

RESEARCH ARTICLE

Open Access



# Ruminant-specific multiple duplication events of PRDM9 before speciation

Abinash Padhi<sup>1\*</sup>, Botong Shen<sup>1</sup>, Jicai Jiang<sup>1</sup>, Yang Zhou<sup>2,3</sup>, George E. Liu<sup>2</sup> and Li Ma<sup>1\*</sup>

## Abstract

**Background:** Understanding the genetic and evolutionary mechanisms of speciation genes in sexually reproducing organisms would provide important insights into mammalian reproduction and fitness. PRDM9, a widely known speciation gene, has recently gained attention for its important role in meiotic recombination and hybrid incompatibility. Despite the fact that PRDM9 is a key regulator of recombination and plays a dominant role in hybrid incompatibility, little is known about the underlying genetic and evolutionary mechanisms that generated multiple copies of PRDM9 in many metazoan lineages.

**Results:** The present study reports (1) evidence of ruminant-specific multiple gene duplication events, which likely have had occurred after the ancestral ruminant population diverged from its most recent common ancestor and before the ruminant speciation events, (2) presence of three copies of PRDM9, one copy (lineages I) in chromosome 1 (chr1) and two copies (lineages II & III) in chromosome X (chrX), thus indicating the possibility of ancient inter- and intra-chromosomal unequal crossing over and gene conversion events, (3) while lineages I and II are characterized by the presence of variable tandemly repeated C2H2 zinc finger (ZF) arrays, lineage III lost these arrays, and (4) C2H2 ZFs of lineages I and II, particularly the amino acid residues located at positions -1, 3, and 6 have evolved under strong positive selection.

**Conclusions:** Our results demonstrated two gene duplication events of PRDM9 in ruminants: an inter-chromosomal duplication that occurred between chr1 and chrX, and an intra-chromosomal X-linked duplication, which resulted in two additional copies of PRDM9 in ruminants. The observation of such duplication between chrX and chr1 is rare and may possibly have happened due to unequal crossing-over millions of years ago when sex chromosomes were independently derived from a pair of ancestral autosomes. Two copies (lineages I & II) are characterized by the presence of variable sized tandem-repeated C2H2 ZFs and evolved under strong positive selection and concerted evolution, supporting the notion of well-established Red Queen hypothesis. Collectively, gene duplication, concerted evolution, and positive selection are the likely driving forces for the expansion of ruminant PRDM9 sub-family.

**Keywords:** Speciation genetics, Genetic incompatibility, PRDM9, Gene duplication, Positive selection, Molecular evolution, Ruminants

## Background

Ever since the theory of genetic incompatibility (Bateson-Dobzhansky-Muller Model) was independently proposed by three eminent evolutionary biologists [1–3], researchers across the disciplines have been devoted to characterizing the evolutionary impacts of reproduction-associated genes on speciation and species diversity. Understanding the molecular diversity of speciation genes

would unravel the underlying mechanisms by which species diversity drives speciation and the latitudinal gradient of taxonomic groups as species diversity decreases with latitude [4, 5]. Further, in-depth understanding of the genetic and evolutionary mechanisms of speciation genes would not only provide important insights into an organism's fitness and/or reproduction but also promote conservation of threatened mammalian species through genetic re-engineering, a technique that has recently been used to reverse hybrid sterility in mice by editing the zinc fingers (ZFs) of a widely known speciation gene, PRDM9 [6]. This landmark experiment further signified the

\* Correspondence: bshen218@umd.edu; lima@umd.edu

<sup>1</sup>Department of Animal and Avian Sciences, University of Maryland, College Park, MD 20742, USA

Full list of author information is available at the end of the article



important role of PRDM9 in fertility and reproductive compatibility [6]. Nevertheless, the reports of genome-wide non-random distributions of DNA binding motifs and the corresponding clustering of meiotic recombination hotspots, together with the Red Queen model of evolution of these DNA-binding motifs provide convincing evidence of the dominant role of PRDM9 in metazoan speciation [7–25]. Red Queen Hypothesis, which is based on the metaphors in Lewis Carroll's "Through the Looking Glass" [26], was first used by VanValen [27] to explain speciation dynamics and extinction of species. Since then this metaphor has been widely used as the key hypothesis to test the continual adaptation of species in order to survive in the face of competition and changing environment, including the evolution of ZFs of PRDM9 by treating PRDM9 ZFs as "species" and genome background as "environment" [16, 25]. Nevertheless, the absence of functional PRDM9 in canids [28–30] and presence of single copies of PRDM9 in rodents but multiple copies (i.e., PRDM 7/9) in primates, ruminants and other metazoan lineages [31–33] indicate an interesting yet complex evolutionary history of the PRDM9 gene family.

PRDM9 has been reported to play a dominant role in meiotic recombination in a wide range of mammalian groups [8–10, 13–18, 20, 21, 23, 34–37]. It is a member of the PRDM gene family [33] and encodes a protein with a KRAB, a SSXRD, a PR/SET histone H3(K4) trimethyl transferase domain and a DNA-binding domain consisting of a variable-sized tandemly repeated array of C2H2 ZFs at the C-terminal [18]. The C-terminal array of the C2H2 ZFs domain possesses a DNA-binding function, shows a high diversity and fast evolutionary rate, and hence is likely to have evolved extremely rapidly by positive Darwinian selection [16, 21, 25, 38–40]. However, the N-terminal KRAB, SSXRD and SET domains have evolved at a very slow rate [18], thus making it an ideal genetic marker to trace the evolutionary history of PRDM9 in each metazoan lineage.

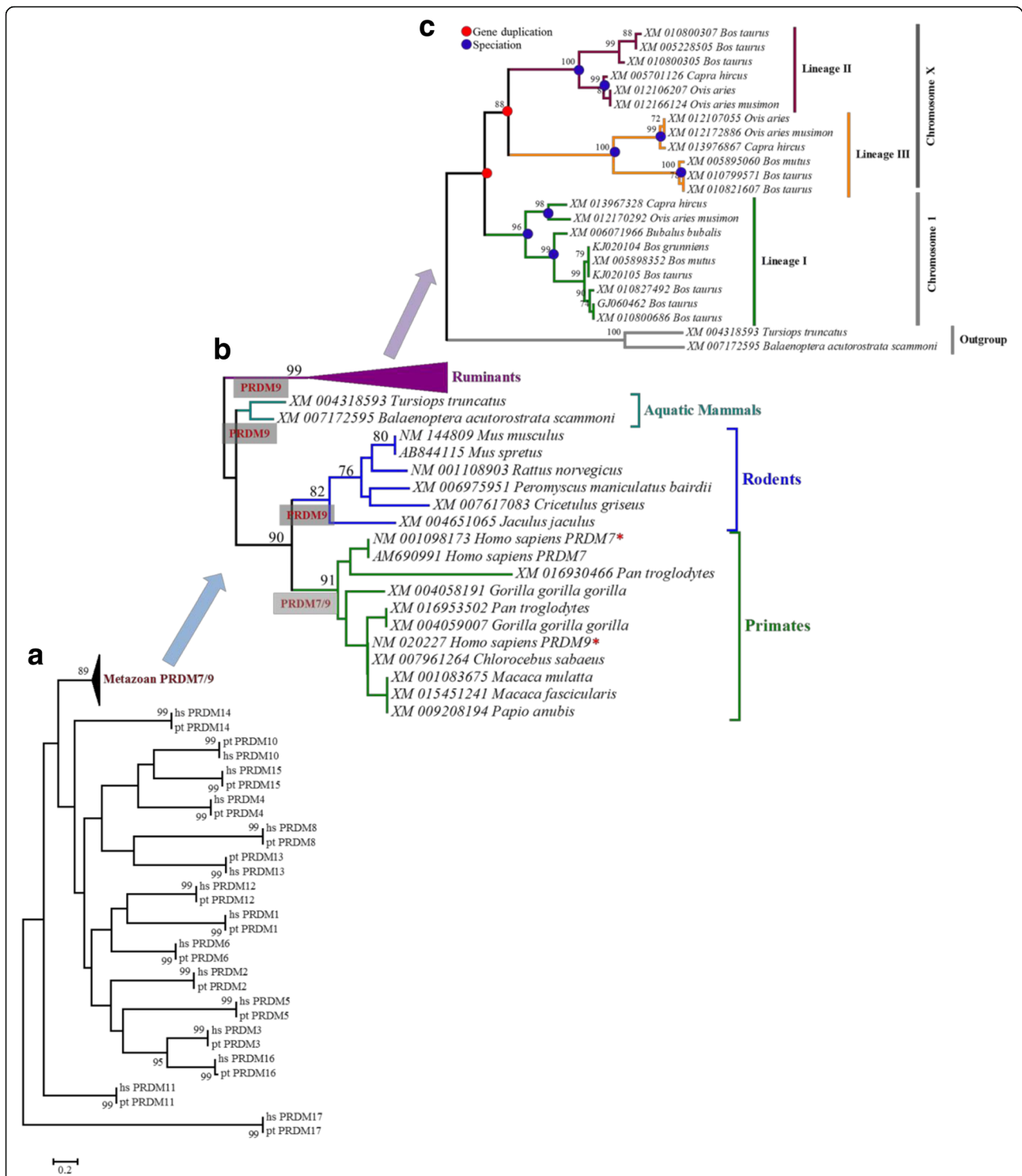
Despite the critical role of the PRDM gene family in early development and reproduction [41] little is known about the evolutionary history of these genes. Two recent studies [31, 33] reported the evolution of PRDM gene family and suggested that while primate PRDM9 has a higher similarity of gene structure and protein domain organization with the non-primate co-orthologs and likely retains the features of the ancestral locus, PRDM7 appears to be primate-specific and may have undergone major structural arrangements that decreased the number of ZFs [31]. Vervoort et al. [33] reported that PRDM7 and PRDM9 gene trees do not form separate monophyletic groups and these gene trees are highly incongruent with the species tree, suggesting an unusual evolution of these genes in primates. Further, those studies concluded that PRDM7/9 phylogenetic analysis may

be unreliable for positioning the duplication events that have occurred in the primate lineage [33]. Given such unusual evolutionary patterns of PRDM7/9, in particular a non-monophyletic grouping of PRDM9 and PRDM7 in primates [31, 33], one might speculate that PRDM9 and PRDM7 have evolved independently in different metazoan lineages. Therefore, it is unclear if these form monophyletic groups in other metazoan, and we might need to revise the nomenclature of these gene copies.

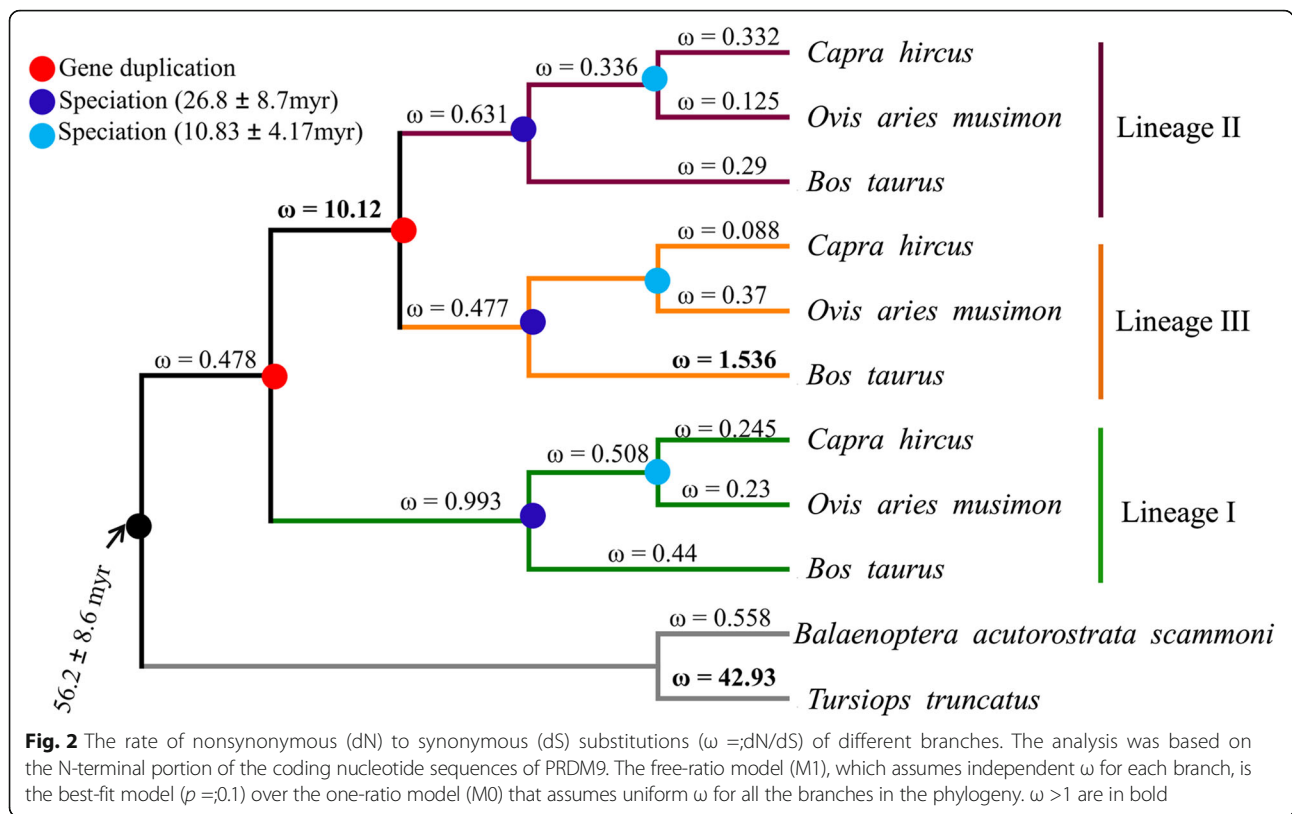
Utilizing the N-terminal portion of the PRDM9 nucleotide and protein sequences the objective of this study is to investigate the origin and evolution of the multiple copies of PRDM9 in ruminants, to determine the phylogenetic congruencies of gene trees from these novel gene copies with the ruminant species tree, and to assess the underlying genetic and evolutionary forces that shaped the evolution of these gene copies in ruminants. Furthermore, given the fact that each functional domain of the PRDM9 gene is associated with different functions [18], these functional domains are expected to show different evolutionary trajectories. Thus, another objective of this study is to unravel the different evolutionary forces that shape the evolution of N-terminal and that are responsible for a variable-sized tandem-repeat array of C2H2 ZFs at the C-terminal in each lineage. Finally, we propose a model that explains the evolution of PRDM9 and its multiple copies in the ruminant species.

## Results

We first give an overview of the main results and then provide more detailed explorations in the following paragraphs. The present study reports (1) evidence of ruminant-specific multiple gene duplication events which likely have had occurred before the ruminant speciation events and after the ancestral ruminant population diverged from its most recent common ancestor (Figs. 1 and 2), (2) the presence of three copies of PRDM9 (Figs. 1 and 2), two copies (lineage II and III; Fig. 1) in chromosome X (chrX) and one copy (lineage I; Fig. 1) in chromosome 1 (chr1) with variable-sized tandemly repeated arrays of C2H2 ZFs at the C-terminal (Fig. 3) thus indicating the possibility of ancient inter- and intra-chromosomal unequal crossing over and gene conversion events, (3) while lineages I and II are characterized by the presence of variable tandemly repeated C2H2 ZFs arrays, lineage III lost these arrays (Fig. 3), (4) C2H2 ZFs of lineages I and II, particularly amino acid residues located at positions -1, 3, and 6 have likely evolved under strong positive selection (Fig. 4; Table 1) thus supporting the notion of previously established Red Queen hypothesis [16, 25], and finally, (5) together with the evidence of positive selection (Fig. 4 Table 1) relatively higher diversities at the nonsynonymous sites (Fig. 5) the presence of identical arrays yet located at different



**Fig. 1** Phylogenetic trees inferred from PRDM genes. **a** Phylogenetic tree inferred from the SET domain amino acid sequences (alignment length: 203 aa) of the PRDM genes. **b** PRDM7/9 gene tree inferred from the SET domain (alignment length: 118 aa). **c** Phylogenetic tree inferred from the N-terminal portion of the amino acid sequences (alignment length: 351 aa) of PRDM9 depicting duplication and speciation events in ruminants. Bootstrap values greater than 70 are shown at the base of the nodes. GenBank accession numbers and scientific names of the species are shown. SET domain sequences representing all the 17 PRDM genes that were previously reported by Fumasoni et al. [31] were used as reference sequences. Asterisks (\*) indicate reference sequences for human PRDM7 [52] and human PRDM9 [9, 10]. *hs*: *Homo sapiens*; *pt*: *Pan troglodytes*

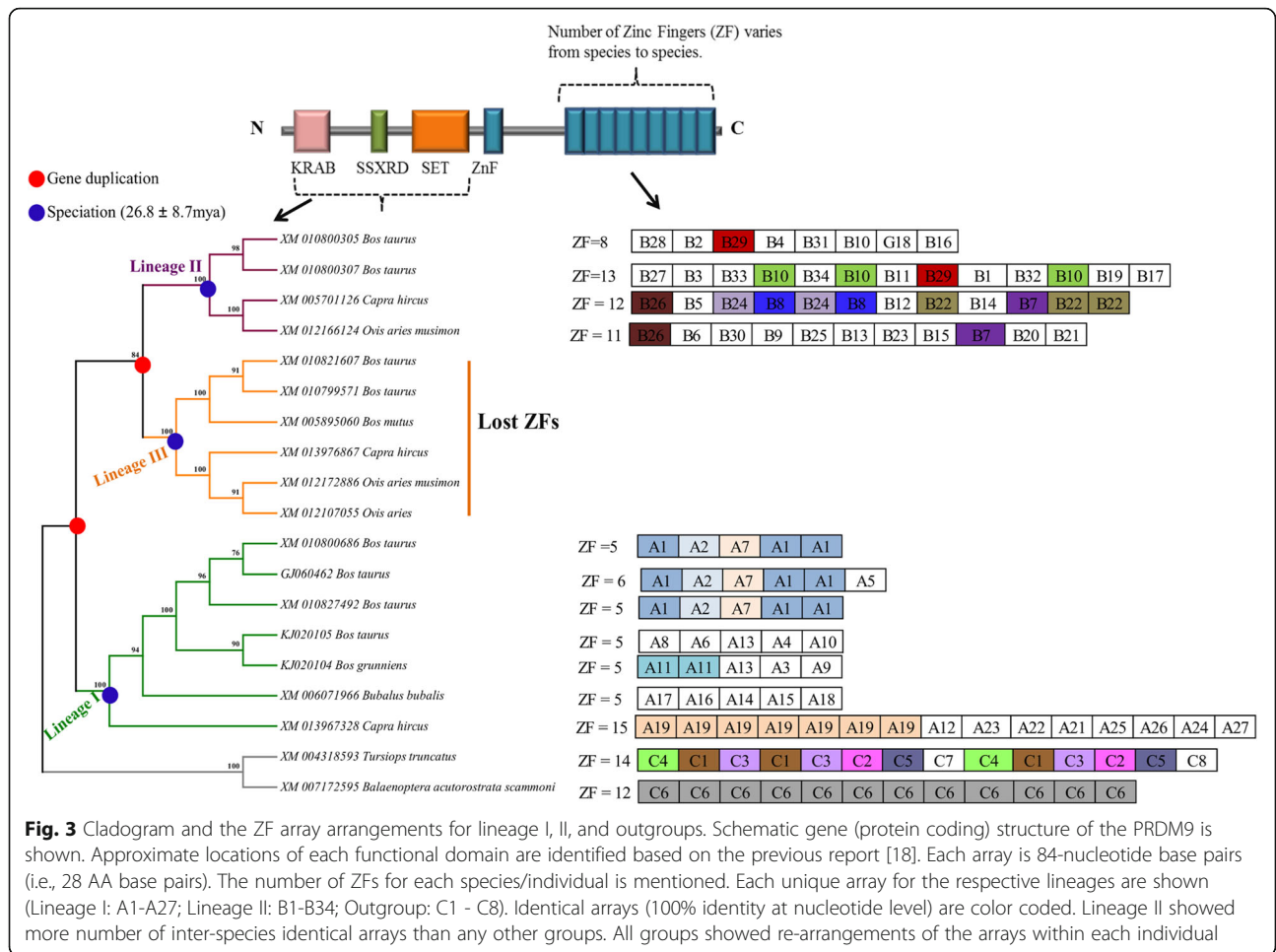


alignment positions in the sister-species (Fig. 3) as well as the observation of variable length of binding motifs for each ruminant species (Fig. 6) support both the concerted evolution [16] and a cyclical back-and-forth evolution of C2H2 ZFs arrays throughout the ruminant evolution spanning millions of years regardless of positive frequency-dependent or negative frequency-dependent selection, a dynamic evolutionary pattern that was recently proposed for host-parasite co-evolution [42, 43].

To evaluate the phylogenetic positioning (Fig. 1a) and clustering (Fig. 1b) of PRDM7 and PRDM9 in the PRDM7/9 gene tree and to assess the evolutionary origin of multiple copies of PRDM9 in ruminants (Fig. 1c) we reconstructed the phylogenetic trees using the amino acid sequences of the PR domains located at the N-terminal region (Fig. 1a b, and c). Consistent with previous studies [31, 33], our analyses revealed that PRDM7/9 form unique clusters (Fig. 1a) and that PRDM7 is primate-specific (Fig. 1b). The N-terminal amino acid sequence-based phylogeny showed that each PRDM9 copy (lineage I-III) of ruminants formed a separate monophyletic group and showed the evidence of two gene duplication events prior to the ruminant speciation (Fig. 1c). Multiple paralog copies of PRDM9 in ruminants (e.g., genus: *Bos*, *Capra*, and *Ovis*) support gene duplications before the speciation events. Based on the previous reports [44–47], these three species (genus: *Bos*, *Capra*, and *Ovis*) had a shared

ancestry. *Bos* diverged from the common ancestral population approximately 26.8 ( $\pm 8.7$ ) million years ago (mya), and the split between *Capra* and *Ovis* was estimated to be approximately 10.83 ( $\pm 4.17$ ) mya. Concurrently, the presence of all three PRDM9 copies in each species provides strong evidence of the gene duplication events before ruminant speciation (i.e., 26.8  $\pm$  8.7 mya) (Fig. 2).

One of the striking observations is the presence of two copies of PRDM9 (lineage II and III) on chrX (Fig. 1). While one X-linked copy is characterized by the presence of variable-sized tandemly repeated C2H2 ZFs (lineage II) the other copy completely lost its ZFs (lineage III). Interestingly, the dN/dS ratio ( $\omega$ ), for the branch leading to X-linked lineages (i.e., II and III) was estimated to be 10.12, indicating the evidence of positive selection, a typical characteristic of novel gene copies after a duplication event [48, 49]. We also found that the C-terminal C2H2 ZFs of lineages I and II and the outgroup, especially the amino acid residues at the positions -1, 3 and 6 that played crucial roles in DNA binding during meiotic recombination [16], have likely evolved under strong positive selection (Fig. 4). The tandemly repeated arrangement of the ZFs and the presence of identical ZFs (for example, in lineage I: A1, A7, A11 and in lineage II: B7, B8, B10, B22, B26, and B29) showed evidence of concerted evolution of C2H2 ZFs of both X-linked (lineage II) and autosomal (lineages I) PRDM9

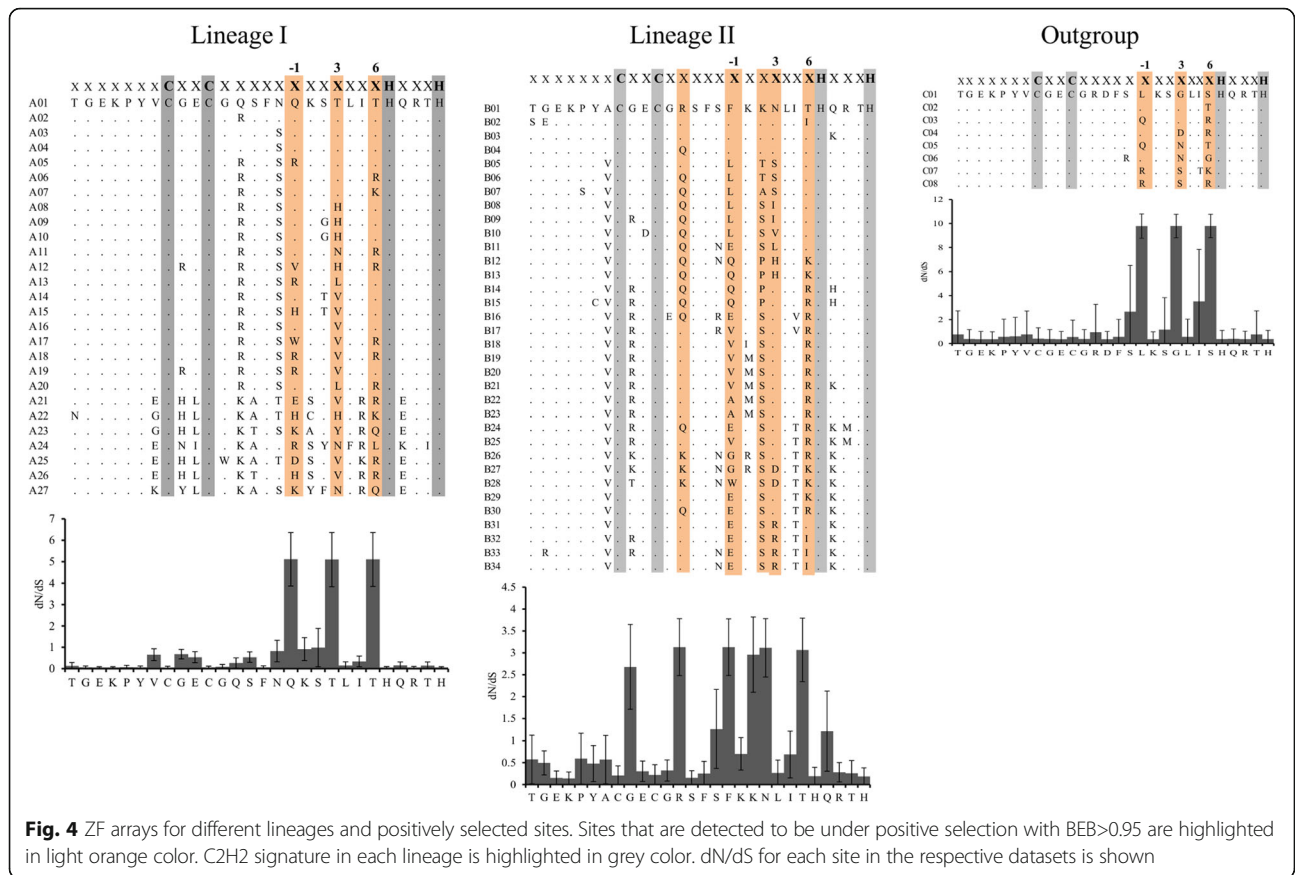


copies (Fig. 3). Further, we observed species-specific, lineage-specific, and individual-level variations of the length of tandemly repeated C2H2 ZFs as well as variations in the predicted binding motifs (Fig. 6). Finally, taking all the possible evolutionary forces (e.g., concerted evolution, gene duplications, and positive selection) that likely affected the evolution of PRDM9 and maintained genetic variations even at the individual levels in these economically important ruminant species (genus: *Bos*, *Capra*, and *Ovis*) into consideration, we presented a schematic model to describe how the multiple copies of PRDM9 are derived and evolved in the ruminant species (Fig. 7).

**Discussion**

Despite the fact that PRDM9 is a key regulator of meiotic recombination [7–18, 20, 21, 34, 35, 37, 50, 51] and plays a dominant role in hybrid incompatibility [6], little is known about the underlying genetic and evolutionary mechanisms that generated multiple copies of PRDM9 in many metazoan lineages. The present study elucidates the underlying evolutionary genetic mechanisms that

shaped the evolution of PRDM9, an important speciation gene [16, 18], in the economically important ruminants species (genus *Bos*, *Capra*, and *Ovis*). These domesticated ruminants are estimated to have diverged from a common ancestor approximately 26.8 (±8.7) mya [44–47]. In contrast to primate’s PRDM7 and PRDM9 gene copies that form non-separate monophyletic groups and show ambiguities concerning the phylogenetic positioning of the gene duplication events in the primate phylogeny [33], the observation of deep-split among the three lineages together with a strong statistical support for monophyletic groups provide convincing evidence of two gene duplication events before the ruminant speciation. Taken together with the results of a previous study [33], our study suggests that the PRDM9 duplication event in ruminants, which is estimated to have had occurred sometime between 27 and 56 mya, is ruminant-specific and likely occurred after the split of the ruminants ancestral populations from the most recent common ancestor. Nevertheless, based on these results, one might speculate that PRDM9 of other mammalian lineages may also exhibit unique phylogenetic histories. Further, together with the



results of a previous study [33], we ascertained that the primate-specific PRDM7 [31, 33, 52] is not phylogenetically closely related with the novel copies of ruminant PRDM9, therefore, warrants separate nomenclature of PRDM9 copies belonging to lineage II and III.

Although gene duplication events through inter-chromosomal especially, autosomal crossing-overs are common across the mammalian groups [53], the observations of gene duplications between sex chromosomes and autosomes is a unique event. Interestingly, a previous study has also reported inter-chromosomal duplications of the adrenoleukodystrophy (ADL) locus from chrX to chromosomes 2p11, 10p11, 16p11 and 22q11 in humans [54]. However, to our knowledge, so far no such

inter-chromosomal duplications between chrX and autosomes have been reported for any other mammalian taxa. We previously found a strong association between PRDM9 on chr1 and recombination phenotypes in cattle [13]. Sandor et al. [19] have also reported the presence of an X-linked PRDM9 and have detected several polymorphisms in the corresponding C2H2 ZFs. Although PRDM9 is present on both chr1 and chrX in cattle, the genetic and evolutionary mechanisms of the evolution of PRDM9 on the two chromosomes remain unclear. The presence of X-linked PRDM9 copies in ruminants could possibly be a rare event and be explained by some unique evolutionary mechanisms. Sex chromosomes were derived from a pair of ancestral autosomes [55]

**Table 1** Tests for positive selection for the ZFs of each lineage

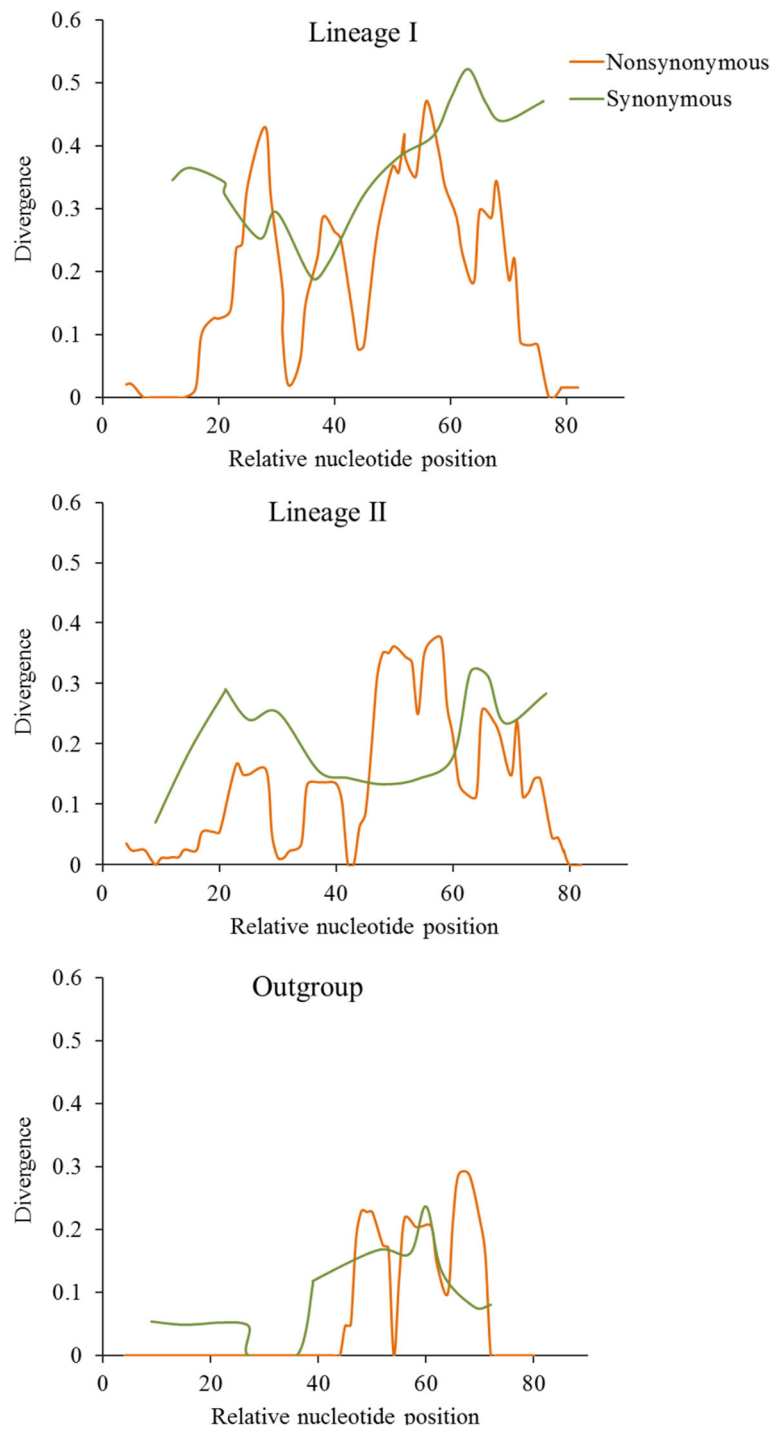
| Model Comparison | Lineage I |          | Lineage II |         | Outgroup |          |
|------------------|-----------|----------|------------|---------|----------|----------|
|                  | 2ΔI       | p        | 2ΔI        | p       | 2ΔI      | p        |
| M1a vs M2a       | 20.1377   | 0.000042 | 13.8381    | 0.00099 | 22.9932  | 0.000010 |
| M7 vs M8         | 20.4410   | 0.000036 | 17.5405    | 0.00016 | 23.1771  | 0.000009 |
| M8 vs M8a        | 18.3856   | 0.000018 | 12.9539    | 0.00032 | 22.9906  | 0.000002 |

Neutral (null) models: M1a, M7, M8a

Selection (alternative) models: M2a, M8

Degrees of freedom: 2 for M1a-M2a, M7-M8; and 1 for M8-M8a

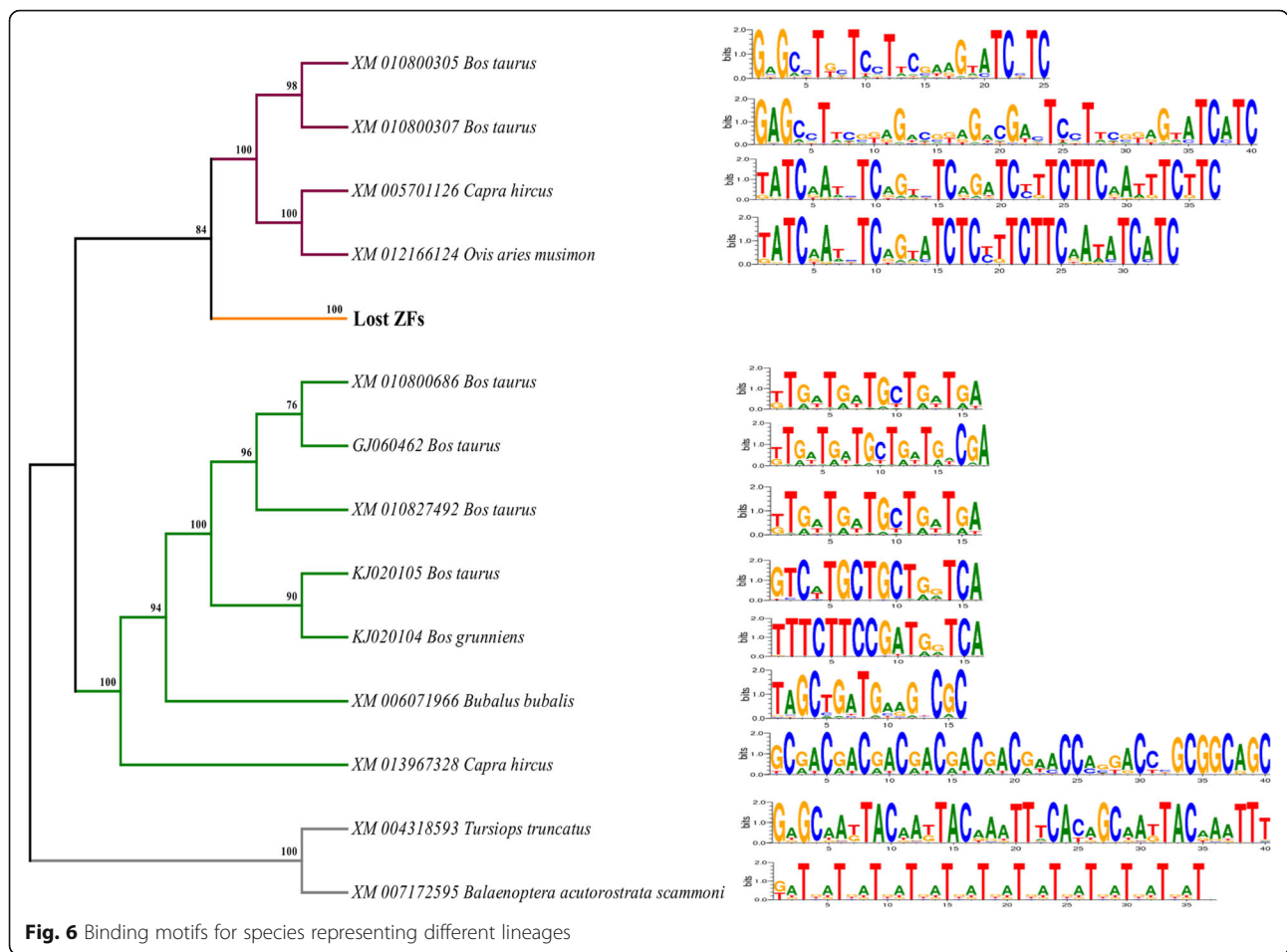
ΔI: Differences between the likelihood scores of null and alternative models



**Fig. 5** Divergence at the synonymous and nonsynonymous sites. Sliding Window Analyses (SWA) showing the divergence at the synonymous and nonsynonymous sites in the ZF array that comprises 84 nucleotide base pairs for different lineages

and have evolved independently many times during the mammalian evolution [56]. Additionally, Ohta [57] proposed that inter-and intra-chromosomal unequal crossing overs, coupled with mutation and random drift, are among the fundamental forces in the evolution of

multigene families. More importantly, inter-and intra-chromosomal unequal crossing overs have been shown to have a dominant effect on the contraction and expansion of genes in a given family [57, 58]. Therefore, it could be possible that the ancestral locus of PRDM9,

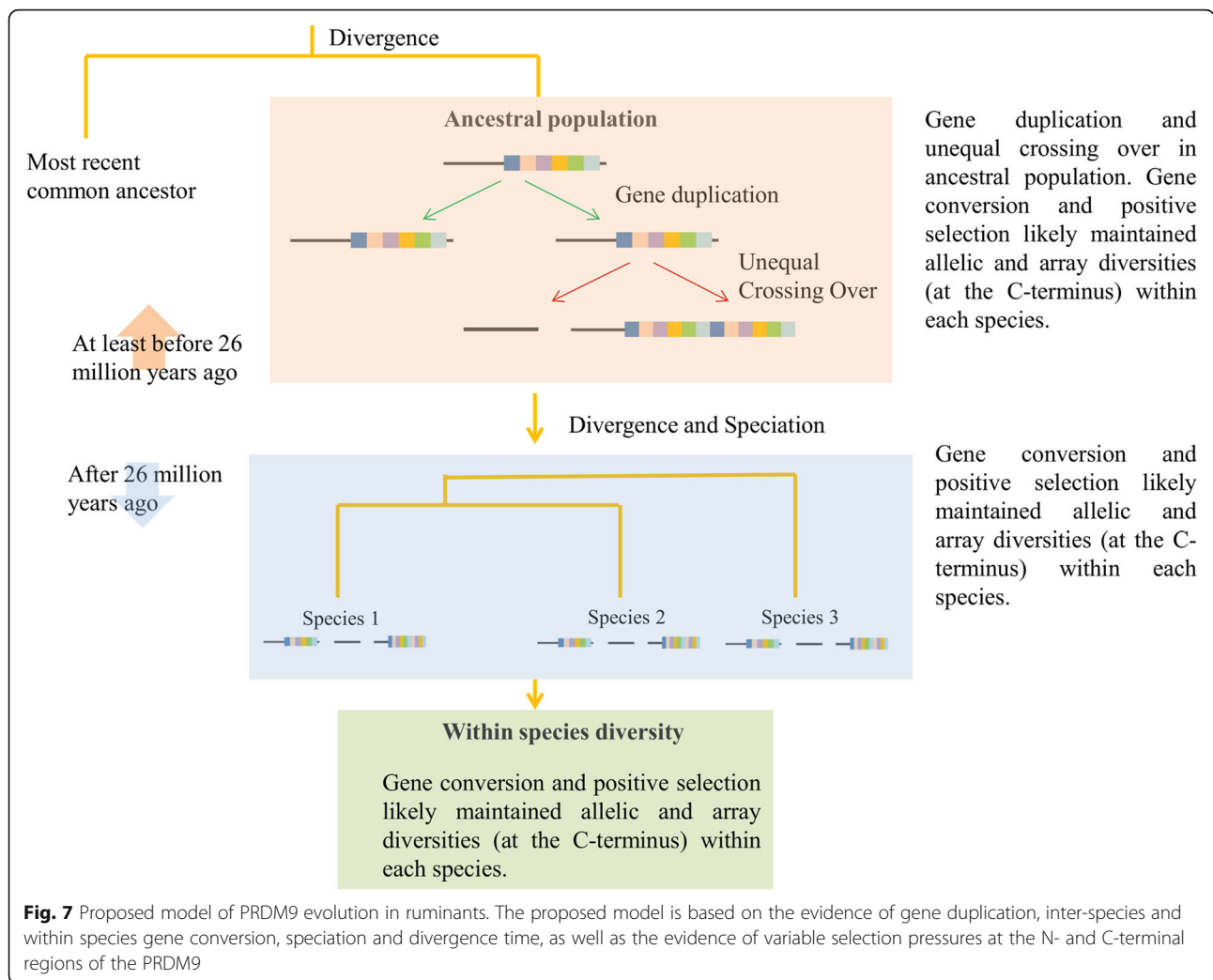


which is originally located at the autosomal region in most of the metazoans, appeared in the ruminant's chrX through unequal crossing overs, which might have happened millions of years ago possibly prior to ruminant's speciation and resulted in two additional copies of X-linked PRDM9. Given the fact that ruminants PRDM9 copies have been in the autosome and in the X chromosome for at least the past 27 million years, these copies are predicted to have differential evolutionary trajectories [56]. Mammalian sex chromosome genes are predicted to evolve at a much higher rate, and the fixation rate of beneficial mutations is predicted to be higher for X-linked genes than that of autosomal genes [56]. Interestingly, the observed elevated dN/dS ratio (i.e., dN/dS > 1), which indicates the evidence of positive selection, further supports the notion of accelerated rate of evolution of novel gene copies after a duplication event [48, 49]. Additionally, these duplicated copies may also have some functional consequences, and three possibilities would be expected [48, 59, 60]: i) the novel copies are likely to have experienced relaxed selection pressure and ultimately may acquire deleterious mutations that lead to loss of function, a process known as non-functionalization; ii) in rare cases

the novel copies can acquire beneficial mutations that differentiate their functions from that of the ancestor, a process known as neo-functionalization; and iii) mutations may occur in both ancestor and duplicated copies of a gene and result in complementary functions which is known as sub-functionalization [59, 60]. The presence of a stop-codon at the KRAB region in three sequences representing the genus *Ovis* and *Capra* of the lineage III (Additional file 1) supports the notion of non-functionalization; however an artifact of sequencing errors cannot be ruled out.

Although it is apparent that PRDM9 of chr1 regulates meiotic recombination in cattle [13] the functional significance of the X-linked PRDM9 is yet to be explored. Nevertheless, even in the absence of gene duplication event, sex chromosome genes are predicted to evolve at a faster rate than autosomal genes [56]. Therefore, the mutation rate of the X-linked PRDM9 is expected to be higher than that of the autosomal copy. However, due to the limited sample size, we could not directly estimate the mutation rate for each lineage, but the observation of incomplete lineage sorting for *Bos* species in chr1 may be an indication of slower mutation rate of





lineage I. This inference, however, should be taken with caution since sequences representing more species are required to test the hypothesis of mutational differences between the X-linked and autosomal PRDM9 copies.

In contrast to the N-terminal portion of PRDM9 which comprises three conserved functional domains [18], the C-terminal C2H2 ZFs of lineages I and II and the out-group, especially the amino acid residues at the positions -1, 3 and 6 that played crucial roles in DNA binding during meiotic recombination [16], have likely evolved under strong positive selection. Although this observation of extremely rapid evolution of ruminant's PRDM9 C2H2 ZFs by positive Darwinian selection is nothing surprising and has been reported for several other mammalian species [16, 21, 32, 38, 39], the evidence of positive selection on the X-linked C2H2 ZFs is one of the most striking observations. This compelling evidence of positive selection on the X-linked C2H2 ZFs PRDM9 indicates some unknown functional significance, thus warrants further investigation on the functional significance of the X-linked PRDM9

C2H2 ZFs. Consistent with a previous study [16], the present study has also showed evidence of concerted evolution of both X-linked (lineage II) and autosomal (lineages I) ZFs of PRDM9, which explained the species-specific, even at the individual level, variations in the length of the tandemly repeated C2H2 ZFs and the predicted binding motifs as well.

### Conclusions

In stark contrast to the primate lineage where the PRDM9 duplication mechanism is still an unresolved issue [33] our study provides strong evidence that the autosomal PRDM9 of ruminants has been duplicated to the X chromosome in the ruminants, which likely happened before the ruminant speciation events. The presence of X-linked PRDM9 copies in ruminants could possibly be a rare event and may be explained by some unique evolutionary mechanisms, possibly, through unequal crossing-overs. Nevertheless, the inter-chromosomal duplications before the ruminant's speciation together with the persistent positive selection and

concerted evolution of ZFs, at both species and individual levels, shaped the evolution of autosomal and X-linked PRDM9 in ruminants. Collectively, this study reports the unique evolutionary mechanism of PRDM9 in ruminants, including the presence of duplicated copies of PRDM9 on chr1 and chrX both with active C2H2 ZFs under positive selection. Concomitantly, a recent study has also reported extensive diversity of PRDM9 in several ruminant species [40]. Nevertheless, given such lineage-based unique evolutionary trajectories of the PRDM9, as demonstrated in the present study as well as in previous studies (eg., [16, 33]), taking more taxonomic lineages into consideration, future studies should be carried out to unravel the evolutionary trajectory of this important speciation gene across the metazoans.

## Methods

### Phylogenetic analyses

To unravel the evolutionary dynamics of PRDM9 and its novel copies in ruminants using the previously characterized cattle PRDM9 as reference sequences ([61] GenBank accession numbers: GJ060462 KJ020105), all the available complete coding nucleotide sequences of ruminant PRDM9 were retrieved from GenBank [62] (Additional file 4). Since the PRDM gene sequences have varying numbers of Zinc Finger (ZF) repeat sequences at their C-terminal domain, to avoid non-specific hits, we used the N-terminal portion of the PRDM amino acid sequences of the reference genomes and subsequently retrieved the complete DNA sequence of each PRDM7/9. Using the well-characterized and annotated human [9, 10, 52], mouse [32], and cattle [61] PRDM 7/9 protein sequences, we also retrieved the PRDM7/9 protein sequences representing primates, rodents, ruminants, and aquatic mammalian groups from GenBank. The conserved SET domain that comprises 118 amino acids was used for phylogenetic reconstruction of PRDM7/9 gene tree and specifically to assess the phylogenetic positioning of the human PRDM7. We aligned the protein sequences and manually checked the sequence quality using MEGA7 [63]. Amino acid alignments of the N-terminal functional domains representing different taxonomic groups (primates, rodents, ruminants, and aquatic mammals) are shown in Additional files 1, 2 and 3. To reconstruct the PRDM gene tree, amino acid sequences of the SET domain of 17 PRDM genes [31] were used as reference sequences. Based on the previous reports, functional domains of PRDM9 were identified [16, 18]. Sequences were aligned using the MUSCLE algorithm implemented in MEGA7 [63]. All the sequences were manually visualized to ensure high quality. Since the N-terminal portion of the PRDM9 that comprises three functional domains has slower evolutionary rate and evolutionarily conserved across the metazoan lineages

[18], we used this portion of the sequences to infer evolutionary history and phylogenetic relatedness among the novel copies of the ruminant's PRDM9. Protein alignment of the N-terminal portion of the PRDM9 and its novel copies are shown in Additional files 1 and 3. Aquatic mammals seemed to have close phylogenetic affiliation with ruminants [46], therefore PRDM9 of aquatic mammals were used as outgroup. Nucleotide and amino acid based maximum-likelihood (ML) phylogenies were reconstructed under appropriate substitution models in MEGA7 [63]. Appropriate models of nucleotide and amino acid substitutions for the respective datasets were selected under the Bayesian Information Criterion (BIC) implemented in MEGA7. JTT (Jones–Taylor–Thornton) + G (gamma distribution shape parameter) [64] and TrN93 (Tamura–Nei) + G [65], respectively, were the best-fit amino acid and nucleotide substitution models selected by BIC. Using the same program, nodal supports were estimated with 1000 bootstrap replicates. The time of divergence of the respective clades/species that were previously estimated based on the fossil based molecular clock calibration [44–47] were used to determine the timing of ruminant's speciation and ruminant's PRDM9 duplication events. The ZF arrays in each PRDM9 sequences were identified according to the previously defined nomenclature [9]. The putative DNA binding motifs for each PRDM9 C2H2 were predicted using the software (available at: <http://compbio.cs.princeton.edu/zf/>) [66, 67], which has been previously used in the prediction of PRDM9 binding motifs in primates [14, 68, 69].

### Test for positive selection

Given the fact that the presence of recombinant sequences in the data set could potentially affect the selection analyses [70, 71] using the recombination detection programs (RDP) implemented in RDP ver. 3 [72], we performed recombination detection analyses to ensure there are no recombinant sequences in the respective data sets used in selection analyses. The ratio of nonsynonymous (dN) to synonymous (dS) substitutions ( $\omega = dN/dS$ ), which has been widely used to measure the strength of selection on a protein-coding gene [73, 74], was used to measure the selection pressures in each dataset under five codon-based substitution models (neutral models: M1a, M7, M8a; selection models: M2a M8) that are implemented in the CODEML of the PAML 4.7 package [75], and their performances were evaluated using Likelihood Ratio Tests (LRTs) [73, 76]. Codon sites with Bayes-Empirical Bayes (BEB) posterior probability  $\geq 0.95$  were considered to be under positive selection. The inferred unrooted ML trees for the respective datasets that were used as input trees for the CODEML program were reconstructed using the PhyML ver. 3 [77]. To know whether  $\omega$  varies across the

branches, using the inferred phylogeny we compared the free-ratios model (M1), which assumes an independent  $\omega$  for each branch, with the one-ratio model (M0) that assume uniform  $\omega$  across the branches [76]. LRT was used to select the best-fit model. To check consistency of the selection results, we performed selection analyses using different input trees that were built under different tree-building methods implemented in MEGA [63] and PhyML [77]. Our selection results are very much consistent and are not biased by different tree building methods. To know the patterns of nonsynonymous and synonymous variations across the ZFs in respective lineages, using the DNAsp ver 5.0 [78], we also performed Sliding Window (window length = 5bp step size = 1bp) Analyses (SWA).

### Additional files

**Additional file 1:** Amino acid alignment of the N-terminal region of the PRDM9. (PDF 187 kb)

**Additional file 2:** Amino acid alignment of the SET domain of PRDM7/9. (PDF 100 kb)

**Additional file 3:** The PR domains of human PRDM7 and PRDM9 are aligned with the corresponding sequences of each lineage. (PDF 96 kb)

**Additional file 4:** GenBank accession numbers of the PRDM9 nucleotide sequences analyzed in this study. (PDF 80 kb)

### Abbreviations

BDM: Bateson-Dobzhansky-Muller; BIC: Bayesian information criterion; dN: Nonsynonymous; dS: Synonymous; JTT: Jones–Taylor–Thornton; LRT: Likelihood ratio test; ML: Maximum-likelihood; PRDM9: PR domain9; SWA: Sliding window analyses; TrN93: Tamura-Nei; ZFs: Zinc fingers

### Acknowledgements

We thank two anonymous reviewers for the insightful comments which greatly improved the manuscript.

### Funding

This work was supported in part by Agriculture and Food Research Initiative Competitive Grant 2016-67015-24886 from the USDA National Institute of Food and Agriculture and by MAES Competitive Grant from the Maryland Experimental Station.

### Availability of data and materials

The datasets analyzed in this study are publicly available in NCBI GenBank with accession numbers provided in Additional file 4.

### Authors' contributions

Conceived and designed the experiment: AP, LM; Analyzed the data: AP, BS, JJ, YZ; Supplied tools and reagents: LM, GEL; Prepared the first draft: AP; Revised the manuscript: AP, LM, GEL. All authors read and approved the final draft of the manuscript.

### Competing interests

The authors declare that they have no competing interests.

### Consent for publication

Not applicable.

### Ethics approval and consent to participate

Not applicable.

### Author details

<sup>1</sup>Department of Animal and Avian Sciences, University of Maryland, College Park, MD 20742, USA. <sup>2</sup>Animal Genomics and Improvement Laboratory, Agricultural Research Service, USDA, Beltsville, MD 20705, USA. <sup>3</sup>College of Animal Science and Technology, Northwest A & F University, Shaanxi Key Laboratory of Agricultural Molecular Biology, Yangling, Shaanxi 712100, China.

Received: 1 July 2016 Accepted: 26 January 2017

Published online: 14 March 2017

### References

- Bateson W. Heredity and variation in modern lights. *Darwin and Modern Science*. 1909;85–81.
- Dobzhansky T. Studies on hybrid sterility. I. Spermatogenesis in pure and hybrid *Drosophila pseudoobscura*. *Z Zellforsch Mikrosk Anat*. 1934;21:169–221.
- Muller HJ. Isolating mechanisms, evolution, and temperature. *Biol Symp*. 1942;6:71–125.
- Emerson BC, Kolm N. Species diversity can drive speciation. *Nature*. 2005;434(7036):1015–7.
- Rolland J, Condamine FL, Jiguet F, Morlon H. Faster speciation and reduced extinction in the tropics contribute to the Mammalian latitudinal diversity gradient. *PLoS Biol*. 2014;12(1):e1001775.
- Davies B, Hatton E, Altemose N, Hussin JG, Pratto F, Zhang G, Hinch AG, Moralli D, Biggs D, Diaz R, et al. Re-engineering the zinc fingers of PRDM9 reverses hybrid sterility in mice. *Nature*. 2016;530(7589):171–6.
- Baker CL, Kajita S, Walker M, Saxl RL, Raghupathy N, Choi K, Petkov PM, Paigen K. PRDM9 drives evolutionary erosion of hotspots in *Mus musculus* through haplotype-specific initiation of meiotic recombination. *PLoS Genet*. 2015;11(1):e1004916.
- Baudat F, Buard J, Grey C, de Massy B. Prdm9, a key control of mammalian recombination hotspots. *Med Sci*. 2010;26(5):468–70.
- Baudat F, Buard J, Grey C, Fledel-Alon A, Ober C, Przeworski M, Coop G, de Massy B. PRDM9 is a major determinant of meiotic recombination hotspots in humans and mice. *Science*. 2010;327(5967):836–40.
- Berg IL, Neumann R, Lam KW, Sarbajna S, Odenthal-Hesse L, May CA, Jeffreys AJ. PRDM9 variation strongly influences recombination hot-spot activity and meiotic instability in humans. *Nat Genet*. 2010;42(10):859–63.
- Berg IL, Neumann R, Sarbajna S, Odenthal-Hesse L, Butler NJ, Jeffreys AJ. Variants of the protein PRDM9 differentially regulate a set of human meiotic recombination hotspots highly active in African populations. *Proc Natl Acad Sci U S A*. 2011;108(30):12378–83.
- Hochwagen A, Marais GA. Meiosis: a PRDM9 guide to the hotspots of recombination. *Curr Biol*. 2010;20(6):R271–4.
- Ma L, O'Connell JR, VanRaden PM, Shen B, Padhi A, Sun C, Bickhart DM, Cole JB, Null DJ, Liu GE, et al. Cattle sex-specific recombination and genetic control from a large pedigree analysis. *PLoS Genet*. 2015;11(11):e1005387.
- Myers S, Bowden R, Tumian A, Bontrop RE, Freeman C, MacFie TS, McVean G, Donnelly P. Drive against hotspot motifs in primates implicates the PRDM9 gene in meiotic recombination. *Science*. 2010;327(5967):876–9.
- Neale MJ. PRDM9 points the zinc finger at meiotic recombination hotspots. *Genome Biol*. 2010;11(2):104.
- Oliver PL, Goodstadt L, Bayes JJ, Birtle Z, Roach KC, Phadnis N, Beatson SA, Lunter G, Malik HS, Ponting CP. Accelerated evolution of the Prdm9 speciation gene across diverse metazoan taxa. *PLoS Genet*. 2009;5(12):e1000753.
- Parvanov ED, Petkov PM, Paigen K. Prdm9 controls activation of mammalian recombination hotspots. *Science*. 2010;327(5967):835.
- Ponting CP. What are the genomic drivers of the rapid evolution of PRDM9? *Trends Genet*. 2011;27(5):165–71.
- Sandor C, Li W, Coppieters W, Druet T, Charlier C, Georges M. Genetic variants in REC8, RNF212, and PRDM9 influence male recombination in cattle. *PLoS Genet*. 2012;8(7):e1002854.
- Sandovici I, Sapienza C. PRDM9 sticks its zinc fingers into recombination hotspots and between species. *F1000 Biol Rep*. 2010;2:37.
- Schwartz JJ, Roach DJ, Thomas JH, Shendure J. Primate evolution of the recombination regulator PRDM9. *Nat Commun*. 2014;5:4370.
- Segurel L. The complex binding of PRDM9. *Genome Biol*. 2013;14(4):112.
- Segurel L, Leffler EM, Przeworski M. The case of the fickle fingers: how the PRDM9 zinc finger protein specifies meiotic recombination hotspots in humans. *PLoS Biol*. 2011;9(12):e1001211.

24. Smagulova F, Brick K, Pu Y, Camerini-Otero RD, Petukhova GV. The evolutionary turnover of recombination hot spots contributes to speciation in mice. *Genes Dev.* 2016;30(3):266–80.
25. Lesecque Y, Glemin S, Lartillot N, Mouchiroud D, Duret L. The Red Queen model of recombination hotspots evolution in the light of archaic and modern human genomes. *PLoS Genet.* 2014;10(11):e1004790.
26. Carrol L. *Through the looking glass and what Alice found there.* London: Macmillan; 1872.
27. Van Valen L. A new evolutionary law. *Evol Theory.* 1973;1:1–30.
28. Auton A, Rui Li Y, Kidd J, Oliveira K, Nadel J, Holloway JK, Hayward JJ, Cohen PE, Grealis JM, Wang J, et al. Genetic recombination is targeted towards gene promoter regions in dogs. *PLoS Genet.* 2013;9(12):e1003984.
29. Axelsson E, Webster MT, Ratnakumar A, Consortium L, Ponting CP, Lindblad-Toh K. Death of PRDM9 coincides with stabilization of the recombination landscape in the dog genome. *Genome Res.* 2012;22(1):51–63.
30. Munoz-Fuentes V, Di Rienzo A, Vila C. Prdm9, a major determinant of meiotic recombination hotspots, is not functional in dogs and their wild relatives, wolves and coyotes. *PLoS one.* 2011;6(11):e25498.
31. Fumasoni I, Meani N, Rambaldi D, Scafetta G, Alcalay M, Ciccarelli FD. Family expansion and gene rearrangements contributed to the functional specialization of PRDM9 genes in vertebrates. *BMC Evol Biol.* 2007;7:187.
32. Kono H, Tamura M, Osada N, Suzuki H, Abe K, Moriwaki K, Ohta K, Shiroishi T. Prdm9 polymorphism unveils mouse evolutionary tracks. *DNA Res.* 2014; 21(3):315–26.
33. Vervoort M, Meulemeester D, Behague J, Kerner P. Evolution of Prdm genes in animals: insights from comparative genomics. *Mol Biol Evol.* 2016;33(3):679–96.
34. Baudat F, Imai Y, de Massy B. Meiotic recombination in mammals: localization and regulation. *Nat Rev Genet.* 2013;14(11):794–806.
35. Billings T, Parvanov ED, Baker CL, Walker M, Paigen K, Petkov PM. DNA binding specificities of the long zinc-finger recombination protein PRDM9. *Genome Biol.* 2013;14(4):R35.
36. McVean G, Myers S. PRDM9 marks the spot. *Nat Genet.* 2010;42(10):821–2.
37. Patel A, Horton JR, Wilson GG, Zhang X, Cheng X. Structural basis for human PRDM9 action at recombination hot spots. *Genes Dev.* 2016;30(3):257–65.
38. Steiner CC, Ryder OA. Characterization of Prdm9 in equids and sterility in mules. *PLoS one.* 2013;8(4):e61746.
39. Thomas JH, Emerson RO, Shendure J. Extraordinary molecular evolution in the PRDM9 fertility gene. *PLoS one.* 2009;4(12):e8505.
40. Ahlawat S, Sharma P, Sharma R, Arora R, De S. Zinc finger domain of the PRDM9 gene on chromosome 1 exhibits high diversity in ruminants but its paralog PRDM7 contains multiple disruptive mutations. *PLoS one.* 2016;11(5):e0156159.
41. Sun XJ, Xu PF, Zhou T, Hu M, Fu CT, Zhang Y, Jin Y, Chen Y, Chen SJ, Huang QH, et al. Genome-wide survey and developmental expression mapping of zebrafish SET domain-containing genes. *PLoS one.* 2008;3(1):e1499.
42. Rabajante JF, Tubay JM, Ito H, Uehara T, Kakishima S, Morita S, Yoshimura J, Ebert D. Host-parasite Red Queen dynamics with phase-locked rare genotypes. *Sci Adv.* 2016;2(3):e1501548.
43. Rabajante JF, Tubay JM, Uehara T, Morita S, Ebert D, Yoshimura J. Red Queen dynamics in multi-host and multi-parasite interaction system. *Sci Rep.* 2015;5:10004.
44. Hedges SB, Dudley J, Kumar S. TimeTree: a public knowledge-base of divergence times among organisms. *Bioinformatics.* 2006;22(23):2971–2.
45. Kumar S, Hedges SB. TimeTree2: species divergence times on the iPhone. *Bioinformatics.* 2011;27(14):2023–4.
46. Meredith RW, Janecka JE, Gatesy J, Ryder OA, Fisher CA, Teeling EC, Goodbla A, Eizirik E, Simao TL, Stadler T, et al. Impacts of the cretaceous terrestrial revolution and KPg extinction on mammal diversification. *Science.* 2011;334(6055):521–4.
47. Hedges SB, Kumar S. *Discovering the TimeTree of life.* New York: Oxford University Press; 2009.
48. Ohno S. *Evolution by gene duplication.* Heidelberg: Springer; 1970.
49. Zhang J, Rosenberg HF, Nei M. Positive Darwinian selection after gene duplication in primate ribonuclease genes. *Proc Natl Acad Sci U S A.* 1998; 95(7):3708–13.
50. Baker CL, Walker M, Kajita S, Petkov PM, Paigen K. PRDM9 binding organizes hotspot nucleosomes and limits Holliday junction migration. *Genome Res.* 2014;24(5):724–32.
51. Jeffreys AJ, Cotton VE, Neumann R, Lam KW. Recombination regulator PRDM9 influences the instability of its own coding sequence in humans. *Proc Natl Acad Sci U S A.* 2013;110(2):600–5.
52. Blazer LL, Lima-Fernandes E, Gibson E, Eram MS, Loppnau P, Arrowsmith CH, Schapira M, Vedadi M. PR Domain-Containing Protein 7 (PRDM7) is a Histone 3 Lysine 4 Trimethyltransferase. *J Biol Chem.* 2016;291:13509. doi:10.1074/jbc.M116.721472.
53. Bailey JA, Eichler EE. Primate segmental duplications: crucibles of evolution, diversity and disease. *Nat Rev Genet.* 2006;7(7):552–64.
54. Eichler EE, Budarf ML, Rocchi M, Deaven LL, Doggett NA, Baldini A, Nelson DL, Mohrenweiser HW. Interchromosomal duplications of the adrenoleukodystrophy locus: a phenomenon of pericentromeric plasticity. *Hum Mol Genet.* 1997;6(7):991–1002.
55. Charlesworth B. The evolution of sex chromosomes. *Science.* 1991; 251(4997):1030–3.
56. Vicoso B, Charlesworth B. Evolution on the X chromosome: unusual patterns and processes. *Nat Rev Genet.* 2006;7(8):645–53.
57. Ohta T. An extension of a model for the evolution of multigene families by unequal crossing over. *Genetics.* 1979;91(3):591–607.
58. Ohta T. Theoretical population genetics of repeated genes forming a multigene family. *Genetics.* 1978;88(4):845–61.
59. Force A, Lynch M, Pickett FB, Amores A, Yan YL, Postlethwait J. Preservation of duplicate genes by complementary, degenerative mutations. *Genetics.* 1999;151(4):1531–45.
60. Lynch M, Force AG. The origin of interspecific genomic incompatibility via gene duplication. *Am Nat.* 2000;156(6):590–605.
61. Zimin AV, Delcher AL, Florea L, Kelley DR, Schatz MC, Puiu D, Hanrahan F, Pertea G, Van Tassel CP, Sonstegard TS, et al. A whole-genome assembly of the domestic cow, *Bos taurus*. *Genome Biol.* 2009;10(4):R42.
62. Benson DA, Clark K, Karsch-Mizrachi I, Lipman DJ, Ostell J, Sayers EW. GenBank. *Nucleic Acids Res.* 2014;42(Database issue):D32–7.
63. Kumar S, Stecher G, Tamura K. MEGA7: Molecular Evolutionary Genetics Analysis version 7.0 for bigger datasets. *Mol Biol Evol.* 2016;33(7):1870–4. doi: 10.1093/molbev/msw054.
64. Jones DT, Taylor WR, Thornton JM. The rapid generation of mutation data matrices from protein sequences. *Comput Appl Biosci.* 1992;8(3):275–82.
65. Tamura K, Nei M. Estimation of the number of nucleotide substitutions in the control region of mitochondrial DNA in humans and chimpanzees. *Mol Biol Evol.* 1993;10(3):512–26.
66. Persikov AV, Osada R, Singh M. Predicting DNA recognition by Cys2His2 zinc finger proteins. *Bioinformatics.* 2009;25(1):22–9.
67. Persikov AV, Singh M. De novo prediction of DNA-binding specificities for Cys2His2 zinc finger proteins. *Nucleic Acids Res.* 2014;42(1):97–108.
68. Auton A, Fledel-Alon A, Pfeifer S, Venn O, Segurel L, Street T, Leffler EM, Bowden R, Aneas I, Broxholme J, et al. A fine-scale chimpanzee genetic map from population sequencing. *Science.* 2012;336(6078):193–8.
69. Pratto F, Brick K, Khil P, Smagulova F, Petukhova GV, Camerini-Otero RD. DNA recombination. Recombination initiation maps of individual human genomes. *Science.* 2014;346(6211):1256442.
70. Anisimova M, Nielsen R, Yang Z. Effect of recombination on the accuracy of the likelihood method for detecting positive selection at amino acid sites. *Genetics.* 2003;164(3):1229–36.
71. Scheffler K, Martin DP, Seoighe C. Robust inference of positive selection from recombining coding sequences. *Bioinformatics.* 2006;22(20):2493–9.
72. Martin DP, Lemey P, Lott M, Moulton V, Posada D, Lefevre P. RDP3: a flexible and fast computer program for analyzing recombination. *Bioinformatics.* 2010;26(19):2462–3.
73. Swanson WJ, Nielsen R, Yang Q. Pervasive adaptive evolution in mammalian fertilization proteins. *Mol Biol Evol.* 2003;20(1):18–20.
74. Yang Z, Bielawski JP. Statistical methods for detecting molecular adaptation. *Trends Ecol Evol.* 2000;15(12):496–503.
75. Yang Z. PAML 4: phylogenetic analysis by maximum likelihood. *Mol Biol Evol.* 2007;24(8):1586–91.
76. Yang Z, Nielsen R, Goldman N, Pedersen AM. Codon-substitution models for heterogeneous selection pressure at amino acid sites. *Genetics.* 2000;155(1):431–49.
77. Guindon S, Delsuc F, Dufayard JF, Gascuel O. Estimating maximum likelihood phylogenies with PhyML. *Methods Mol Biol.* 2009;537:113–37.
78. Librado P, Rozas J. DnaSP v5: a software for comprehensive analysis of DNA polymorphism data. *Bioinformatics.* 2009;25(11):1451–2.