotype A_1B_1 (Figure 1).

Mathematical analysis detailed in Appendix 2 in supplementary material shows that a higher mutation rate at A helps the transitions $A_1B_2 \rightarrow A_2B_2$ and $A_2B_1 \rightarrow A_1B_1$. Such transitions are mediated by selection f ; mutations here help create the standing variation (A_2B_2 and A_1B_1 haplotypes, respectively) necessary for selection to act and increase the frequency of *hot* allelic combinations (Figure 1). Higher mutation rates at A also act against the transitions $A_1B_1 \rightarrow A_1B_2$ and $A_2B_2 \rightarrow A_2B_1$: When the PRMD9 allele gets mutated,

then its corresponding target allele gets converted less often. On the other hand, higher mutation rate at B slows down the transitions $A_1B_2 \rightarrow A_2B_2$ and $A_2B_1 \rightarrow A_1B_1$, as it removes the necessary standing variation (Figure 1). It also accelerates $A_1B_1 \rightarrow A_1B_2$ and $A_2B_2 \rightarrow A_2B_1$ transitions by acting in the same direction as conversion.

Overall, we see that mutations on A accelerate transitions from *cold* to *hot* haplotypes, while mutations on B accelerate transitions from *hot* to *cold* haplotypes (Figure 1). Therefore, a modifier that increases mutations on A relative to B will, over a full cycle, create an excess of *hot* allelic combinations relative to the ancestral allele. Such a mutant modifier thus will tend to be more often associated with *hot*, fitter genetic background than its ancestral counterpart—and will be indirectly selected for as described above. Reversely, a mutant modifier that increases mutation rates on B relative to A will

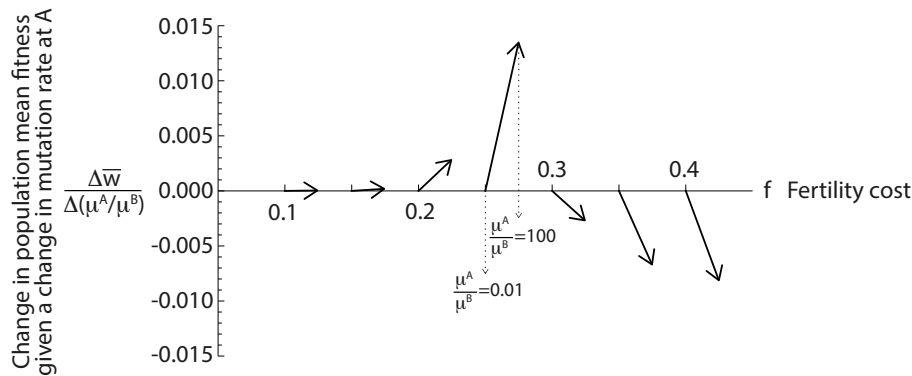


Figure 5. Change in population mean fitness caused by a change in the mutation rate at PRDM9 relative to its target. The x-axis corresponds to fertility costs f . The range of values explored $0 < f < \frac{1}{2}c$ corresponds to the case when recombination hotspots die due to gene conversion and are born due to fertility selection (red-queen dynamics). The y-axis corresponds to the change in population mean fitness $\Delta\bar{w}$ due to a change in mutation rate at PRDM9 relative to its target $\delta\frac{\mu^A}{\mu^B}$. In particular, the change in mutation rate considered is 10^4 from $\frac{\mu^A}{\mu^B} = 10^{-2}$ to $\frac{\mu^A}{\mu^B} = 10^2$ in all cases. Population mean fitness is calculated as the average value across 7,000 generations. This is necessary due to the oscillatory nature of the dynamics. The change in population mean fitness results from the difference between $\bar{w}$ when $\mu^A = 10^{-12}$ and $\mu^B = 10^{-10}$ and $\bar{w}$ when $\mu^A = 10^{-8}$ and $\mu^B = 10^{-10}$. It can be observed that when $f < \frac{1}{4}c$, increments in mutation rates at A result in increments in the population mean fitness. However, when $\frac{1}{4}c < f < \frac{1}{2}c$ increments in the mutation rate at A result in reductions in the population mean fitness. Initial conditions are $p = q = 0.9999$, $D_{MA} = D_{MB} = D_{MAB} = D_{AB} = 0$. Other parameter values are $b = c = 1$, $r_{MA} = r_{AB} = \frac{1}{2}$.

become, on average, more often associated with *cold*, less fit allelic combinations than its ancestral counterpart, resulting in selection acting against it.

Greater mutation rate at PRDM9 does not necessarily increase the population mean fitness

In Figure 5, we present the change in population mean fitness as a function of the change in mutation rate at the PRDM9 locus relative to the mutation rate at the target locus. We find that when f is smaller than $\frac{1}{4}c$, higher mutation rate at PRDM9 relative to its target results in increments in the population mean fitness (Figure 5). We find that when f is greater than $\frac{1}{4}c$, higher mutation rate at PRDM9 relative to its target results in reductions in the population mean fitness (Figure 5).

The evolution of higher mutation rates at PRDM9 relative to its target does not always increase the population mean fitness. Naively, it would be expected that unlinked modifiers are selected to increase the population mean fitness. Following this logic, we would expect higher mutation probabilities at PRDM9 to evolve when $f < \frac{1}{4}c$, and higher mutation probabilities at the target to evolve at higher values of $\frac{1}{4}c < f < \frac{1}{2}c$. This stands in contrast with our findings in the previous section where we showed that higher mutation rates at PRDM9 evolve for any value of f .

Discussion

One of the most remarkable features of gene PRDM9 is its unusually high mutation rate, a rate that makes PRDM9 one of the fastest-evolving genes in the human genome (Baker et al., 2017; Ponting, 2011). The estimate of the mutation rate in PRDM9 in humans is roughly 10^{-5} per gamete per generation (Jeffreys et al., 2013), while the mutation rate in target sites is roughly 10^{-7} per gamete per generation—to reflect the observed mutation rate per base pair and the 10 base pairs that determine a target site in humans (Kong et al., 2012). What explains that PRDM9 has a mutation rate roughly two orders of magnitude higher than its target site?

To explore the evolution of mutation rates at PRDM9 and target sites, we formulate a population genetics model for a modifier of mutation rates at either PRDM9 or its target. We find that natural selection favours the evolution of higher mutation rates at PRDM9 and lower mutation rates at PRDM9's target sites. We find that selection for higher mutation rates at PRDM9 is independent of the level of recombination between modifier and PRDM9—notice, however, that the strength of selection is not independent of the level of recombination. Both unlinked and linked modifiers of mutation rates at PRDM9 are favoured when increasing mutation rates. Whether the high mutation rate observed at PRDM9 is the result of an unlinked gene underpinning mutation rates at the sequence recognition domain of PRDM9 or a tightly linked gene acting as alleles segregating in PRDM9, natural selection favours increments in mutation rates at PRDM9. We show that modifiers of mutation rates at PRDM9 and its target spread in the population when they get associated with genetic backgrounds that favour crossover events (hot backgrounds). Mathematical analysis shows that higher mutation rates at PRDM9 accelerate transitions from cold to hot allelic combinations, while lower mutation rates at target sites accelerate transitions from hot to cold (Figure 1). An intuitive explanation of our findings is that natural selection favours an increment in mutation rates taking targets towards hot states—higher mutation rates in PRDM9—and a reduction in mutation rates taking targets away from hot states—lower mutation rates at target sites (Figure 1). Interestingly, we observe that higher mutation rates at PRDM9 do not always mean greater population mean fitness. Increments in mutation rates at PRDM9 increase population mean fitness for lower values of fertility selection, $0 < f < \frac{1}{4}c$, but reduce population mean fitness for higher values of fertility selection, $\frac{1}{4}c < f < \frac{1}{2}c$ (Figure 5).

Our work finds that the strength of selection acting on modifiers of the mutation rate at PRDM9 and target loci is variable depending on the effect of the modifier on mutation rates, the strength of fertility selection, and the linkage between

modifier and modified locus. In general, the greater the effect of the modifier on mutation rates, the greater the strength of selection acting on modifiers (Figures 3 and 4). The greater the strength of fertility selection on crossover, the greater the strength of selection acting on modifiers (Figures 3 and 4). Modifiers of mutation rates at groups of ten base pairs within the genome—target loci—are likely to be determined by modifiers freely recombining with them—that is unlinked modifiers. We show that selection acting on these modifiers is extremely small (Figure 3). Modifiers of mutation rates at PRDM9's sequence recognition domain are likely to be tightly linked. We show that selection acting on modifiers determining the mutation rate at PRDM9 is proportional to the linkage between modifier and PRDM9 (Figures 3 and 4). The greater the linkage, the greater the strength of selection acting on these modifiers (Figures 3 and 4). The maximum strength of selection corresponds to a modifier of mutation rate at PRDM9 perfectly linked to it (Figures 3 and 4).

Another source of asymmetry in the strength of selection acting on modifiers of mutation at PRDM9 and target loci is that PRDM9 can build a genetic associations (linkage disequilibrium) with many target loci, but one target locus can build genetic association with only one PRDM9 locus. For simplicity, we model one target locus only; in reality, there are tens of thousands of targets (Arnheim et al., 2007; Baudat et al., 2013; Myers et al., 2005; Paigen & Petkov, 2010). A rare modifier increasing mutation rates at PRDM9 will be selectively advantageous due to its linkage disequilibria with each and every one of the tens of thousands of target loci PRDM9 interacts with. This is a selection force of significant strength for a gene, PRDM9, that appeared in the population hundreds of millions of years ago—before the rise and diversification of vertebrates (Cavassim et al., 2022). However, a rare modifier decreasing mutation rates at a target site will be selectively advantageous due to its linkage disequilibrium with that target locus only. This is a weak selection force. Putting together this and the previous considerations, we can conclude that the only evolutionary process able to overcome random genetic drift (roughly $s > 1/N_e = 10^{-5}$ in humans) is selection for higher mutation rates at PRDM9. The most likely outcome of the evolution of mutation rates in PRDM9 and target loci is higher mutation rates at PRDM9 determined by linked modifiers. Here, selection coefficients are well above the drift barrier $s > 10^{-5}$ even when a single target interaction is considered. Interestingly, this is what is observed in nature, with the mutation rate of PRDM9 being two orders of magnitude higher than that at the target sites (and the average for the genome). The sequence recognition domain in PRDM9 corresponds to a zinc finger structure that is characterized by the repetition of sequences. It is the repetitive structure of the zinc finger that causes the high mutation rate of PRDM9's sequence recognition domain. One potential modifier of the mutation rate would be a modifier changing the number of repetitions in an ancestral nonrepetitive sequence recognition domain. When this modifier is tightly linked, it would really be acting as alleles with different numbers of repetitions segregating in PRDM9. Enough repetitions might be a necessary condition for the evolution of a zinc-finger structure at the sequence recognition domain of PRDM9.

Our findings are a rare case in the field of the evolution of mutation rates, although they are not unique. Substantial work on the evolution of mutation rates shows that selection favours a reduction in mutation rates because mutations

are mostly deleterious (Altenberg et al., 2017; Feldman & Liberman, 1986; Kimura, 1967; Liberman & Feldman, 1986). The ultimate limit to this reduction in mutation rate is the power of random genetic drift (Lynch et al., 2016). In our work, we find an exception to this reduction in mutation principle. We find that natural selection does favour increments in mutation rates at PRDM9. Another exception to the reduction in mutation principle is when fluctuating environments impact fitness (Liberman et al., 2011). In fluctuating environments, natural selection favours increments in mutation rates (Liberman et al., 2011). We think it is possible to draw some parallels between the exception we find in mutation rates at PRDM9 and the exception of mutation rates at a gene in fluctuating environments. Because the coevolution between PRDM9 and its targets results in oscillations between hot and cold states, selection on a modifier of mutation rates at PRDM9 fluctuates with the temperature at the target. In cold states, mutations at PRDM9 are favoured because they speed up the transition to a hot state, thus increasing fitness (Úbeda et al., 2023). In hot states, however, mutations at PRDM9 do not favour any transition but increase the linkage between nonmatching alleles, thus reducing fitness (Úbeda et al., 2023). It is the red-queen dynamics that generates a form of fluctuating environment in which mutations at PRDM9 are favoured.

Overall, here we uncovered a selective force acting on modifiers of mutation rate at PRDM9 and target sites indirectly through the association of the modifier with the target site. This selective force favours the increment of mutation rates at PRDM9 but not its target sites. Our predictions match well the observation that mutation rates at PRDM9 are two orders of magnitude higher than the average mutation rate in the genome (Jeffreys et al., 2013). Our work provides a potential evolutionary explanation for the abnormally high mutation rate observed in the binding domain of PRDM9, mutation rate that makes PRDM9 one of the fastest-evolving genes in the human genome (Baker et al., 2017; Ponting, 2011). The evolution of a tightly linked modifier of mutation rates at PRDM9 would help to explain the evolution of a zinc-finger structure—structure that is highly mutable on account of its repetitive sequence structure (Ellegren, 2004)—to code for the binding domain of PRDM9.

Supplementary material

Supplementary material is available at *Journal of Evolutionary Biology* online.

Data availability

The data underlying this article are available in the article and in its online supplementary material.

Author contributions

Frédéric Fyon (Conceptualization [equal], Formal analysis [equal], Investigation [equal], Methodology [equal], Validation [equal], Writing—original draft [lead], Writing—review & editing [equal]), and Francisco Úbeda (Conceptualization [equal], Formal analysis [equal], Funding acquisition [lead], Investigation [equal], Methodology [equal], Project administration [lead], Supervision [lead], Validation [equal],

Visualization [lead], Writing—original draft [supporting], Writing—review & editing [equal])

Funding

This work was supported by NSFDEB-NERC Research Grant NE/T009322/1. For the purpose of open access, the author has applied a Creative Commons Attribution (CC BY) licence to any Author Accepted Manuscript version arising.

Conflicts of interest

None declared.

References

- Altenberg, L., Liberman, U., & Feldman, M. W. (2017). Unified reduction principle for the evolution of mutation, migration, and recombination. *Proceedings of the National Academy of Sciences of the United States of America*, 114(12), E2392–E2400.
- Arnheim, N., Calabrese, P., & Tiemann-Boege, I. (2007). Mammalian meiotic recombination hot spots. *Annual Review of Genetics*, 41, 369–399. <https://doi.org/10.1146/annurev.genet.41.110306.130301>
- Bachtrog, D. (2013). Y-chromosome evolution: Emerging insights into processes of Y-chromosome degeneration. *Nature Reviews Genetics*, 14(2), 113–124. <https://doi.org/10.1038/nrg3366>
- Baker, Z., Przeworski, M., & Sella, G. (2023). Down the Penrose stairs, or how selection for fewer recombination hotspots maintains their existence. *eLife*, 12, e83769. <https://doi.org/10.7554/eLife.83769>
- Baker, Z., Schumer, M., Haba, Y., & Przeworski, M. (2017). Repeated losses of PRDM9-directed recombination despite the conservation of PRDM9 across vertebrates. *eLife*, 6, e24133. <https://doi.org/10.7554/eLife.24133>
- Barton, N. H., & Charlesworth, B. (1998). Why sex and recombination? *Science*, 281(5385), 1986. <https://doi.org/10.1126/science.281.5385.1986>
- Baudat, F., Buard, J., Grey, C., & De Massy, B. (2010). PRDM9 is a major determinant of meiotic recombination hotspots in humans and mice. *Science*, 327(5967), 836–840. <https://doi.org/10.1126/science.1183439>
- Baudat, F., Imai, Y., & De Massy, B. (2013). Meiotic recombination in mammals: Localization and regulation. *Nature Reviews Genetics*, 14(11), 794. <https://doi.org/10.1038/nrg3573>
- Boulton, A., Myers, R. S., & Redfield, R. J. (1997). The hotspot conversion paradox and the evolution of meiotic recombination. *Proceedings of the National Academy of Sciences*, 94(15), 8058–8063. <https://doi.org/10.1073/pnas.94.15.8058>
- Burt, A. (2000). Perspective: Sex, recombination and the efficacy of selection—was Weismann right? *Evolution*, 54(2), 337–351.
- Cavassim, M. I. A., Baker, Z., Hoge, C., & Przeworski, M. (2022). PRDM9 losses in vertebrates are coupled to those of paralogs ZCWPW1 and ZCWPW2. *Proceedings of the National Academy of Sciences of the United States of America*, 119(9), e2114401119. <https://doi.org/10.1073/pnas.2114401119>
- Coop, G., Wen, X., Ober, C., & Przeworski, M. (2008). High-resolution mapping of crossovers reveals extensive variation in fine-scale recombination patterns among humans. *Science*, 319(5868), 1395–1398. <https://doi.org/10.1126/science.1151851>
- Ellegren, H. (2004). Microsatellites: Simple sequences with complex evolution. *Nature Reviews Genetics*, 5(5), 435–445. <https://doi.org/10.1038/nrg1348>
- Feldman, M., & Liberman, U. (1986). An evolutionary reduction principle for genetic modifiers. *Proceedings of the National Academy of Sciences of the United States of America*, 83(13), 4824–4827. <https://doi.org/10.1073/pnas.83.13.4824>
- Fledel-Alon, A., Wilson, D. J., Broman, K., & Przeworski, M. (2009). Broad-scale recombination patterns underlying proper disjunction in humans. *PLoS Genetics*, 5(9), e1000658. <https://doi.org/10.1371/journal.pgen.1000658>
- Fyon, F., Cailleau, A., & Lenormand, T. (2015). Enhancer runaway and the evolution of diploid gene expression. *PLoS Genetics*, 11(11), e1005665–e1005665. <https://doi.org/10.1371/journal.pgen.1005665>
- Genestier, A., Duret, L., & Lartillot, N. (2024). Bridging the gap between the evolutionary dynamics and the molecular mechanisms of meiosis: A model based exploration of the PRDM9 intra-genomic Red Queen. *PLoS Genetics*, 20(5), e1011274. <https://doi.org/10.1371/journal.pgen.1011274>
- Haig, D., & Grafen, A. (1991). Genetic scrambling as a defence against meiotic drive. *Journal of Theoretical Biology*, 153(4), 531–558. [https://doi.org/10.1016/S0022-5193\(05\)80155-9](https://doi.org/10.1016/S0022-5193(05)80155-9)
- Hudson, R. R., & Kaplan, N. L. (1995). Deleterious background selection with recombination. *Genetics*, 141(4), 1605–1617. <https://doi.org/10.1093/genetics/141.4.1605>
- Imai, Y., Baudat, F., Taillepierre, M., & de Massy, B. (2017). The PRDM9 KRAB domain is required for meiosis and involved in protein interactions. *Chromosoma*, 126(6), 681–695. <https://doi.org/10.1007/s00412-017-0631-z>
- Jeffreys, A. J., Cotton, V. E., Neumann, R., & Lam, K.-W. G. (2013). Recombination regulator PRDM9 influences the instability of its own coding sequence in humans. *Proceedings of the National Academy of Sciences of the United States of America*, 110(2), 600–605. <https://doi.org/10.1073/pnas.1220813110>
- Jeffreys, A. J., & Neumann, R. (2002). Reciprocal crossover asymmetry and meiotic drive in a human recombination hot spot. *Nature Genetics*, 31, 267. <https://doi.org/10.1038/ng910>
- Jeffreys, A. J., & Neumann, R. (2005). Factors influencing recombination frequency and distribution in a human meiotic crossover hotspot. *Human Molecular Genetics*, 14(15), 2277–2287. <https://doi.org/10.1093/hmg/ddi232>
- Kimura, M. (1967). On the evolutionary adjustment of spontaneous mutation rates. *Genetical Research*, 9, 23–34. <https://doi.org/10.1017/S0016672300010284>
- Kong, A., Frigge, M. L., Masson, G., & Stefansson, K. (2012). Rate of de novo mutations and the importance of father's age to disease risk. *Nature*, 488(7412), 471–475. <https://doi.org/10.1038/nature11396>
- Lenormand, T., Engelse, J., Johnston, S. E., & Haag, C. R. (2016). Evolutionary mysteries in meiosis. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*, 371(1706), 20160001. <https://doi.org/10.1098/rstb.2016.0001>
- Lenormand, T., Fyon, F., Sun, E., & Roze, D. (2020). Sex chromosome degeneration by regulatory evolution. *Current Biology*, 30(15), 3001–3006.e5.
- Liberman, U., & Feldman, M. (1986). Modifiers of mutation-rate—a general reduction principle. *Theoretical Population Biology*, 30(1), 125–142. [https://doi.org/10.1016/0040-5809\(86\)90028-6](https://doi.org/10.1016/0040-5809(86)90028-6)
- Liberman, U., Van Cleve, J., & Feldman, M. W. (2011). On the evolution of mutation in changing environments: Recombination and Phenotypic Switching. *Genetics*, 187(3), 837–851. <https://doi.org/10.1534/genetics.110.123620>
- Lynch, M., Ackerman, M. S., Gout, J. F., & Foster, P. L. (2016). Genetic drift, selection and the evolution of the mutation rate. *Nature Reviews Genetics*, 17(11), 704–714. <https://doi.org/10.1038/nrg.2016.104>
- Myers, S., Bottolo, L., Freeman, C., & Donnelly, P. (2005). A fine-scale map of recombination rates and hotspots across the human genome. *Science*, 310(5746), 321–324. <https://doi.org/10.1126/science.1117196>
- Myers, S., Bowden, R., Tumian, A., & Donnelly, P. (2010). Drive against hotspot motifs in primates implicates the PRDM9 gene in meiotic recombination. *Science*, 327(5967), 876–879. <https://doi.org/10.1126/science.1182363>
- Myers, S., Freeman, C., Auton, A., & McVean, G. (2008). A common sequence motif associated with recombination hot spots and genome instability in humans. *Nature Genetics*, 40(9), 1124–1129. <https://doi.org/10.1038/ng.213>

- Nachman, M. W., & Payseur, B. A. (2012). Recombination rate variation and speciation: Theoretical predictions and empirical results from rabbits and mice. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 367(1587), 409–421. <https://doi.org/10.1098/rstb.2011.0249>
- Oliver, P. L., Goodstadt, L., Bayes, J. J., & Ponting, C. P. (2009). Accelerated evolution of the Prdm9 speciation gene across diverse Metazoan Taxa. *PLoS Genetics*, 5(12), e1000753. <https://doi.org/10.1371/journal.pgen.1000753>
- Otto, S. P., & Barton, N. H. (2001). Selection for recombination in small populations. *Evolution*, 55(10), 1921–1931.
- Otto, S. P., & Lenormand, T. (2002). Resolving the paradox of sex and recombination. *Nature Reviews Genetics*, 3(4), 252–261. <https://doi.org/10.1038/nrg761>
- Paigen, K., & Petkov, P. (2010). Mammalian recombination hot spots: Properties, control and evolution. *Nature Reviews Genetics*, 11(3), 221. <https://doi.org/10.1038/nrg2712>
- Paigen, K., & Petkov, P. M. (2018). PRDM9 and its role in genetic recombination. *Trends in Genetics*, 34(4), 291–300. <https://doi.org/10.1016/j.tig.2017.12.017>
- Petes, T. D. (2001). Meiotic recombination hot spots and cold spots. *Nature Reviews Genetics*, 2(5), 360. <https://doi.org/10.1038/35072078>
- Ponting, C. P. (2011). What are the genomic drivers of the rapid evolution of PRDM9? *Trends in Genetics*, 27(5), 165–171. <https://doi.org/10.1016/j.tig.2011.02.001>
- Powers, N. R., Parvanov, E. D., Baker, C. L., & Paigen, K. (2016). The meiotic recombination activator PRDM9 trimethylates both H3K36 and H3K4 at recombination hotspots in vivo. *PLoS Genetics*, 12(6), 1–24. <https://doi.org/10.1371/journal.pgen.1006146>
- Ptak, S. E., Hinds, D. A., Koehler, K., & Pääbo, S. (2005). Fine-scale recombination patterns differ between chimpanzees and humans. *Nature Genetics*, 37(4), 429. <https://doi.org/10.1038/ng1529>
- Ren, H., Ferguson, K., Kirkpatrick, G., & Ma, S. (2016). Altered crossover distribution and frequency in spermatocytes of infertile men with azoospermia. *PLoS One*, 11(6), e0156817. <https://doi.org/10.1371/journal.pone.0156817>
- Stevison, L. S., Woerner, A. E., Kidd, J. M., & Wall, J. D. (2015). The time scale of recombination rate evolution in great apes. *Molecular Biology and Evolution*, 33(4), 928–945. <https://doi.org/10.1093/molbev/msv331>
- Úbeda, F., & Wilkins, J. F. (2011). The Red Queen theory of recombination hotspots. *Journal of Evolutionary Biology*, 24(3), 541–553. <https://doi.org/10.1111/j.1420-9101.2010.02187.x>
- Úbeda, F., Fyon, F., & Bürger, R. (2023). The recombination hotspot paradox: Co-evolution between PRDM9 and its target sites. *Theoretical Population Biology*, 153, 69–90. <https://doi.org/10.1016/j.tpb.2023.07.001>
- Úbeda, F., & Haig, D. (2005). On the evolutionary stability of Mendelian segregation. *Genetics*, 170, 1345–57. <https://doi.org/10.1534/genetics.104.036889>
- Úbeda, F., Patten, M. M., & Wild, G. (2015). On the origin of sex chromosomes from meiotic drive. *Proceedings. Biological Sciences*, 282(1798), 20141932–20141932.
- Úbeda, F., Russell, T. W., & Jansen, V. A. (2019). PRDM9 and the evolution of recombination hotspots. *Theoretical Population Biology*, 126, 19–32. <https://doi.org/10.1016/j.tpb.2018.12.005>
- Winckler, W., Myers, S. R., Richter, D. J., & Altshuler, D. (2005). Comparison of fine-scale recombination rates in humans and chimpanzees. *Science*, 308(5718), 107–111. <https://doi.org/10.1126/science.1105322>
- Wu, H., Mathioudakis, N., Diagouraga, B., & Kadlec, J. (2013). Molecular basis for the regulation of the H3K4 methyltransferase activity of PRDM9. *Cell Reports*, 5(1), 13–20. <https://doi.org/10.1016/j.celrep.2013.08.035>