Prdm9 is an Anti-Speciation Gene

Donald R. Forsdyke

Donald R. Forsdyke
forsdyke@queensu.ca

Department of Biomedical and Molecular Sciences, Queen's University, Kingston, ON, Canada

Abstract Mechanisms initiating a branching process that can lead to new species are broadly classified as chromosomal and genic. Chromosomal mechanisms were supported by breeding studies involving exchanges of individual chromosomes between mouse subspecies. There were also studies of the rapidly mutating mouse *PR/SET-domain 9* (*prdm9*) gene, which encodes PRDM9, a protein targeting DNA recombination hotspots. When PRDM9 is bound symmetrically with equal strength, the meiotic repair of mutations in one parental strand, based on information on the allelic strand (conversion), would seem to be unbiased in discriminating between strands. So mismatches detected between pairing paternal and maternal DNA strands (heteroduplexes) would undergo unbiased conversions (to homoduplexes). This would leave uncertainty on whether a mutation had been corrected or compounded. However, a hypothetical tagging of mismatch regions, so that both strands are epigenetically marked as uncertain, would make it possible over numerous generations for mutations to be corrected (biased conversions) whenever *asymmetry* is detected. Thus variation would decrease and members of a species would remain within its bounds. Intriguingly, new experimental studies show that, when chromosomally interpreted, PRDM9 also works through *asymmetrical* epigenetic labelling to confine members to species bounds. To the extent that the experimentally observed and hypothetical anti-speciation asymmetries can be related, chromosomal mechanisms are further supported.

Keywords: Chromosomal speciation, Epigenetic marking, Haldane's rule, Hotspot paradox, Hybrid sterility, PRDM9

Introduction

Hypotheses on the initiation of a branching process that can lead to new species are broadly classified as chromosomal and genic (Forsdyke 2001, 2010; Coyne and Orr 2004; Nei and Nasawa 2011; Nevo 2012). Agreement is sought as to which initiation mechanisms are actually, rather than hypothetically, capable of originating species, and which are most likely to have operated in the general case (Kliman et al. 2001; Forsdyke 2004; Johannesson 2010). Chromosomal hypotheses invoke sequence disparities between parental chromosomes such that meiotic pairing of 'homologous' chromosomes fails within the gonad of their offspring (hybrid). Thus, there can be no exchange of DNA segments (recombination) and the production of gametes ceases (hybrid sterility). In the light of "fresh evidence for a genetic connection between recombination and hybrid sterility," there is now growing recognition of "the intriguing possibility that recombination and speciation are mechanistically coupled" (Payseur 2016). However, the 'speciation genes' commonly invoked to explain this have become increasingly elusive.

Specific genes concerned with the initiation of speciation have been proposed for fruit fly (Mallet 2006). While doubt is cast on the existence of similar genes in other species (Schartl 2008; Louis 2009; Kao et al. 2010), a gene expressed in the early stages of germ cell maturation (Hayashi et al. 2005; Mihola et al. 2009) is now thought a likely "first mammalian candidate for a speciation gene" (Flachs et al. 2012). More definitely, it is "the only mammalian speciation gene yet identified" (Davies et al. 2016). However, studies to be reviewed here suggest that the major role of this gene (PR/SET-domain 9; *prdm9*) is to encode a genome maintenance protein that works to *retain a line of organisms within species bounds*. Thus, it is better regarded as an *inhibitor* of speciation – an anti-speciation gene – that opposes speciation when initiated chromosomally (Reese and Forsdyke 2016; Forsdyke 2016). Whatever the outcome, "cracking the curious case of PRDM9 promises to provide important insights into large swaths of biology, from human genetics to speciation" (Ségurel et al. 2011).

The paper has four parts. The first describes the role of hybrid sterility in chromosomal hypotheses and draws attention to the long-held view that meiosis, apart from being a widely-recognized generator of diversity, can also *reduce* diversity. The second part turns to the *prdm9*

gene and the role of its protein product, PRDM9, in adding epigenetic methylation marks to histones. This designates the chromosomal location of recombination hotspots where there can be an asymmetric conversion of information in one parental genome to that of the other. However, like some other genes whose products target DNA, *prdm9* is on a mutational 'treadmill,' being forced to change rapidly to keep pace with DNA sequence changes. Its DNA target 'calls the tune,' yet in a paradoxical way. Hotspots appear self-destructive in that a given PRDM9 protein faces, generation after generation, diminishing target availability, thus 'obliging' its gene to rapidly mutate in response to an unexplained evolutionary pressure for the designation of new targets. The third part deals with the new light cast on chromosomal speciation by experimental modifications of PRDM9 that confers on it the ability to tune-call, so shifting hotspot locations. Finally, the possibility is considered that these observations relate to a postulated asymmetry requirement of the epigenetic marking of sequences whose accuracy is in doubt (Reese and Forsdyke 2016).

1. Chromosomal Basis of Hybrid Sterility

Hybrid Sterility Can Only Be Primary

A state of reproductive isolation exists between two organisms when they are either unable to produce offspring, or their offspring are sterile. The 'spark' that initiates a new species involves mechanisms for achieving reproductive isolation. This ensures that the speciation process is not subverted by recombination between the genomes of diverging types – such recombination would tend to blend (homogenize), rather than retain, differences. Reproductive isolation begins with interruption of the germ-line (germinal) cycle – gamete, zygote, embryo, meiotic adult gonad, gamete, and so on. Being a recursive cycle, any point, be it before or after union of gametes to form a zygote (i.e. be it either pre- or post-zygotic), will serve to mediate the primary interruption. Whatever the point, for successful branching evolution, two independent cycles – two species – must eventually emerge (Fig. 1).

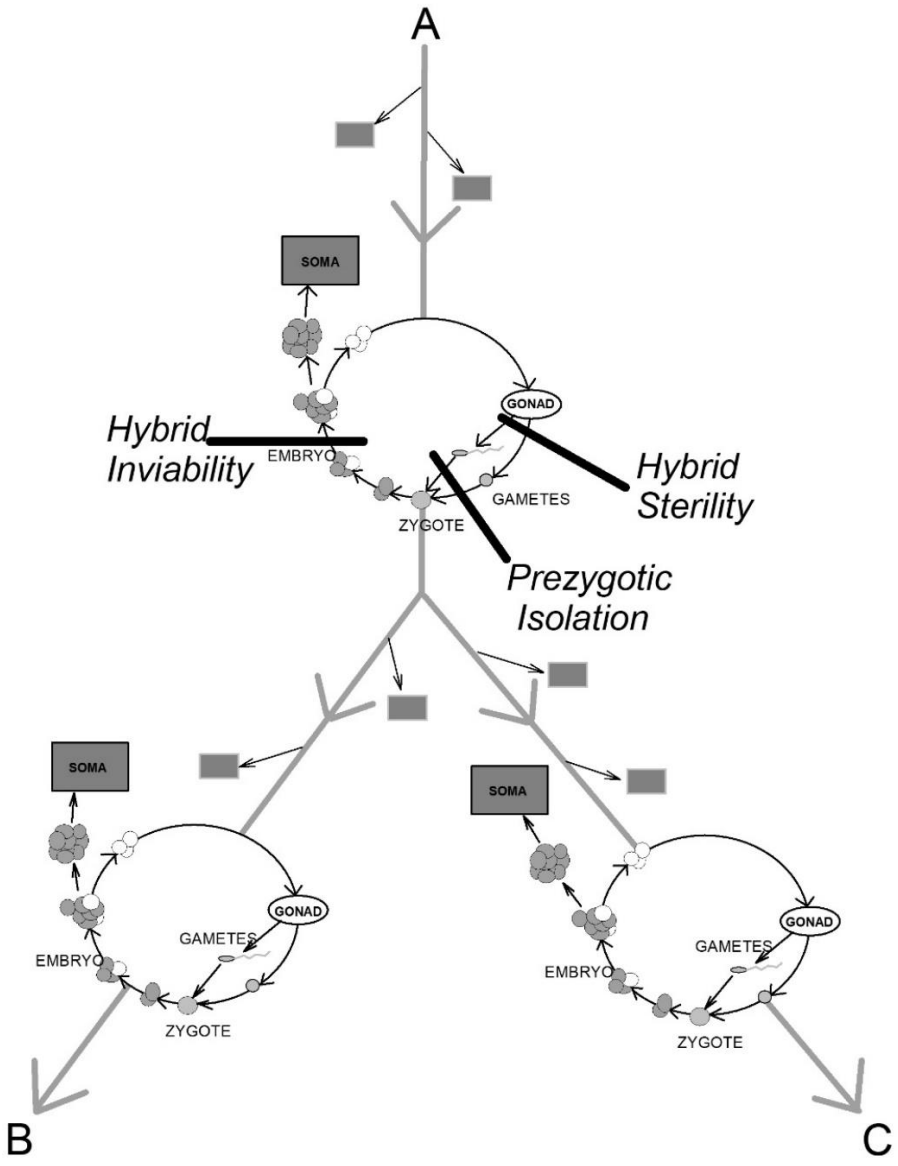

Fig. 1 Branching of species A into new species (B and C). The mortal soma (large grey boxes) provides support for the gonad (white ovals), but is discarded in each generation (small grey boxes). The germinal cycle operates continuously to facilitate within-species sexual crossing until there appears one of three barriers (black bars). These have the potential to stop production of fertile offspring by preventing either fertilization (prezygotic isolation), or development (hybrid inviability), or gamete formation (hybrid sterility). When speciation is complete, two independent germinal cycles maintain the continuity of species B and C through the generations. To simplify, male and female gametes here share a common cycle.

There are many possible mechanisms for initiating cycle interruption. If the primary interruption is at meiosis, so that gamete formation in an individual's gonad is impaired, then the parents of that individual (who is their hybrid) are reproductively isolated from each other. The hybrid sterility of their *otherwise normal* offspring is a manifestation of *their* partner-specific genome phenotypes. Disparities in the DNA sequences they have contributed have been sufficient to impair the chromosome pairing needed for the error-correction that can oppose sequence variation. Their offspring has been denied the meiotic pairing (synapsis) that can work to *decrease* intra-species diversity and thus retain organisms within species bounds (Forsdyke 1996; 2001; 2016; 2017a, b; Reese and Forsdyke 2016)**.** When the disparity within their offspring's gonad is not correctable, then only within the bounds of an emerging new species will each parent be likely to find a non-disparate partner.

A *primary* disparity at the meiotic level – usually of non-genic origin (see later) – may serve as a barrier that, for a while, secures some independence for the emerging cycles. But usually the barrier is incomplete, so that some mixture of gene pools remains possible. Later a *secondary* disparity– a barrier that can more securely achieve reproductive isolation – may arise at another level, eventually replacing the primary barrier. Thus, partially shielded from blending by the primary barrier, members of one incipient species may accept mutations in genes which consequently no longer interact properly with those of members of the other incipient species with whom they cross. This secondary barrier between the species may affect the development of their offspring, which hence are *not normal* (hybrid inviability; Fig. 1). The parents are now more profoundly reproductively isolated.

In some circumstances the latter developmental failure may also be a primary cause of cycle interruption. So hybrid inviability can be either primary or secondary. However, hybrid sterility can *only* be primary, since if a hybrid is inviable then *there can be no adult hybrid to manifest sterility*. The onus is upon those believing a case of hybrid inviability to be primary, to show that it had not been preceded by a period, however brief, of hybrid sterility. By the same token, the onus is upon those holding certain cases of prezygotic isolation to be primary (e.g. elephant cannot copulate with mouse), to show that they had not been preceded by periods of postzygotic hybrid inviability (genically determined) or sterility (chromosomally determined).

Thus, regarding the speciation process there is a "multi-dimensional … continuum" (Mérot et al. 2017). Barriers are hierarchically arranged and although an early member of a hierarchical sequence may sometimes be forestalled (preempted) by a later, there must be evidence that the early member had not already acted, before concluding that it was not primary. A determination that the hybrid sterility barrier was absent at species initiation may be difficult since, once liberated by a secondary barrier from its reproductive isolation role, it becomes free to respond to other pressures (Forsdyke 2001, p. 22).

Chromosomal Differences and Haldane's Rule

A primary role for hybrid sterility was suggested from the generalization that, prior to the development of full sterility among offspring (F1 hybrids), the sterility may be partial in that only one sex (male in mice) is affected (Haldane 1922). Furthermore, experimental exchanges ("introgressions") of individual chromosomes between "closely related mouse subspecies" have shown that "heterospecific autosomal pairs in sterile hybrids are more prone to asynapsis than the homospecific pairs in which both homologs came from the same species" (Bhattacharyya et al. 2013). Even among heterospecific pairs, some chromosomes appear more disposed to asynapsis than others. Thus, Bhattacharyya et al. (2014) note that "the number of unsynapsed autosomes per cell varies, indicating the same type of cis-acting mechanism operates on individual autosomes." Their observations are attributed, not to some mobile ("trans-acting") genic factor with the ability to single out individual chromosome pairs, but to "their fast evolving non-genic divergence," which could have affected some chromosome pairs more than others (Bhattacharyya et al. 2013).

Being *already* disparate, the sex chromosomes in male mice (XY) regularly fail to pair along most of their lengths. Relative to females that have acquired homologous sex chromosomes (XX), this gives males (the "heterogametic" sex) a head-start along the path towards full sterility (Haldane's rule for hybrid sterility; Forsdyke 2000). Since both *prdm9* and a hemizygous gene on the male X chromosome (*hstx2*; derived from the female parent) play a role, Bhattacharyya et al. (2013) conclude that "variation in pairing failure is under genic control," but the sterility itself "is chromosomal, caused by heterospecific pairing incompatibilities." They deem this supportive

of similar suggestions regarding sterility in both fruit fly (Naviera and Maside 1998; Moehring 2011), and yeast ("simple sequence divergence acted upon by the mismatch repair system;" Louis 2009).

Whether sex chromosomes are the same size (homomorphic) or of different size (heteromorphic), these considerations also apply to plants with independent sexes (Ironside and Filatov 2005). Indeed, Delph and Demuth (2016), claiming to have "the best explanation for male rarity in some *Silene* hybrids," consider that "although the original chromosomal mechanism … largely fell out of favor, recent work has argued for its importance on theoretical grounds." Furthermore, demonstrating the progressive involvement of autosomes *after* a primary involvement of sex chromosomes, Hu and Filatov (2016) find that a "large-X effect in plants" has "increased species divergence and reduced gene flow on the *Silene* X-chromosome," but gene flow involving autosomal loci is still "sufficient to homogenize the gene pools of the two species."

Meiosis Restrains Diversity

That meiosis, apart from generating diversity by shuffling parental genomes, also has the potential to restrain diversity, has a long history. Montgomery (1901, p. 223) proposed that meiosis would "rejuvenate" the colorfully staining material – chromatin – in disparate chromosomes:

> When the spermatozoon conjugates with the ovum there is a mixture of cytoplasm with cytoplasm, of karyolymph with karyolymph, … but there is no intermixture of chromatin, for the chromosomes … remain more separated from one another than at any other stage … . But after this beginning stage of the germinal cycle, … in the synapsis stage … we find a positive attraction between paternal and maternal chromosomes. The reason for the final union of these chromosomes is obvious; it is evidently to produce a rejuvenation of the chromosomes. From this standpoint the conjugation of the chromosomes in the synapsis stage may be considered the final step in the process of conjugation of the germ cells. It is a process that effects the rejuvenation of the chromosomes; such rejuvenation could not be produced unless chromosome of different parentage joined together, and there would be no apparent reason for chromosomes of like parentage to unite.

That the rejuvenation could involve exchange of linear information between disparate chromosomes was outlined by Winge (1917) and later elaborated in terms of DNA mismatch repair (Bernstein and Bernstein 1991; Forsdyke 2001, 2007; Gorelick and Heng 2010). If we assume rejuvenation to imply a restoration ("correction") of DNA sequences that have strayed from the species norm (i.e. they are to be returned closer to their formerly 'pristine' base composition and order), then it follows that a *lack* of such rejuvenation would allow DNA differences to accumulate, so possibly leading to reproductive isolation and speciation. Rejuvenation would restrain diversification, and hence, speciation. Fundamental to our understanding of this are mechanisms of meiotic recombination.

"Pair First" is First

For homologous recombination two DNA duplexes must pair and exchange segments. This requires both recognition of some degree of sequence similarity and strand breakage. The temporal order of these events is contentious. A popular model postulates cutting to produce a single strand that *then* seeks a pairing partner in the homologous duplex (Szostak et al. 1983). However, a growing view, summarized by Zickler and Kleckner (2015), suggests otherwise:

> A prominent, but still mysterious, feature of chromosome biology is the ability of homologous chromosomes, or chromosomal regions, to specifically recognize and pair with one another in the apparent absence of DNA lesions (DSBs) or recombination. … Recombination-independent pairing … plays prominent roles for premeiotic and meiotic programs, where it is defined as pairing that occurs before and/or without … DSBs"

A "cut first" model implies a *localized commitment* prior to pairing. A "pair first" model should more reliably afford reversible genome-wide homology *testing*, without commitment (McGavin 1977; Wilson 1979; Boeteng et al. 2013). An initial alignment through "kissing" interactions between the loops of extruded DNA stem-loop structures requires only loop-loop base-pair complementation (Forsdyke 2007). In the absence of base-pair complementarity in stem regions adjacent to loops, the pairing would not extend from the loops; hence the "kissing" would not be meiotically consummated. Thus, initial *breakage-independent* homology recognition might not lead to strand breakage and segment exchange. This would depend on

whether dispersed loop homologies were, or were not, interspersed with stem sequences that were also homologous.

Indeed, from studies of homology-directed DNA changes in fungi (repeat-induced point mutation; RIP), Gladyshev and Kleckner (2014) provide "a new perspective … for … the breakage-independent recognition of homology that underlies RIP and, potentially, other processes where sequence-specific pairing of intact chromosomes is involved." Thus, "the nucleotide composition of participating DNA molecules is identified as an important factor," and "homology recognition is modulated by the underlying sequence" (Gladyshev and Kleckner 2016). Accordingly, "sequence information can be compared directly between double-stranded DNA molecules during RIP," and there is the potential for application to "other processes where homologous pairing of intact DNA molecules is observed." This view is supported by later studies (Gladyshev and Kleckner 2017).

However, the "pair first" and "cut first" views on the process by which genomes of paternal and maternal origin exchange information in the gonad of their child, should not necessarily be seen as mutually exclusive. A limited number of pair-first sites might suffice (i.e. provide anchor points) to assist the close apposition of homologous chromosomes. Once this alignment was achieved, the homology-search task of single-strands liberated by a "cut first" mechanism could be much easier.

2. PRDM9 and Recombination Hotspots

Histone Modifications by PRDM9

Although controversial, the above studies indicate that nucleic acid pair-first mechanisms dominate recombination in modern organisms. Subversion of such mechanisms by sequence disparity would create conditions favorable for speciation. Subversion would arise from retention of a strict homology-dependence requirement in the face of increasing sequence disparity. Furthermore, while cellular nucleic acids are closely associated with proteins that could affect recombination-related nucleic acid interactions, many aspects of DNA pair-first mechanisms can

be reproduced in vitro in the absence of proteins (Danilowicz et al. 2009; Forsdyke 2016, pp. 175-192).

Assuming nucleic acid evolution to have preceded that of proteins (Reanney 1979; Poole and Logan 2005), one can envisage the kinetics of recombination – entirely dependent on nucleic acid chemistry (Forsdyke 2007) – to have been enhanced and modified by later evolving proteins. Thus, in practice, pair-first mechanisms cannot really be first. There is a need for a prior modification of proteins, such as the histones around which DNA is organized as 'nucleosome' complexes. Modification of such nucleoprotein (chromatin) complexes to improve access to DNA is one of the functions of PRDM9, the 'zinc finger' part of which recognizes with high specificity a short DNA sequence motif at the nucleosome surface, so defining a 1-2 kb recombination 'hotspot' region. This recognition activates the protein's PR/SET domain that catalyzes the local addition of methyl groups to certain lysine residues of one of the histone proteins (H3). Trimethylation of H3 generates a 'nucleosome-depleted region' where recombination is facilitated (i.e. homologous strands can pair, there is branch migration to increase the area scanned, and the enzyme causing double-strand breaks, SpoII, can be recruited).

Conversion Decreases Diversity

Meiosis within a parental gonad generally provides for the unbiased distribution of grandpaternal and grandmaternal genomic information among offspring. Sometimes detected as a failure of characters to follow a Mendelian pattern of distribution, the *asymmetrical* conversion of information from one genome into that of the other has been called "gene conversion." But, in principle, the asymmetry would also apply to non-genic sequences. Given two formerly homologous segments, A and B, which have come to differ slightly (i.e. are now heterologous), information in A can be converted into that of B, or vice-versa. Thus, regarding individual offspring, diversity is diminished. However, the direction of the conversion may be *generally* unbiased in that in one individual A may have converted to B, and in another B may have converted to A. Thus, at the population level, in the absence of conversion bias there will be no change in the frequencies of A and B. Diversity is not diminished.

On the other hand – rather than let natural selection be the arbiter of frequencies should one segment prove more advantageous than the other – *if there were sufficient information*, a parent could 'place an educated bet' on the conversion direction to adopt. For example, if it were 'known' that A was the species norm and/or that B had experienced a recent mutation, then there could be an appropriate conversion bias. Since most mutations are detrimental, resulting in negative selection of the organisms within which they occur, the best bet would be to convert B into A. If all parents were to bet similarly then diversity within the population would be diminished. Intriguingly, various studies show that a sequence motif recognized by PRDM9 (so defining a recombination hotspot) is directionally converted in mice heterozygous for that hotspot in favor of the sequence *less strongly* recognized by that protein ('meiotic drive'). I consider later the possibility of 'educated bets' in the light of studies of incipient speciation in mice.

DNA Inconstancy Drives Protein Treadmill

Proteins often act as enzymes catalyzing changes in other molecules – their specific substrates – which may be micromolecules (e.g. glucose), or macromolecules (e.g. DNA). In the course of evolution there are changes in proteins as a consequence of mutations in the genes that encode those proteins. Their substrates do not usually participate in this directly. For example, as organisms evolve, specific enzymes may change and so improve the utilization of glucose. That's natural selection. *But glucose itself has no say in this*. It does not change to lighten, or impede, the task of the enzymes. For some enzymes, however, DNA is a specific substrate. DNA both encodes the enzymes (locally in the regions of their genes), and is either their local or general substrate (target). However DNA, by virtue of base changes, can change its character. Thus, it has the power to lighten or impede the tasks of the enzymes that act on it. More than this, the *constancy* of glucose means that the enzymes of glucose metabolism can count on glucose being the *same* from generation to generation. The *limited constancy* of DNA means that some of the proteins that react with it are on an evolutionary treadmill. They are "red queens" that must change as their substrate changes (Lesecque et al. 2014).

While some proteins target nucleic acids generically by virtue of their unchanging phosphate-ribose structures, others are not indifferent to the composition or order of bases. So, as nucleic acid bases change, the proteins that target nucleic acids more specifically may need to adapt functionally to ensure optimal binding to their substrate. This *adaptive treadmill* implies that, as lines diverge, the mutation rate in nucleic acid recognition proteins must be greater than in most other proteins (Paz et al. 2006). Thus there is positive selection in that, if its genes encoding nucleic acid recognition proteins do not rapidly mutate, an organism may not survive (negative selection). While here the DNA "dog" may wag the protein "tail," sometimes the converse may apply. Changes in proteins encoded by certain "mutator" genes can exert a genome-wide effect on base composition (Cox and Yanofsky 1967). Likewise a mutated *prdm9* gene can, in one fell swoop, exert a genome-wide effect on the nature and location of target sites that focus recombination events (see below). Paradoxically, however, over evolutionary time it is usually these target sites that "call the tune" resulting in the positive selection of *prdm9* mutants: "PRDM9 has been under selective pressure to switch to new targets. However, the reasons for this selective pressure remain mysterious" (Lesecque et al. 2014).

.

The Hotspot Conversion Paradox

The genomic locations of recombination events are usually not randomly distributed. In microorganisms there are specific, mainly gene-located, "crossover hotspot instigator" (Chi) sequences that are recognized by enzymes with a role in recombination (Lao and Forsdyke 2000). Multicellular organisms also have sequence hotspots where recombination most frequently initiates. Such hotspots may relate to genic boundaries (e.g. transcription start sites), either as a primary feature (Forsdyke 2011a), or as a default option when sites (motifs) recognized by PRDM9 are not operative (Brick et al. 2012). In organisms possessing a functional *prdm9* gene (e.g. primates, rodents) the primary hotspot option is often extra-genic or within introns. These DNA regions have more potential than exons to extrude kissing-loop structures that might mediate recombination (Forsdyke 1996, 2007).

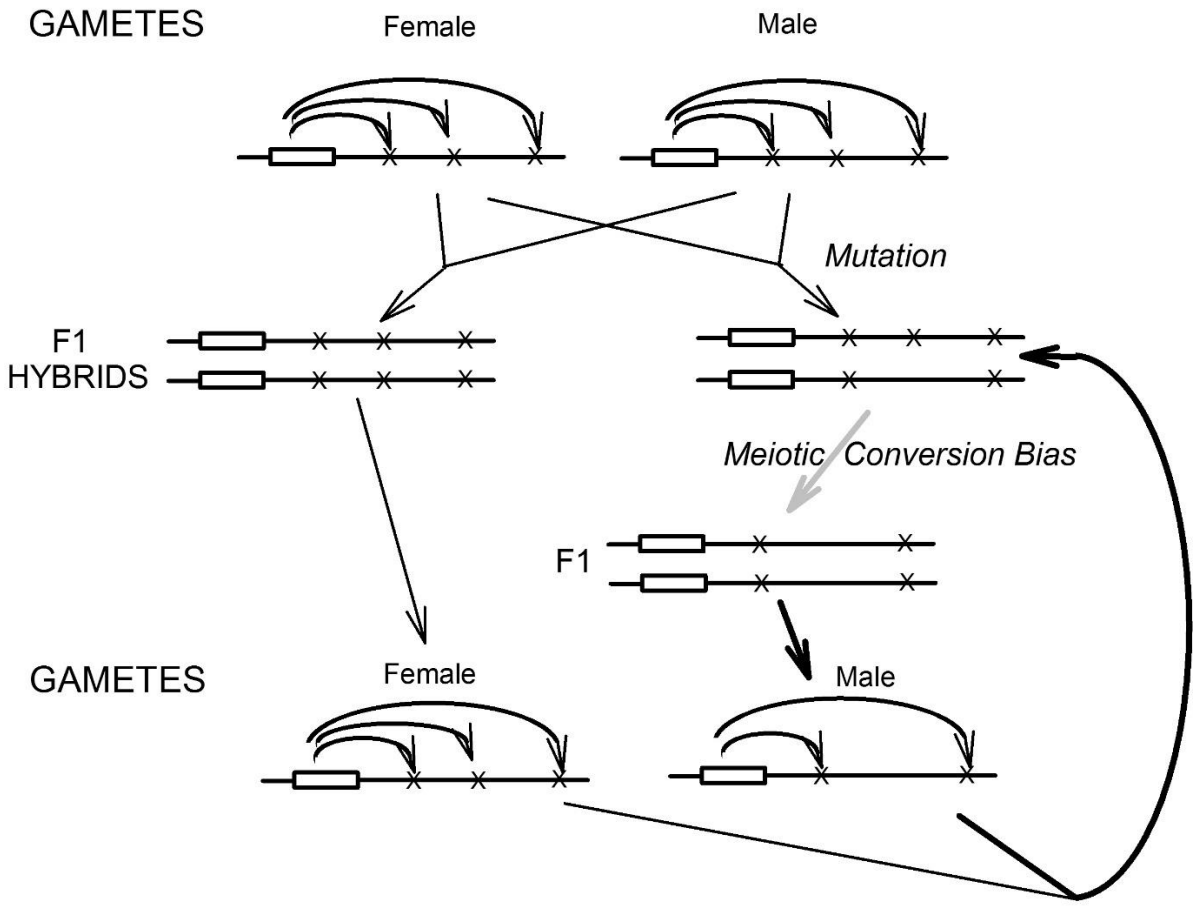

Fig. 2 Cyclic erosion of PRDM9 binding sites during repeated within-species crosses. Homologous *prdm9* genes are shown as white rectangles flanked by horizontal lines (genomes). At top, curved arrows indicate potential PRDM9 targets (X) at various genomic locations. At left, parental gametes unite to form normal F1 hybrids with symmetrical target sites. These hybrids, in turn, form parental-type gametes (the choice of sex is here arbitrary). At right, a mutation has changed a target site in one parental gamete (loss of X). Despite the local asymmetry, both normal and mutant sites are recognized at meiosis by PRDM9, and chromosomal pairing and recombination occur, albeit perhaps less efficiently. However, there is a conversion bias in favor of the more weakly recognized mutant parent genome. While, through the generations, members of the majority population continue to produce unmutated target sites (scheme at left), recurrent crosses with members of the growing mutant population (the cycle on the right; thick lines), erode this lead. Thus, the *prdm9* gene, in its X-recognizing form, becomes progressively irrelevant.

Chi being a conserved sequence in bacteria, it was at first thought that there would be a conserved, specifically targeted, binding sequence for PRDM9. Much attention was paid to a degenerate 13-base consensus motif that was often found in human hotspots. However, although recombination initiation hotspots may endure for many generations, eventually they disappear (erode) due to an unexplained *directional* conversion bias that accompanies recombination (Fig. 2). Thus, Wahls and Davidson (2011) observed that "hotspots seed their own destruction," and hence, to maintain recombination, fresh sites with different sequence characteristics must emerge. Focusing on yeast, which do not have the *prdm9* gene, they noted that "within the genome is a collection of hotspot-active DNA sites and a reservoir of 'cryptic' DNA sequence motifs that can be rendered active by as little as a single-base-pair substitution." Under this "equilibrium dynamic model" the substitution refers to a DNA target motif, not to the enzyme that recognizes that motif. However, the possibility was entertained that for organisms with the *prdm9* gene, novel motifs could be so designated by its PRDM9 product if the gene were appropriately mutated ("*prdm9* shift model"). Accordingly, the widely distributed hotspot sequences that had eroded could be viewed as having 'called the tune,' and the gene had 'responded' by mutation (i.e. the mutation conferred some selective advantage). The latter view is now supported by extensive studies, such as those on the infertility of the male offspring of crosses between mouse subspecies carried out in the Forejt laboratory (see later).

When comparing diverging sequences it is argued that that which is conserved between the sequences is evolutionarily important, whereas that which is less conserved (e.g. so-called "junk DNA") is less important – indeed, that is why the 'hand of evolution' may have 'chosen' to discard it. Thus, when confronting the "hotspot conversion paradox" (Fig. 2), Boulton et al. (1997) noted that since "biased gene conversion is a typical consequence of recombination at hotspots," then "the sites thought to initiate crossing over cannot be maintained by the benefits of the events they cause". In other words, change *itself* seemed to have an adaptive value.

So there are circumstances under which non-conservation seems to have a *greater* adaptive value than conservation – as has been argued in another context for immunologically relevant RNA species (Forsdyke 2016, p. 279-303). Likewise, proteins such as PRDM9 are themselves transient, displaying extensive variation over time. Their genes are on a treadmill, having to

mutate to keep up with the constantly changing hotspot landscape. Thus there is high within-species variation (e.g. polymorphism), and even greater between-species (e.g. human-mouse) variation, in the DNA recognition regions (zinc fingers) of *prdm9* genes. Grey et al. (2011) noted:

> The PRDM9 gene is well-conserved among metazoans, however the domain encoding the zinc finger array experiences an accelerated evolution in several lineages, including rodents and primates. This accelerated evolution is restricted to codons responsible for the DNA-binding specificity of PRDM9 zinc fingers, which appear to have been subjected to positive selection.

3. PRDM9 and Speciation

Subspecies as Incipient Species

As indicated above, studies from the Forejt laboratory (Bhattacharyya et al. 2013), where members of two mouse subspecies (*Mus musculus musculus* (*Mmm*) and *Mus musculus domesticus* (*Mmd*)), were crossed, support chromosomal, rather than genic, disparity as a basis for species initiation. The subspecies had not been designated as "species" because the reproductive isolation that defines species was incomplete in that a degree of fertility was still evident among female offspring. The subspecies were incipient species with a potential eventually to attain full species status. Two important strains within the subspecies are PWD (from *Mmm*) and B6 (from *Mmd*). They are estimated to have diverged from a common mouse ancestor about 0.5 million years ago and, while their DNA sequences are closely identical, they differ in their *prdm9* genes and the multiple specific targets of the PRDM9 proteins encoded by those genes. The latter targets can be represented (see Fig. 3, top) as Xs (for PWD) and black dots (for B6). These symbols represent the target sequence motifs *as they exist at the present time* and, largely because of *past* erosions, they are likely to *differ* from versions that were active at earlier times (closer to the time of the initial PWD-B6 divergence).

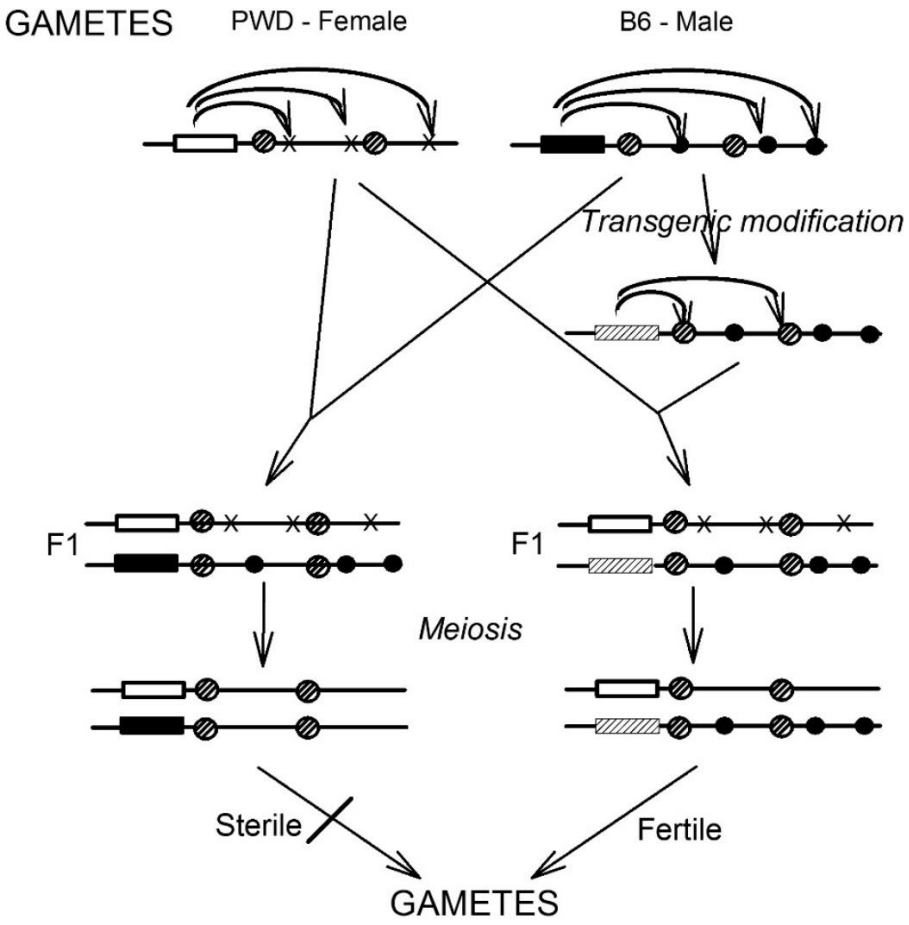

Fig. 3 Hybrid sterility induced by pervasive reciprocal asymmetry of PRDM9 target sites when members of mouse subspecies are crossed (left), and its rescue in a transgene "humanized" strain (right). At top, subspecies-specific *prdm9* genes are shown as rectangles (PWD strain, white; B6 strain, black). Corresponding potential genomic targets (Xs and black dots) are indicated by curve arrows. At left, gametes unite to form F1 hybrids that develop and grow normally. However, due to a pervasive asymmetrical meiotic conversion bias that impairs gamete formation, they are sterile. At right, the B6 version of the *prdm9* gene has been transgenically modified ("humanized") such that the zinc fingers no longer recognize B6 targets (black dots), but do recognize sequences elsewhere in the genome (striped dots) that happen to closely resemble the target of the human *prdm9* gene. When the humanized B6 mouse is crossed with a normal PWD mouse, meiosis proceeds normally due to the symmetry between targets (striped dots). This overcomes the effects of asymmetries related to the presence of the PWD version of *prdm9*, and the mice are fertile.

Erosion of Target Motifs

A major consequence of the erosion of specific target sites (Fig. 2), is that *today's* PWD mice encode PRDM9 proteins with *less* affinity for their target sites on the PWD genome than for target sites at the corresponding positions on the B6 genome (the latter are not symbolised in Fig. 3, top right). In other words, *today's* PWD-encoded PRDM9 proteins would, if given the opportunity, display *stronger* binding to a B6 genome than to a PWD genome. Likewise, *today's* B6-encoded PRDM9 proteins would, if given the opportunity, display *stronger* binding to a PWD genome than to a B6 genome. That opportunity arises when members of the two subspecies are crossed.

First we should note that, for a *within*-subspecies cross (Fig. 2), the decreased affinity does not disturb the general *symmetry* of binding of PRDM9 proteins. As long as appreciable affinity is retained, there should be fertile offspring. However, when there is a *between*-subspecies cross (Fig. 3) the binding is generally *asymmetrical* and the story is very different. When a PWD female is crossed with a B6 male, all male offspring are sterile (Haldane 1922). A likely mechanism for this is shown at the left of Fig. 3. Because of the prior erosion, on the genome donated by the PWD gamete the target sites for the PRDM9 encoded by the PWD version of *prdm9* are weaker than the *same* target sites on the genome donated by the B6 gamete. Similarly, on the genome donated by the B6 gamete the target sites for the PRDM9 encoded by the B6 version of *prdm9* are weaker than the *same* target sites on the genome donated by the PWD gamete. The wide spread asymmetry is somehow sufficient to impede meiotic pairing and the individual is sterile.

Humanized Targets Restore Fertility

For within-subspecies crosses, binding site affinity progressively decreases over evolutionary time (Fig.2), with the potential to lead to sterility. Thus, individuals with *prdm9* mutations affecting the PRDM9 zinc fingers (so designating fresh binding sites), should have a selective advantage. Indeed, experimental exchanges of mouse *prdm9* genes indicate restoration of fertility (Mihola et al. 2009). An extreme version of this was to insert a human zinc-finger DNA recognition region into a mouse *prdm9* gene (Fig. 3, right). Given both the great length of their

genomes, and the separation of humans and mice from their common ancestor some 150 million years ago, it was likely that there would be some symmetrical, non-eroded, potential target sequences in the mouse genome (indicated by striped dots in Fig. 3, right). These would be targeted, through its protein product, by the transgenically modified *prdm9* gene. Given that B6 targets (black dots in Fig. 3, right) were denied a corresponding B6 *prdm9*-encoded protein, the finding that the cross was fully fertile strongly indicated a primary role for PRDM9 zinc-finger mutations in preserving within species fertility (Davies et al. 2016).

Thus, the eroding multiple target sites in a genome 'call the tune.' The *prdm9* gene must 'respond' to maintain fertility. This maintenance is in keeping with the observation of Oliver et al. (2009) for humans that "allelic variations at the DNA–binding positions of human PRDM9 zinc fingers show significant association with decreased risk of infertility." Davies et al. (2016) concluded that "the full fertility of humanized mice implies there are unlikely to be any specific essential PRDM9 binding sites."

4. Epigenetic Marking of Suspect Sequences

An Anti-Speciation Gene

From generation to generation, growing DNA sequence diversity between two intraspecies groups makes speciation more likely (Harvey et al. 2017). Conversely, anything that facilitates the transgenerational uniformity of genomic information should make speciation less likely. As mentioned above, in many quarters *prdm9* is considered a 'speciation gene,' indicating its direct *positive* involvement in the speciation process, such as by *increasing* the sequence diversity that can lead to hybrid sterility. Indeed, PRDM9 can work to induce sterility as shown in Fig. 3 (left side). However, concerning this positive involvement, Davies et al (2016) remark that "the rapid evolution of the zinc finger array of PRDM9 implies an unexpected transience of this direct role." In keeping with this, Fig. 3 (right side) raises the possibility that the development of sterility would be forestalled by zinc finger mutations (of which the humanized version is an extreme example) that would work to reduce variation and produce fertile offspring. The sterility could indeed be evolutionarily transient. Supporting this view, human PRDM9 zinc finger

mutations have been found to associate with *protection* against hybrid sterility (Irie et al 2009; Oliver et al 2009).

Thus, PRDM9 can be viewed as normally promoting recombination, hence providing opportunities for repairing mutations and so impeding the sequence divergence that can lead to speciation. It opposes the hybrid sterility that is "characterized by failure of pairing (synapsis) of homologous chromosomes and an arrested meiotic prophase owing to lack of repair of recombination intermediates" (Davies et al 2016). In this light, *prdm9* is best regarded as an *anti-speciation* gene in that it facilitates the correction of meiotic mismatches (heterozygosities). There is now diminishing support for the view that adverse genic interactions (epistasis), referred to as Dobzhansky-Muller incompatibilities, are as fundamental to speciation as meiotic chromosomal mismatches (Kao et al. 2010; Forsdyke 2011b).

A Tale of Two Asymmetries

Yet none of this easily explains the "hot spot paradox" (Boulton et al. 1997), namely the erosion of PRDM9 target sites when asymmetrically disposed between parental genomes (Fig. 2). For some reason the *prdm9* gene is on a treadmill being constantly 'obliged,' through mutation, to reinvent itself, so encoding a novel PRDM9 protein product that can summon up fresh, symmetrical, target sites. These sites then become epigenetically marked through histone methylation.

In another context, the epigenetic symmetry/asymmetry issue also emerged with a hypothetical mechanism for transgenerational error-correction through biased meiotic conversion (Reese and Forsdyke 2016; Forsdyke 2016, pp. 327-350). Here a means by which a repair process could achieve a gradual (multi-generational) bias against mutations was envisioned. When meiotic branch migration encounters a point mutation, a base pair mismatch arises and is repaired, so either the maternal or the paternal base is converted to the complement of its opposing base in the heteroduplex. In the absence of any reason to expect a systematic bias in this repair process, chance alone would seem to determine a mutation's fate.

However, it is supposed that where a meiotic strand exchange first encounters a base pair mismatch (indicating a likely mutation) *both* strands are epigenetically marked to designate

uncertainty. Such a mark if retained within the new organism's germ cell line will convey the message "there is a near 50% chance that this allele is a mutation." If this epigenetic tag could promote subsequent meiotic strand exchange, and invite the next meiotic partner (less likely to harbor a mutation) to serve as the donor strand in the subsequent mismatch repair, this could ensure that over many generations biased conversions will remove mutations. Such a process, by reducing the rate of mutation-driven genetic divergence within breeding populations, would act as a force against speciation.

Fig. 4 outlines the proposed tagging principle, which has some resemblance to the epigenetic tagging of damaged DNA in cell lines (Kafer et al. 2016). When meiotic conversion is *randomly* biased, a mutation in one of the parental genomes can either be repaired or compounded (Fig 4, outcomes 1 and 3). However, this *uncertainty* in the repair process can be registered by adding an epigenetic tag symmetrically to *both* strands in the repaired region (Fig 4, outcomes 2 and 4). Subsequently, the tagged gametes can be compared with normal gametes provided by partners from the main, non-mutant, population. It would be an 'educated bet' that the best direction of conversion should be from normal to mutant. So there could be a selection pressure for directing the conversion from untagged to tagged DNA strands. With PRDM9 the direction of conversion from donor to recipient is found experimentally to be from the strand that more strongly binds the protein to the weaker binding strand (Fig. 3). Could the experimental and hypothetical asymmetries, which both seem to oppose speciation, be related? Could the two asymmetries be functionally linked with retention of appropriate directionalities? Indeed, could one be the evolutionary raison d'être of the other?

PRDM9 Cannot Cover All Potholes

Given the great length of genome sequence in need of screening for errors, it would seem reasonable to suppose that recombination "hotspots might massively increase search efficiency by directing homology search to PRDM9 binding sites" (Davies et al. 2016). Given this supposed advantage, why should hotspots erode? Some 25,000 to 50,000 hotspots were detected in the human genome by Myers et al. (2005), perhaps extending to 80,000 (Khil and Camerini-Otero 2010).

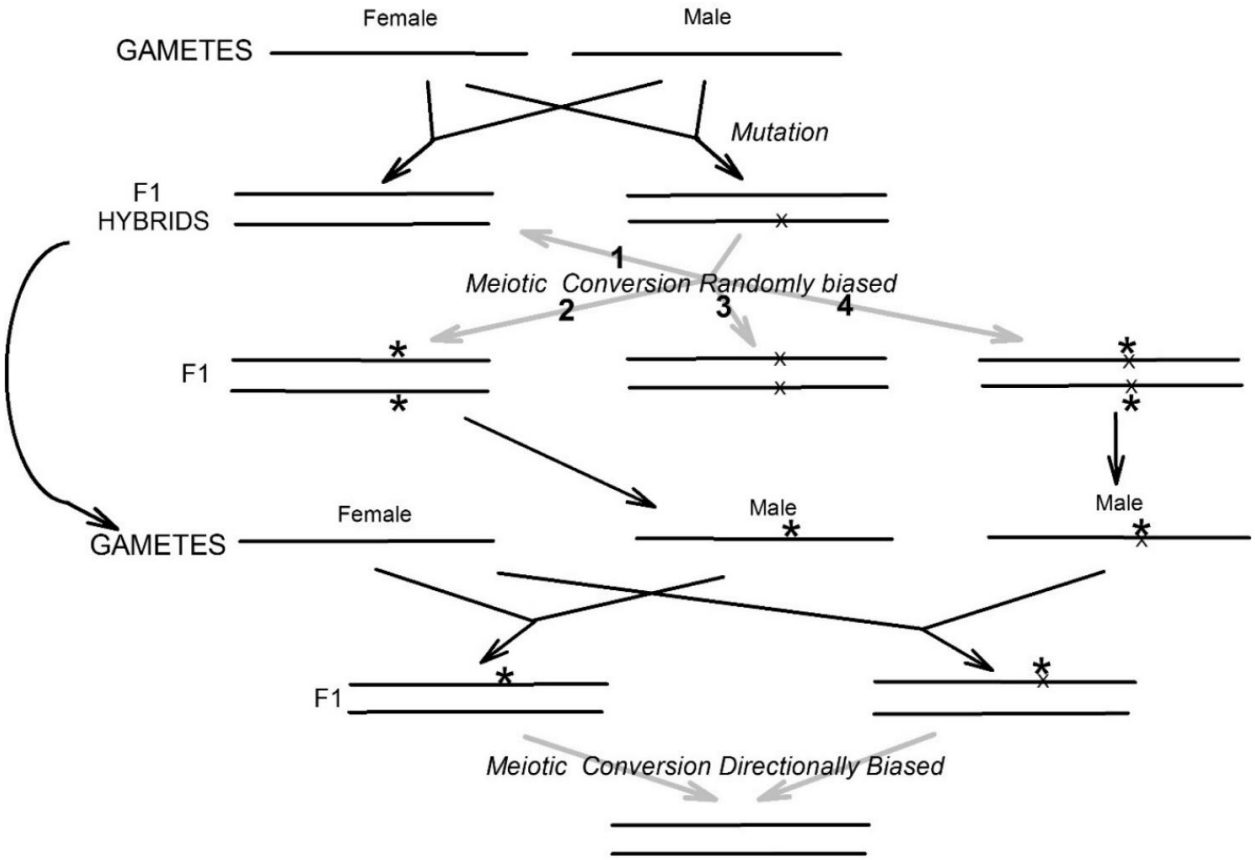

Fig. 4 Marking as "suspect" a sequence whose accuracy is in doubt facilitates remedy (albeit delayed) by directional conversion bias. Resulting from a normal cross (top left), healthy homozygous F1 hybrids are representative of the species, members of which produce normal gametes (curved arrow). However, a mutation (X) in one parent (top right) may produce a heterozygous F1 hybrid. If the mutation is such that mutant and non-mutant strands cannot be distinguished, meiotic conversion bias will occur randomly, with four possible outcomes (1, 3, 2 and 4). 1: Bias in favor of the unmutated strand *corrects* the mutation. 3: Bias in favor of the mutated strand *compounds* the mutation. 2 and 4: In the region of the heterozygosity, conversion is accompanied by epigenetic tagging as "suspect" (asterisks). This affects both parental strands, irrespective of whether the mutation has been corrected (2), or compounded (4). At middle, the tagged parents produce tagged gametes (the sex here is arbitrary). At bottom, crosses of tagged parents with members of the general population produce unilaterally tagged hybrids. Directionally biased meiotic conversion in favor of non-tagged DNA strands removes both the mutation (if present) and the tag.

However, only a small percentage of potential PRDM9 binding sites are used in any one meiotic cell (Balcova et al. 2016; Davies et al. 2016; supplementary information, section 3). Thus, at any point in time it is not the *number* of hotspots that is limiting: "Not all PRDM9 binding sites become hotspots, and the reasons for this remain unclear" (Altemose et al. 2017). Increasing PRDM9 dosage experimentally appears to increase hotspot usage, suggesting that the availability of PRDM9 can be rate-limiting (Davies et al. 2016; supplementary information, section 12). One unexplained implication of this is that an allelic pair of homologous hotspots could have PRDM9 molecules asymmetrically bound, not because of the erosion effect (Fig. 2), but because there was insufficient PRDM9. Recent work suggesting that their proper functioning requires aggregation (multimerization) of individual PRDM9 molecules, complicates this further (Altemose et al. 2017). Even if the PRDM9 concentration were sufficient, should heterologies (mismatches) not be present, the binding of various repair proteins (used in some hotspot assays) might not occur. So many problems, both theoretical and technical, remain.

However, another implication is that, under normal circumstances, widely dispersed, epigenetically tagged, heterologies (e.g. polymorphisms) may need to persist *for many generations* if they are to come under the surveillance of PRDM9. Failing their removal from the population by natural selection (or drift), the hotspot erosion process (Figs. 2) would serve to expose cryptic PRDM9 targets and hence, over the generations, bring different genomic segments under surveillance, so ultimately achieving repair (Fig. 3).

Apart from this there is the pot-hole repair analogy. If a town has roads with many pot-holes and few repair vans, it makes sense to direct those vans to the most pot-holed roads. Thus, simply speeding up the pace of genome-wide surveillance might not be enough. It would seem better if close-knit collectives of epigenetically marked suspect sites could somehow (i.e. by being designated as hotspots) be recognized as needing more urgent treatment than isolated sites. Consistent with this, Arbeithuber et al. (2015) note that "regions in close vicinity to these PRDM9 binding sites also showed a significant enrichment of polymorphisms in humans." While some polymorphisms are stable (presumably having escaped tagging), it seems possible that a local collective of polymorphisms that are enriched for epigenetic 'suspect' tags (Fig. 4), might somehow be able to attract some of the limited number of PRDM9 molecules to an

otherwise cryptic hotspot in its proximity (Fig. 5). A hypertagged region would influence hotspot placement. Alternatively, such a collective of tagged heterologies might be able to trigger local hotspot mutation to increase affinity for PRDM9, or send a long-range signal to the *prdm9* gene that it is time to mutate its zinc fingers in order to reconfigure its global control. Whatever the 'pothole' search mechanism, this underlines the value of focusing recombination to hotspots, rather than randomly.

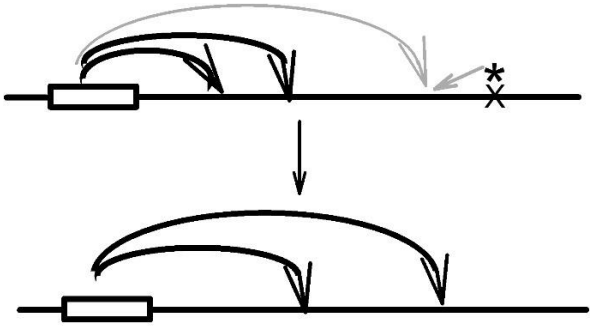

Fig. 5 Hypothetical recruitment of a fresh PRDM9 target site by a chromosome sequence that has been epigenetically marked (asterisk) to bias meiotic conversion. Details are as in Fig. 2. Top: The curved grey arrow indicates a presumptive target site with the same sequence as those that have been targeted in previous generations (curved black arrows). The small grey arrow indicates epigenetic sequence modification prompted by some aspect of the asterisk-marked region. Bottom: The fresh target site is recognized. Given that PRDM9 concentration is limiting, one of the previous target sites is vacated. The suspect sequence with a mutation (X) has been directionally converted and the tag has been eliminated.

Some of the multitude of recombination-related proteins so far discovered – most not mentioned here – could be involved. Noting that increase in PRDM9 dosage only "partially rescues hybrid sterility of PWD x B6 F1 males," Balcova et al. (2016) distinguish the *identification* of hotspot targets from the *rate* of events subsequent to that identification, which implicate the *hstx2* gene. Their results are interpreted as "strongly indicating an independent control of global crossover rate variation and genomic crossover placement." Given the current pace of research in this area, a role for the *hstx2* gene product could soon emerge.

Conclusions

Speciation is a possible outcome of increasing variation between individuals in a species. While variation can be diminished, usually in local genome regions, by selective sweeps, there are numerous internal mechanisms operating genome-wide to maintain DNA sequence integrity (e.g. Bernstein and Bernstein 1991). This could account for the "mysterious" selective pressure on *prdm9*. The quicker *prdm9* can mutate to facilitate symmetrical recognition of new hotspot sequences, the quicker and more comprehensive can be the repair of heterozygosities, and the less likely the triggering of a speciation event. Since this now appears as the main role of PRDM9, then *prdm9* is best regarded as an anti-speciation gene. The case for this, which includes explaining the hotspot conversion paradox, is made here in the context of a chromosomal basis for hybrid sterility. To this extent, the chromosomal hypothesis is supported over genic alternatives as governing the initiation of speciation in the general case. Of course, numerous genes participate in the orchestration of meiosis and malfunction of only one can produce hybrid sterility (Forsdyke 2010). However, in general, it is malfunction of the most distinctive feature of meiosis – the sequence-dependent pairing of homologous chromosomes – that is likely to be decisive in sparking divergence into new species.

Acknowledgements Virgil Reese provided valuable advice. Queen's University hosts my web pages that display items from the early speciation literature (http://post.queensu.ca/~forsdyke/evolutio.htm or https://archive-it.org/collections/7641).

References

Altemose N, Noor N, Bitoun E, Tumian A, Imbeault M, Chapman JR, Aricescu AR, Myers SR (2017) Human PRDM9 can bind and activate promoters, and other zinc-finger proteins associate with reduced recombination in *cis*. BioRxiv preprint http://dx.doi.org/10.1101/144295

Arbeithuber B, Betancourt AJ, Ebner T, Tiemann-Boege I (2015) Crossovers are associated with mutation and biased gene conversion at recombination hotspots. Proc Natl Acad Sci USA 112: 2109–2114

Balcova M, Faltusova B, Gergelits V, Bhattacharyya T, Mihola O, Trachtulec Z et al (2016) Hybrid sterility locus on chromosome X controls meiotic recombination rate in mouse. PLoS Genet 12:e1005906

Bernstein C, Bernstein H (1991) Aging, sex and DNA repair. Academic Press, San Diego.

Bhattacharyya T, Gregorova S, Mihola O, Anger M, Sebestova J, Denny P et al (2013) Mechanistic basis of infertility of mouse intersubspecific hybrids. Proc Nat Acad Sci USA 110: E468–E477

Bhattacharyya T, Reifova R, Gregorova S, Simecek P, Gergelits V, Mistrik M et al (2014) X chromosome control of meiotic chromosome synapsis in mouse inter-subspecific hybrids. PLoS Genet 10:e1004088

Boateng KA, Bellani MA, Gregoretti IV, Pratto F, Camerini-Otero RD (2013) Homologous pairing preceding SPO11-mediated double-strand breaks in mice. Dev Cell 24:196–1205

Boulton A, Myers RS, Redfield RJ (1997) The hotspot conversion paradox and the evolution of meiotic recombination. Proc Natl Acad Sci USA 94:8058–8063

Brick K, Smagulova F, Khil P, Camerini-Otero RD, Petukhova GV (2012) Genetic recombination is directed away from functional genomic elements in mice. Nature 485:642–645

Cox EC, Yanofsky C (1969) Mutator gene studies in *Escherichia coli*. J Bacteriol 100:390–397

Coyne JA, Orr HA (2004) Speciation. Sinauer, Sunderland, MA

Danilowicz C, Lee CH, Kim K, Hatch K, Coljee VW, Kleckner N, Prentiss M (2009) Single molecule detection of direct, homologous, DNA/DNA pairing. Proc Natl Acad Sci USA 106: 19824–19829

Davies B, Hatton E, Altemose N, Hussin JG, Pratto F et al (2016) Re-engineering the zinc fingers of PRDM9 reverses hybrid sterility in mice. Nature 530:171–176

Delph LF, Demuth JP (2016) Haldane's rule: genetic bases and their empirical support. J Hered 107:383–391

Flachs P, Mihola O, Šimeček P, Gregorová S, Schimenti JC, et al (2012) Interallelic and intergenic incompatibilities of the *Prdm9* (*Hst1*) gene in mouse hybrid sterility. PLoS Genet 8: e1003044

Forsdyke DR (1996) Different biological species "broadcast" their DNAs at different (G+C)% "wavelengths". J Theor Biol 178:405–417

Forsdyke DR (2000) Haldane's rule: hybrid sterility affects the heterogametic sex first because sexual differentiation is on the path to species differentiation. J Theor Biol 204:443–452

Forsdyke DR (2001) The origin of species, revisited. McGill-Queen's University Press, Montreal

Forsdyke DR (2004) Chromosomal speciation: a reply. J Theor Biol 230:189–196

Forsdyke DR (2007) Molecular sex: the importance of base composition rather than homology when nucleic acids hybridize. J Theor Biol 249:325–330

Forsdyke DR (2010) George Romanes, William Bateson, and Darwin's "weak point." Notes Rec RS 64**:**139–154

Forsdyke DR (2011a) The selfish gene revisited: reconciliation of Williams-Dawkins and conventional definitions. Biol Theor 5:246-255

Forsdyke DR (2011b) The B in BDM. William Bateson did not advocate a genic speciation theory. Heredity 106:202

Forsdyke DR (2016) Evolutionary Bioinformatics, 3rd edn. Springer, New York

Forsdyke DR (2017a) Speciation: Goldschmidt's chromosomal heresy, once supported by Gould and Dawkins, is again reinstated. Biol Theor 12:4–12

Forsdyke DR (2017b) Base composition, speciation, and why the mitochondrial barcode precisely classifies. Biol Theor 12: (in press) doi: 10.1007/s13752-017-0267-5

Gladyshev E, Kleckner N (2014) Direct recognition of homology between double helices of DNA in *Neurospora crassa*. Nat Commun 5:3509

Gladyshev E, Kleckner N (2016) Recombination-independent recognition of DNA homology for repeat-induced point mutation (RIP) is modulated by the underlying nucleotide sequence. PLoS Genet 12:e1006015

Gladyshev E, Kleckner N (2017) Recombination-independent recognition of DNA homology for repeat-induced point mutation. Curr Genet 63:389–400

Gorelick R, Heng HHQ (2010) Sex reduces genetic variation: a multidisciplinary review. Evolution 65:1088–1098

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

Haldane JBS (1922) Sex ratio and unisexual sterility in hybrid animals. J Genet 12:101–109

Harvey MG, Seeholzer GF, Smith BT, Rabosky DL, Cuervo AM, Brumfield RT (2017) Positive association between population genetic differentiation and speciation rates in new world birds. Proc Natl Acad Sci USA 114:6328–6333

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

Hu X-S, Filatov DA (2016) The large-X effect in plants: increased species divergence and reduced gene flow on the Silene X-chromosome. Mol Ecol 25:2609–2619

Irie S, Tsujimura A, Miyagawa Y, Ueda T, Matsuoka Y, et al (2009) Single nucleotide polymorphisms in *PRDM9* (*MEISETZ*) in patients with nonobstructive azoospermia. J Androl 30:426–431

Ironside JE, Filatov DA (2005) Extreme population structure and high interspecific divergence of the Silene Y chromosome. Genetics 171:705–713

Johannesson K (2010) Are we analyzing speciation without prejudice? Ann N Y Acad Sci 1206:143–149

Kafer GR, Li X, Horii T, Suetake I, Tajima S, Hatada I, Carlton PM (2016) 5-Hydroxymethylcytosine marks sites of DNA damage and promotes genome stability. Cell Rep 14:1283–1292

Kao KC, Schwartz K, Sherlock G (2010) A genome-wide analysis reveals no nuclear Dobzhansky-Muller pairs of determinants of speciation between *S. cerevisiae* and *S. paradoxus*, but suggests more complex incompatibilities. PLoS Genet 6:e1001038

Khil PP, Camerini-Otero RD (2010) Genetic crossovers are predicted accurately by the computed human recombination map. PLoS Genet 6:e1000831

Kliman RM, Rogers BT, Noor MAF (2001) Differences in (G + C) content between species: a commentary on Forsdyke's "chromosomal viewpoint" of speciation. J Theor Biol 209:131–140

Lao PJ, Forsdyke DR (2000) Crossover hotspot instigator (Chi) sequences in *E. coli* occupy distinct recombination/transcription islands. Gene 243:47–57

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 Genet 10:e1004790

Louis EJ (2009) Origins of reproductive isolation. Nature 457:549–560

Mallet J (2006) What does *Drosophila* genetics tell us about speciation? Trends Ecol Evol 21:386–393

McGavin S (1977) A model for the specific pairing of homologous double-stranded nucleic acid molecules during genetic recombination. Heredity (Edinb) 39:15–25

Mérot C, Salazar C, Merrill RM, Jiggins CD, Joron M (2017) What shapes the continuum of reproductive isolation? Lessons from *Heliconius* butterflies. Proc R Soc B 284:20170335

Mihola O, Trachtulec Z, Vicek C, Schimenti JC, Forejt J (2009) A mouse speciation gene encodes a meiotic histone H3 methyltransferase. Science 323:373–375

Moehring AJ (2011) Heterozygosity and its unexpected correlations with hybrid sterility. Evolution 65:2621–2630

Montgomery TJ (1901) A study of the chromosomes of the germ cells of metazoa, Trans Am Phil Soc 20:154–236

Myers S, Bottolo L, Freeman C, McVean G, Donnelly P (2005) A fine scale map of recombination rates and hotspots across the human genome. Science 310:321–324

Naviera HF, Maside XR (1998) The genetics of hybrid male sterility in *Drosophila*. In: Endless forms: species and speciation. (Howard DJ, Berlocher SH, eds.), Oxford University Press, Oxford, pp. 330–338

Nei M, Nozawa M (2011) Roles of mutation and selection in speciation: from Hugo de Vries to the modern genomic era. Gen Biol Evol 3:812–829

Nevo E (2012) Speciation: chromosomal mechanisms. Encyclopedia of Life Sciences**.** John Wiley, Chichester, doi: 10.1002/9780470015902.a0001757.pub3

Oliver PL, Goodstadt L, Bayes JJ, Birtle Z, Roach KC, et al. (2009) Accelerated evolution of the *Prdm9* speciation gene across diverse metazoan taxa. PLoS Genet 5:e1000753

Payseur BA (2016) Genetic links between recombination and speciation. PLoS Genet 12:e1006066

Paz A, Kirzhner V, Nevo E, Korol A (2006) Coevolution of DNA-interacting proteins and genome 'dialect.' Mol Biol Evol 23:56–64

Poole AM, Logan DT (2005) Modern mRNA proofreading and repair: clues that the last universal common ancestor possessed an RNA genome. Mol Biol Evol 22:1444–1455

Reanney DC (1979) RNA splicing and polynucleotide evolution. Nature 277:598–600

Reese VR, Forsdyke DR (2016) Meiotic pairing inadequacies at the levels of X chromosome, gene, or base: Epigenetic tagging for transgenerational error-correction guided by a future homologous duplex. Biol Theor 11:150–157

Schartl M (2008) Evolution of *Xmrk*: an oncogene, but also a speciation gene? BioEssays 30:822–832

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

Szostak JW, Orr-Weaver TL, Rothstein RJ, Stahl FW (1983) The double-strand-break repair model for recombination. Cell 33:25–35

Wahls WP, Davidson MK (2011) DNA sequence-mediated, evolutionarily rapid, redistribution of meiotic recombination hotspots. Genetics 189:685–694

Wilson JH (1979) Nick-free formation of reciprocal heteroduplexes: a simple solution to the topological problem. Proc Natl Acad Sci USA 76:3641–3645

Winge Ö (1917) The chromosomes, their number and general importance. Compt Rend Trav Lab Carlsberg 13:131–275

Zickler D, Kleckner N (2015) Recombination, pairing, and synapsis of homologs during meiosis. Cold Spring Harb Pers Biol 7(6). pii: a016626.