PLoS one

The Expansion of the *PRAME* Gene Family in Eutheria

Ti-Cheng Chang[1,2], Yang Yang[1¤], Hiroshi Yasue[3], Arvind K. Bharti[4], Ernest F. Retzel[4], Wan-Sheng Liu[1,2]*

1 Department of Dairy and Animal Science, The Center for Reproductive Biology and Health, College of Agricultural Sciences, The Pennsylvania State University, University Park, Pennsylvania, United States of America, 2 The Integrative Biosciences Program, Bioinformatics and Genomics Option, The Huck Institute of Life Sciences, The Pennsylvania State University, University Park, Pennsylvania, United States of America, 3 National Institute of Agrobiological Sciences, Tsukuba, Japan, 4 National Center for Genome Resources, Santa Fe, New Mexico, United States of America

Abstract

The *PRAME* gene family belongs to the group of cancer/testis genes whose expression is restricted primarily to the testis and a variety of cancers. The expansion of this gene family as a result of gene duplication has been observed in primates and rodents. We analyzed the *PRAME* gene family in Eutheria and discovered a novel Y-linked *PRAME* gene family in bovine, *PRAMEY*, which underwent amplification after a lineage-specific, autosome-to-Y transposition. Phylogenetic analyses revealed two major evolutionary clades. Clade I containing the amplified *PRAMEY*s and the unamplified autosomal homologs in cattle and other eutherians is under stronger functional constraints; whereas, Clade II containing the amplified autosomal *PRAME*s is under positive selection. Deep-sequencing analysis indicated that eight of the identified 16 *PRAMEY* loci are active transcriptionally. Compared to the bovine autosomal *PRAME* that is expressed predominantly in testis, the *PRAMEY* gene family is expressed exclusively in testis and is up-regulated during testicular maturation. Furthermore, the sense RNA of *PRAMEY* is expressed specifically whereas the antisense RNA is expressed predominantly in spermatids. This study revealed that the expansion of the *PRAME* family occurred in both autosomes and sex chromosomes in a lineage-dependent manner. Differential selection forces have shaped the evolution and function of the *PRAME* family. The positive selection observed on the autosomal *PRAME*s (Clade II) may result in their functional diversification in immunity and reproduction. Conversely, selective constraints have operated on the expanded *PRAMEY*s to preserve their essential function in spermatogenesis.

Introduction

Cancer/testis (CT) genes comprise a group of genes involved primarily in immunity and reproduction. They are expressed in various types of cancers when abnormally activated, whereas, the normal expression of CT genes is restricted mainly to the testis, but it has been detected also in other tissues such as fetal ovary [1–3]. CT genes have more than 240 members from 70 families. Twenty-four of these families are located on the human X-chromosome (CT-X) and two families, the *TSPY* (*testis-specific protein Y-linked*) gene family and the Y-linked *TPTE* (*transmembrane phosphatase with tensin homology*) pseudogene family, are on the Y-chromosome (Y-chr) [4]. Interestingly, many amplified CT gene families are located within direct or inverted repeats on the sex chromosomes (chrs) [1,4]. The autosomal CT genes were conserved during evolution and play roles in spermatogenesis, fertilization, and apoptosis in malignant cells [5–7]. However, knowledge about the CT genes on the sex chrs is still limited. A comparative study suggested that the CT-X genes were subject to positive selection and evolved faster than the autosomal CT genes [4]. The Y-linked *TSPY* gene family is conserved among most mammalian species, and has 30–60 copies on the human Y-chr [8]

and 50–200 copies on the bovine Y-chr (BTAY) [8,9]. This family has a typical CT tissue-restricted expression pattern with functions in immunity and spermatogenesis [10]. In this study, we identified a novel Y-linked CT gene family, *preferentially expressed antigen in melanoma, Y-linked* (*PRAMEY*), and examined its evolution in Eutheria.

PRAME, as one of the CT genes, first identified as an antigen-encoding gene related to immunity in a melanoma cell line [2], is expressed predominantly in normal testis and melanoma, lung squamous cell carcinoma, and acute leukemia, and at much lower levels in the ovary and other tissues [2,11]. The human *PRAME* gene, located on chromosome 22 (HSA22), encodes a protein with seven leucine-rich (LXXLL) motifs through which PRAME interferes with the retinoic acid receptor (RAR) pathway, and leads to the inhibition of RA-induced differentiation, growth arrest, and apoptosis [12]. Thus, PRAME functions as a transcriptional repressor in the signaling cascade, and the over-expression of *PRAME* results in tumorigenesis [12]. Similar to the other multi-copy CT genes, *PRAME* went through expansion and constituted a large gene family in most mammalian species [13,14]. A previous phylogenetic analysis of the primate *PRAME* family has revealed that the expansion of the human paralogs is

hominin-specific and occurred within the past three million years [13]. Several potential surface-accessible sites of the human PRAME protein have been identified under positive selection during evolution [13]. Even though the evolutionary pattern and oncogenic roles of the PRAME family have been studied in the human and rodent [2,11–13,15–17], the phylogeny of the PRAME orthologs in other mammalian species and the function of PRAME in normal tissues, such as testis, remain unclear.

To delineate the macro-evolution of PRAME, we analyzed the PRAME gene family in Eutheria. We discovered a bovine Y-linked PRAME family, namely PRAMEY, which was derived from an autosome-to-Y transposition and underwent amplification after the transposition. A phylogenetic analysis of PRAME/PRAMEY orthologs in Eutheria identified two major clades, which were subject to diverse selection pressures. The origination of the PRAMEY family and its unique expression patterns in spermatids suggest that it plays an important role in spermatogenesis.

Results

Discovery of the PRAMEY Family

Two PRAMEY transcripts (PRAMEY1 and PRAMEY2) were identified through a large-scale direct testis cDNA selection using a micro-dissected, PCR amplified BTAY probe. PRAMEY1 is 99% identical to a predicted mRNA (GenBank acc. no. XM_001253165.1) located in a non-annotated bovine bacterial artificial chromosome (BAC) (GenBank acc. no. AC234911.1). This clone was validated as a Y-linked BAC by a male-specific PCR (Fig. 1). PRAMEY2 is 99% identical to an mRNA (GenBank acc. no. NM_001077979) located in a bovine Y-BAC (GenBank acc. no. AC234853.4). Full-length mRNAs of both transcripts were obtained by RACE (rapid amplification of cDNA ends) (Fig. 2). The mRNA of PRAMEY1 (GenBank acc. no. GU144301) is 2747 bp, with an open reading frame (ORF) from nucleotide (nt) 895 to 2436, and it encodes a peptide of 513 amino acids (aa). The mRNA of PRAMEY2 (GenBank acc. no. GU144302) is shorter (1888 bp), with an ORF from nt 104 to 1639, encoding a peptide of 511 aa (Fig. 2). The similarity between the coding regions of PRAMEY1 and PRAMEY2 is 88% at the nucleotide level and 90% at the protein level.

To address the question whether more loci of PRAMEY are present on BTAY, we searched PRAMEY1/2 against the bovine Y-BACs (available in NCBI) and identified a total of 10 potentially active PRAMEY paralogs (named PRAMEY1-10, Table S1) and 6 pseudogenes. The active- and pseudo-genes were mapped to a total of 11 Y-BACs, each containing one or two copies (Table S1). The pairwise similarity of the 10 active PRAMEY loci was >86%, with a 100% similarity between PRAMEY2 and PRAMEY3 in AC234853.4 (Table S2). PRAMEY1 contains 4 exons whereas PRAMEY2 contains 5 exons because the first exon of PRAMEY1 reads through the second exon, resulting in a single, larger exon (Fig. 2). The first two introns in the coding regions are conserved across all the PRAMEY loci, with a slight difference in length (1289–1371 bp and 274–284 bp) (Fig. 2). A major difference is present in the last intron (Fig. 2): the size is 758 bp in PRAMEY2/3/8, compared to 1161–1212 bp in the remaining PRAMEYs. This difference is the result of an indel of 403–454 bp that is specific to BTAY.

The putative PRAMEY protein isoforms share an identity of ≥82%. Seven important leucine-rich motifs have been identified in the human PRAME protein [12]. The alignment of the bovine PRAMEY/PRAME with the human PRAME on HSA22 revealed that these motifs are highly conserved (Fig. S1).

In addition, we found a predicted gene (GenBank acc. no. XR_082974.1) located on BTA17. This gene shares ~87% similarity with the identified Y-linked PRAMEYs (Table S2). Gene-specific PCR and sequencing (Table S3) confirmed the predicted PRAME on BTA17. This autosomal gene encodes a putative peptide of 410 aa and is located at 74.35 Mb close to two zinc-finger genes, ZNF280A (also known as SUHW1) and ZNF280B (SUHW2).

Expression analysis of the bovine PRAMEY

Expression of the putative PRAMEY loci was investigated by deep-sequencing of the selected testis cDNAs using the Illumina GAIIx (see methods) and aligning the short sequence reads (pair-ends, 2×36 bp) against unique coding regions of the PRAMEY genes (Table S2). Seven of the 10 PRAMEY loci are active at the transcription level (PRAMEY2/3/6-10), and six of the seven loci have exactly matched read-pairs (Fig. 1B); in contrast, PRAMEY1/4/5 have no matched reads. Further, PRAMEY2/3/6 have more uniquely matched reads (>20), suggesting a higher expression level at these loci. Taken together with the RACE result, at least eight of the 10 loci on BTAY have been confirmed to be active at the transcription level.

RT-PCR analysis (Table S3) across nine tissues revealed that PRAMEY2 was expressed specifically in the testis. In contrast, the autosomal PRAME gene on BTA17 was expressed highly in the testis, and low in the kidney, brain and muscle (Fig. 1A). In situ hybridization (ISH) of PRAMEY2 cRNA probes (Table S4) revealed that both sense and antisense transcripts of PRAMEY2 were expressed in adult testis (Fig. 3). The sense RNA of PRAMEY2 was expressed specifically in spermatids (Fig. 3A), whereas the antisense RNA was expressed in all cell types in the seminiferous tubules, with the highest expression occurring in spermatids (Fig. 3B). Quantitative (q) RT-PCR analysis of PRAMEY2 indicated that the expression of the sense RNA was low in 5-11-day and 3-month-old testes, but up-regulated in 8-month- and 24-month-old testes (Fig. 3E); the expression of antisense PRAMEY2 RNA increased slightly with age.

Phylogenetic tree of the PRAME/PRAMEY family

To investigate the evolution of PRAME/PRAMEY, the sequences of multiple PRAME loci in the human, chimpanzee, orangutan, mouse, rat and cattle were retrieved from NCBI (Table S2) [18]. A single autosomal ortholog was found in dog and horse. Multiple PRAME loci were detected on the pig chr 6 (SSC6), similar to the expansions observed in primates (HSA1, PTR1 and MMUL1) and rodents (MMU4 and RNO5/14) [13]. Since SSC6 has not been well-annotated, the corresponding matched regions were collected and aligned with the HSA22 ortholog by Splign [19] to confirm gene structures and splicing signals/sites, which gave rise to 10 swine orthologs containing long ORFs (ranging from 470 to 528 aa) with corresponding splicing sites (Table S2). In addition to the autosomal copies, we found X-linked PRAME (PRAMEX) in rodents and horses. However, we did not identify any ortholog of PRAME in the non-eutherian lineages examined, including opossum, platypus, chicken, frog and zebrafish, all of which have a genome sequence coverage of ≥6X, implying that the PRAME gene family is present in eutherian mammals only.

The coding regions of the retrieved PRAME sequences were used to establish phylogenetic trees using Maximum-likelihood (ML), Bayesian-inference (BI) and Neighbor-joining (NJ) methods. All the tree topologies were consistent and contained two major clades (Fig. 4). The first clade (Clade I) included the syntenic orthologs of the BTA17 PRAME on human (HSA22), macaque (MMUL10), chimpanzee (PRT22), dog (CFA26), horse (ECA8) and pig (SSC14). Interestingly, all the active bovine PRAMEY loci and PRAME on BTA17 were clustered on the same branch with a strong bootstrap support value (100%) (Fig. 4). This clade also
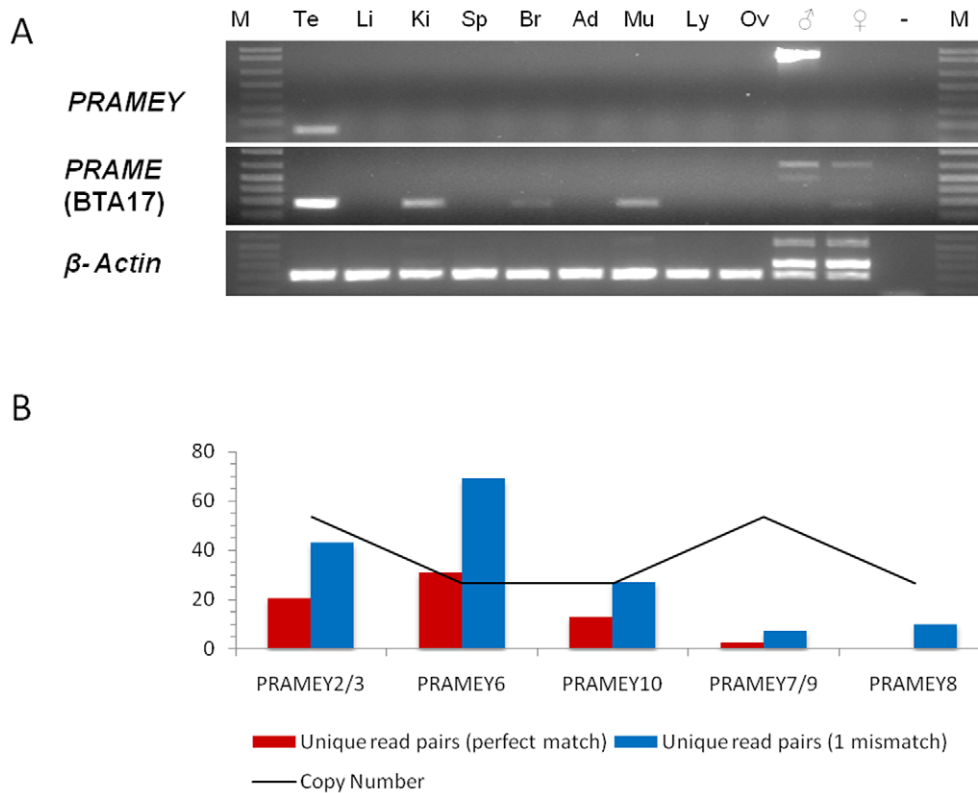
Figure 1. Expression patterns of *PRAME/PRAMEY* in cattle. A. RT-PCR results (lanes 2-10). *PRAMEY* is expressed specifically in the testis, whereas the autosomal *PRAME* is expressed in the testis (predominantly), kidney, brain and the muscle tissues. Bovine male genomic DNA-specific PCR (lanes 11–12) confirmed that *PRAMEY* is Y-specific. Te, testis; Li, liver; Ki, kidney; Sp, spleen; Br, brain (cerebrum); Ad, adrenal gland; Mu, muscle; Ly, lymph node; Ov, ovary; ♂, bovine male genomic DNA control; ♀, bovine female genomic DNA control; -, negative control (water); M, 1 kb DNA ladder. **B.** The expression of the *PRAMEY* loci by deep-sequencing analysis. The alignment of reads derived from deep-sequencing of selected cDNAs against coding regions of the *PRAMEY* loci (Table S2) reveals that seven of the 10 active *PRAME* genes are expressed differentially, six of which have significant numbers of both read-pairs matching exactly to the specific loci.
doi:10.1371/journal.pone.0016867.g001

included the orthologs on the horse and mouse X-chrs (ECAX and MMUX), which have a closer evolutionary distance to Clade I (0.713) than Clade II (0.814) (Maximum-Composite-Likelihood method) [20]. In Clade I, only the *PRAMEY* gene contains multiple copies, whereas the other homologs are all single-copy genes. Since no Y-linked ortholog was identified among the available Y-chrs of the other eutherian mammals, we propose that the bovine *PRAMEY* was derived by a lineage-specific, autosome-to-Y transposition event.

Clade II included the remaining orthologs with three internal clusters (Fig. 4). The first cluster (IIa) comprised the orthologs in

Artiodactyla, including those on BTA16 and SSC6. The second cluster (IIb) included all the orthologs on chr 1 in primates, where the human orthologs were intermingled with chimpanzee and orangutan orthologs as demonstrated previously [13]. The autosomal orthologs in Rodentia constituted the third cluster (IIc) and the mouse and rat orthologs were intermingled within the cluster. The X-linked orthologs in rats were also nested within this cluster. The orthologs in Clade II were all located in a chromosomal region syntenic to HSA1 except for the rat X-orthologs. The *PRAME* gene tree was reconciled with a species tree to reveal potential duplication and speciation events (Fig. 5) [21,22].
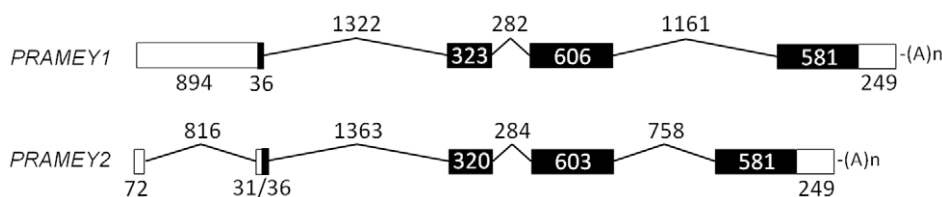
Figure 2. Genome structures of the bovine *PRAMEY* genes. Schematic representations of *PRAMEY1* and *PRAMEY2*. Compared to the *PRAMEY2* (GenBank acc. no. GU144302) that contains five exons, the first exon of the *PRAMEY1* (GenBank acc. no. GU144301) reads through to the second exon and forms a larger exon. The introns are drawn to scale. The open boxes represent UTR regions and the filled black boxes are coding segments (CDS). The numbers denote the length of exons, introns and CDS in bp. The polyA [(A)n] sites are indicated.
doi:10.1371/journal.pone.0016867.g002

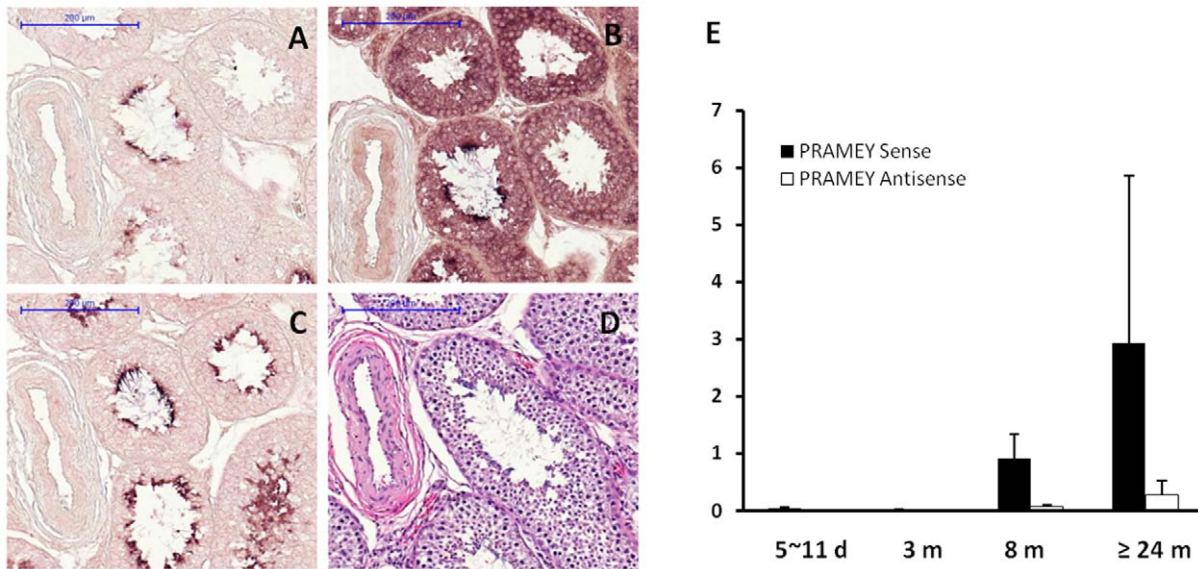**Figure 3. Spatial and temporal expression patterns of the sense and antisense RNA of the bovine *PRAMEY2* in adult bovine testis. A.** The sense RNA of *PRAMEY2* is expressed specifically in spermatids. **B.** The antisense RNA of *PRAMEY2* is expressed broadly across seminiferous tubules with a predominant expression in spermatids. Sense and antisense RNAs of *PRAMEY2* were detected by DIG-labeled cRNA probes. **C.** The bovine *PRM1* gene was used as positive control, and there is no antisense mRNA of *PRM1* detected in the bovine testis [38]. **D.** Haematoxylin and Eosin (H&E) staining is shown. Scale: bar = 200 μm. **E.** Temporal expression pattern of *PRAMEY2*. The relative expression levels of the *PRAMEY2* sense and antisense transcripts at different ages (X-axis), measured by the strand-specific qPCR, were normalized by the 18S rRNA (Y-axis). The *PRAMEY2* sense RNA is expressed very low in earlier stage, but up-regulated in the 8 months and 2 years-old testis. Similarly, antisense RNA of *PRAMEY2* is detected in the 8 months and 2 years-old testis. Values are means ± SD of the three biological replicates.
doi:10.1371/journal.pone.0016867.g003

Selection forces acting on the PRAME genes

The lineage-specific selection test using the PAML (Phylogenetic Analysis by Maximum Likelihood) package revealed that the dN/dS ratios varied significantly among different lineages (p<0.001, fixed ratio/free ratio branch model) [23]. We applied the branch-site models (model A null/model A) to examine whether any lineage is under positive selection [24]. In Clade I, we observed two branches, leading to the primate homologs and the bovine *PRAME* on BTA17, which were subject to positive selection (Fig. 5). Three positively selected sites were found along these two branches (probability >0.8, Table S5) [24]. We also tested different pairs of site-specific models (see methods) in a dataset containing only the homologs in Clade I (Table S2 and S6) [23], and the results were all negative (p>0.1). It is noteworthy that the homologs in Clade I had a significant lower median dN/dS ratio when compared to the three clusters in Clade II (p<0.001, Fig. 6A). Taken together, these data suggest that Clade I was under stronger functional constraints.

In Clade II, we detected a total of 17 sites from 14 different branches under positive selection (Fig. 5). Four sites and three branches were observed in Artiodactyla (Clade IIa), five sites/five branches in Primate (Clade IIb), and nine sites/six branches in Rodentia (Clade IIc) (Fig. 5). Our findings support a previous report that the primate and rodent *PRAME* homologs were subject to positive selection [13]. In this study, we further examined the potential selected sites in the homologs in Artiodactyla (Clade IIa) using the site-specific models (see methods) and detected eight more positively selected sites (model M8, probability >0.8, Table S6 and S7). Therefore, 12 sites in total were found under positive selection in Clade IIa. We built a PRAME protein homology model using the *PRAME* gene on BTA16 (GenBank acc. no. XM_001256020) as the template, and mapped the positively selected sites on the model (Fig. 5). In contrast to the primate and

rodent PRAME, in which the positively selected sites were clustered in the outer surface of the protein [13], the majority (8/12) of the positively selected sites in the bovine PRAME were located in the inner concave region (Fig. 6B). Furthermore, a DNA binding site was predicted in this protein model. This could be important as one of the positively selected sites (329M) and two of the leucine-rich motifs were located in this region. In addition, we also investigated whether or not the bovine paralogs, including the pseudogenes, were subject to gene conversion during evolution using the GENECONV program [25]. The results did not indicate any gene conversion events.

Discussion

Lineage-specific amplification of PRAMEs

PRAME is one of the most amplified gene families in mammals and is considered the third largest gene family in the mouse genome [26]. In the present study, we found that the *PRAME* gene family is present only in Eutheria, indicating that this family may have originated *de novo* in the eutherian lineages [27]. The birth-and-death model of gene duplication, instead of concerted evolution, has been suggested to be the major evolutionary mechanism accounting for the expansion of autosomal *PRAME* and the resemblance between each copy [13]. Our analysis revealed that: 1) during eutherian evolution, the expansion of *PRAME* genes was not limited to autosomes, but also occurred in sex chrs; 2) the expansion of *PRAMEs* is lineage-dependent. This conclusion was based upon the finding that the *PRAME* gene was transposed to and amplified on BTAY, but not on the other mammalian Y-chrs; 3) the intra- (cis-) and inter- (trans-) chromosomal duplications occurred during the expansion of the *PRAME* gene family. The cis-duplications occurred mainly for the syntenic *PRAMEs* in Clade II and the bovine *PRAMEYs* in Clade I
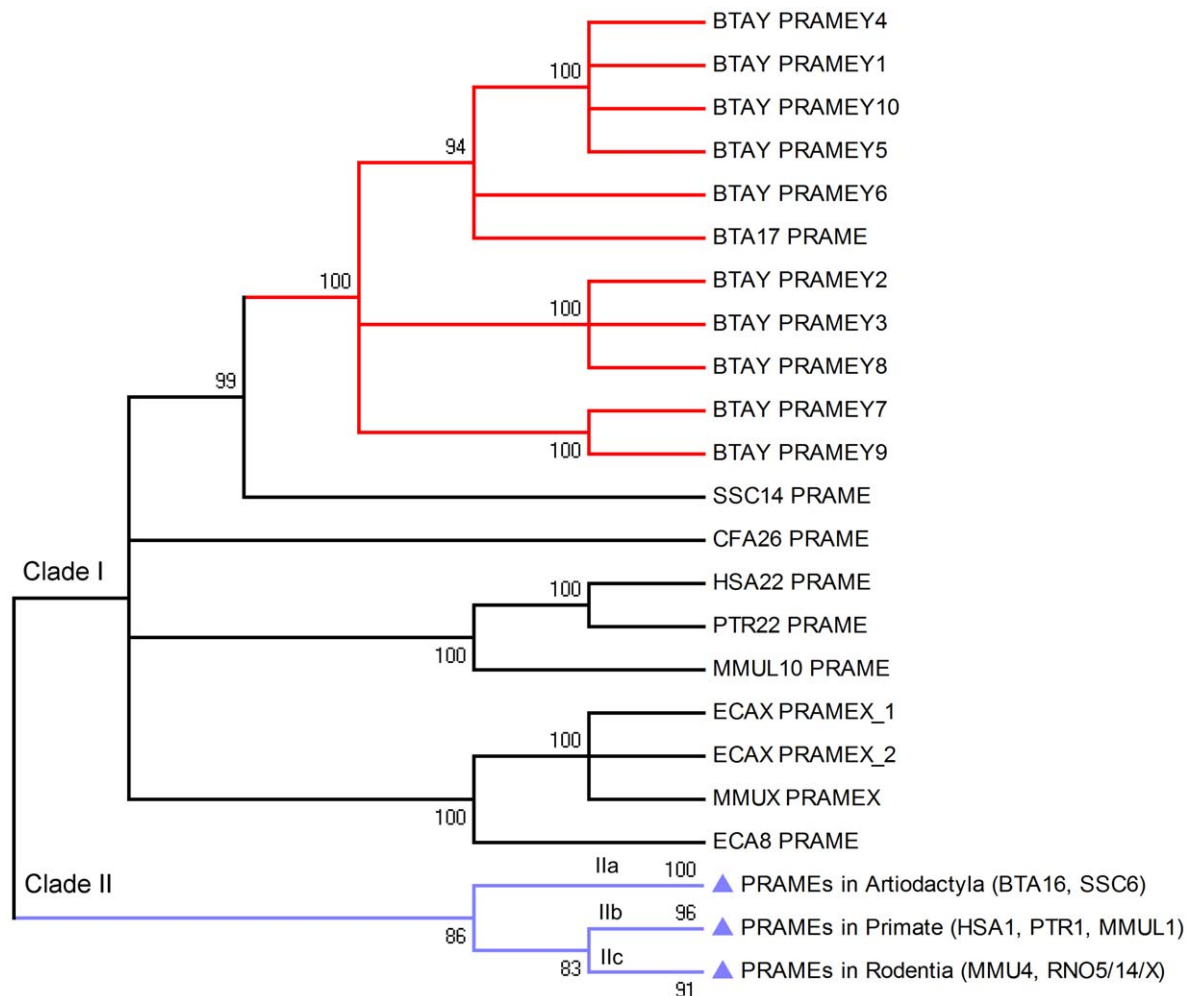
Figure 4. Phylogenetic tree of the *PRAME* gene family. Two major *PRAME/PRAMEY* clades are shown in this tree. The *PRAME* locus on HSA22 and its syntenic orthologs in other species are clustered with the bovine *PRAME* and *PRAMEY* loci in Clade I (branches in red). The orthologs on the X-chrs of horse and mouse are also clustered with Clade I. The *PRAME* orthologs syntenic to HSA1 are clustered in Clade II (branches in light blue), which contains three sub-clusters, IIa (Artiodactyla), IIb (Primates) and IIc (Rodentia). The tree was built based on the ML method and bootstrap values (1000 replicates) are shown above the branches. The branches corresponding to partitions reproduced in less than 80% bootstrap replicates are collapsed.
doi:10.1371/journal.pone.0016867.g004

(Fig.4, Table S2). The rat X-orthologs may be derived from the trans-duplication of the autosomal paralogs on RNO14, but the origin of the mouse X-ortholog is unclear (Fig. 4). It is noteworthy that the *PRAME* genes appear to (cis-) duplicate largely only on those chromosomal regions syntenic to HSA1 in Clade II. In contrast, the orthologs clustered with the *PRAME* gene on HSA22 tend to be maintained as single-copy genes in the respective genome, except for the bovine *PRAMEY* family, which could be a consequence of abundant reorganization and duplication events that occurred during the evolution of the Y-chr [28]. We observed five BACs, each containing two *PRAMEY* loci (Table S1), suggesting that the expansion of *PRAMEY* occurred in tandem on BTAY and gene duplication was the predominant process during the expansion. However, we cannot exclude the possibility that concerted evolution may also have contributed to the similarity between the *PRAMEY* genes because of potential Y-Y gene conversions [29,30]. The mechanism behind the frequent cis-duplications and limited trans-duplications of the *PRAME* gene family in Eutheria may be related to genomic contexts on each chromosome, including local gene density, repeat density, GC content and recombination rate [31].

Selective pressures on *PRAME(Y)*

Positive selection tends to increase the frequency of advantageous mutations; negative selection eliminates the deleterious mutations resulting in less genetic variation. A previous study found a large number of positively selected sites in both human and mouse *PRAME* orthologs on HSA1 and MMU4 [13]. In the present study, we found several branches leading to the orthologs in primates, rodents and artiodactyls in Clade II under positive selection (Fig. 5), which supports the previous report [13]. The selection test for the homologs on BTA16 and SSC6 detected 12 sites that were subject to positive selection (Fig. 6B, Table S5 and S7). Unlike the primate [13], the positively selected sites in Artiodactyla were clustered in the inner concave region, suggesting that the functional accommodations of *PRAMEs* are lineage-dependent. The protein structure of the bovine PRAME model (Fig. 6B) is close to the ribonuclease inhibitor (PDB: 1DFJ), which interacts with its substrate through a similar concave region [32]. Thus, the modifications of PRAME in Artiodactyla appear to occur along the regions essential for protein interaction during evolution. Further, the difference in the median dN/dS ratios between Clade I and Clade II (Fig. 6A)
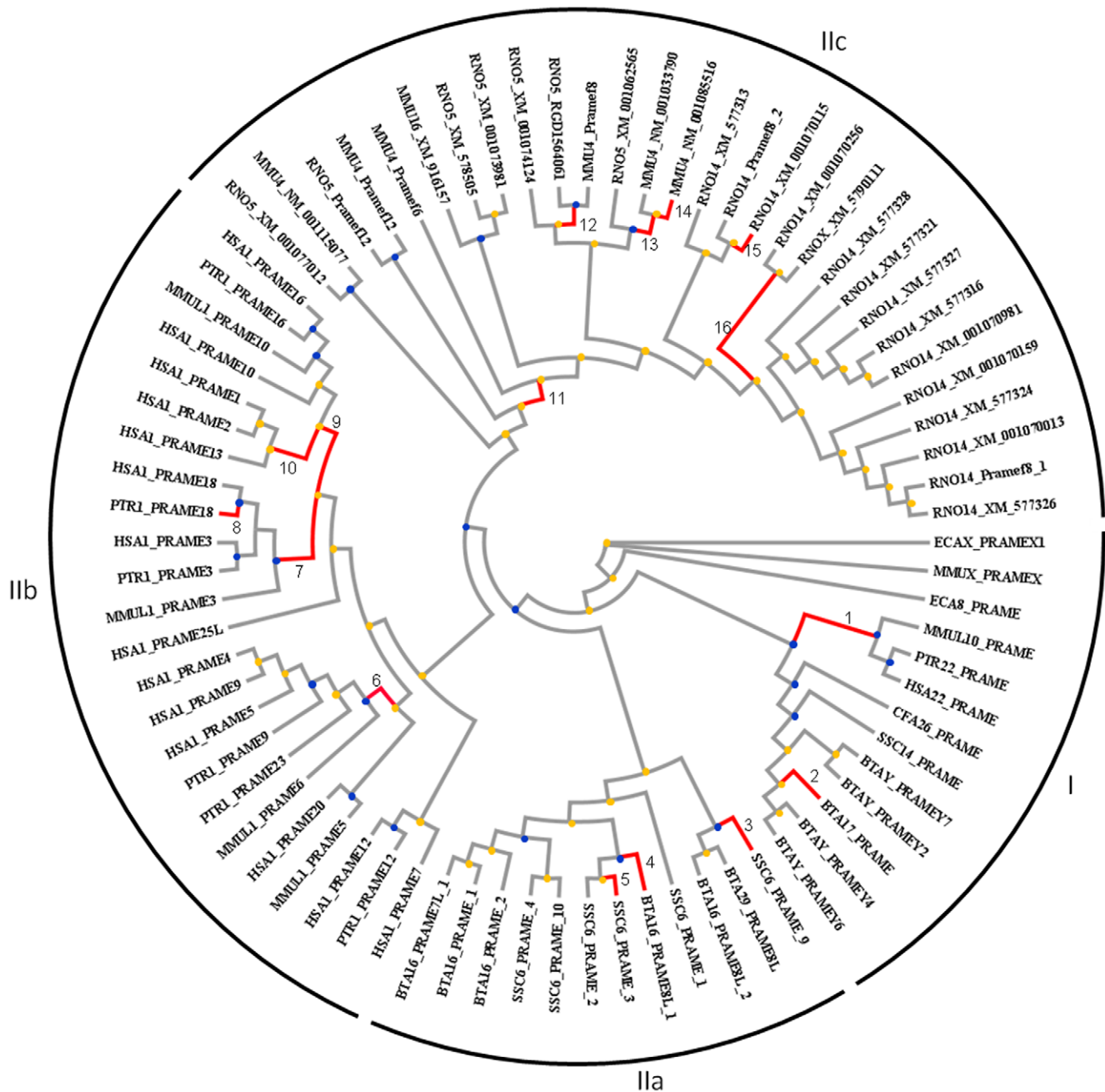
Figure 5. Positive selection on the *PRAME* and *PRAMEY* orthologs. Two branches in Clade I and 14 branches in Clade II are under positive selection (red) based on the branch-site model tests (Model A versus Model A null). The branches under positive selection are numbered and the selected sites along each foreground lineage are detailed in Table S5. The nodes underwent duplication are marked with a yellow circle and speciation with a blue circle.

doi:10.1371/journal.pone.0016867.g005

suggested differential selection pressures acting on the *PRAME* gene family.

Origin of *PRAMEY* in cattle

Our recent study [29] in cattle has shown that a gene block containing *ZNF280BY* and *ZNF280AY* was transposed from BTA17 and duplicated on the Y-chr after the transposition. In the present study, we found a *PRAME* on BTA17, which is linked to *ZNF280B/ZNF280A* within a 60 kb region (74.30–74.36 Mb). Meanwhile, the same gene order (*ZNF280BY -ZNF280AY-*

PRAMEY) was observed in two non-overlapping Y-BACs (GenBank acc. no. AC234853.4 and AC233215.5), leading us to hypothesize that the *PRAMEY*s were derived from the transposition of the block on BTA17. Unlike the human *DAZ* and feline *TETY1* and *FLJ36031* genes, in which the translocation was involved in a single autosomal gene, the bovine *ZNF280B-ZNF280A-PRAME* was transposed to the Y-chr as a block. However, the established phylogenetic tree of *PRAME/PRAMEY* in this study was not clear because the BTA17 locus was nested within the *PRAMEY* cluster (Fig. 3), raising an alternative but
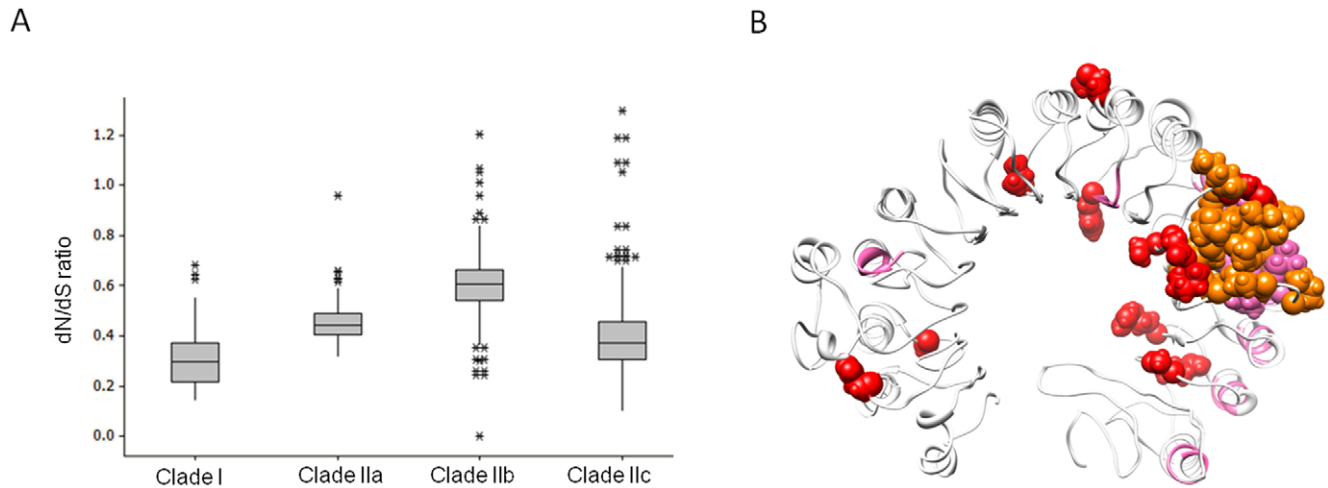
Figure 6. Selective pressures on the *PRAME* family. A. The dN/dS ratio distributions in different clades. Clade I has the lowest mean and median dN/dS ratios. The vertical axis represents the dN/dS ratio. The asterisk (*) represents the outliers of the data. **B.** Map of the positively selected sites detected in Clade IIa to the PRAME protein model. The selected sites derived from PAML analyses are mapped to the protein homology model. Eight of the 12 selected sites (red) are clustered in the inner concave region of the protein model. The model was built based on the *PRAME* gene (GenBank acc. no. XM_001256020.1) on BTA16. The predicted DNA binding site is highlighted in orange. The LXXLL motifs are highlighted in pink.
doi:10.1371/journal.pone.0016867.g006

unlikely hypothesis that the *PRAME* on BTA17 was derived from the loci on BTAY. If we assume a "Y-to-autosome" transposition occurred during evolution, we would expect this gene block to be conserved on the Y-chr of most, if not all, eutherians, but not conserved on autosomes. However, this block is highly conserved on autosomes (Fig. S2) instead of the Y, which apparently conflicts with the alternative hypothesis. Thus, we proposed that the *PRAMEY* genes in cattle were derived from the transposition of the *ZNF280B/ZNF280A/PRAME* on BTA17 and duplicated separately thereafter.

Furthermore, based on the tree topology (Fig. 3), it appears that *PRAMEY*s were clustered into two subgroups and could be derived from two transposition events. However, several observations led us to postulate that *PRAMEY*s were derived from a single transposition of the BTA17 gene block. First, all *PRAMEY*s are highly similar (>86%) and amplify tandemly in a narrow genomic region just like the *PRAME* expansion within 740 kb on HSA1 [13]. Several Y-BACs contain two copies of *PRAMEY*, such as *PRAMEY2* and *PRAMEY3*, which are identical and located in a BAC with a distance of 22 kb. More importantly, *PRAMEY6* and *PRAMEY7*, falling into different subgroups, are also located in one Y-BAC with a distance of 97 kb (Table S1). The narrow distance and high similarity of each copy indicated that the gene duplication is the major evolutionary mechanism of *PRAMEY*s after transposition. Two separate transpositions occurring within a narrow genomic region are implausible. Thus, we propose that the distinct clusters of *PRAMEY*s are the synergic consequence of a higher mutation rate on the non-recombining Y-chr [33] and Y-Y gene conversions [29,30]. The diversity of the duplicated *PRAMEY* sequences reflects a response of Y-chr to diverse selection pressures.

Potential roles of PRAME/PRAMEY

Several lines of evidence have indicated a close relationship between *PRAME* and tumorigenesis [12,13,15,16,34]. *PRAME* acts as a ligand-dependent co-repressor in the important retinoic acids receptor (RAR) pathway [12,34]. When PRAME is absent, the activation of the RAR pathway by retinoids will lead to proliferation arrest, cell differentiation and apoptosis [12].

Conversely, the RAR pathway is inhibited when *PRAME* is abnormally present, resulting in incessant cell proliferation and tumorigenesis.

In addition to tumor development, P*RAME* is implicated in germ cell development. In the mouse, an autosomal *Prame*-like gene, *Oogenesin*, is expressed in oocytes and early cleavage-stage embryos with a role in oogenesis [13,35], suggesting that the duplicated *PRAME* genes on autosomes are related to rapid cell mitosis. The mouse X-linked *Prame-like 3* (*Pramel3*) is expressed specifically in spermatogonia and may function in early stage of spermatogenesis [36]. Since maintaining and amplifying male fertility factors on the Y-chr may provide selection advantages during evolution [37], the origin and retention of these Y-linked copies are expected to be crucial for spermatogenesis. The exclusive expression of *PRAMEY* (Fig. 1 and 3) in spermatids provides a strong support for this hypothesis. We validated that at least eight of the 10 predicted *PRAMEY* loci are active at the transcription level, and differentially expressed in the testis (Fig. 1B). Future research is needed to investigate the biological meanings behind this differential expression. It is worth noting that the predominant expression of the bovine *PRAMEY2* antisense transcript in spermatid may be essential biologically (Fig. 3E). Our previous works demonstrated that the antisense RNAs of three other Y-related and testis-expressed genes (*ZNF280BY*, *DDX3Y* and *CDYL*) in cattle all appear to be expressed in late stage spermatocytes and/or spermatids, indicating that antisense RNA is crucial in the regulation of bovine spermiogenesis [29,38,39].

Recent and extensive duplications of *PRAME* and other CT genes in human are consistently involved in adaptive functions including reproduction and immunity [13,40]. *PRAME* and neighboring *ZNF280BY/ZNF280AY* on HSA22 are reportedly associated with immune responsiveness [41,42]. Thus, the *PRAME/PRAMEY* gene family may also participate in auto-immunity to sperm, which is prevented by the blood-testis barrier in normal males [43]. Anti-sperm immunity is considered as one of the causes of infertility in humans [44] and it is thus important to clarify the immunological roles of *PRAME* in male-related functions.

In conclusion, we have identified a lineage-specific *PRAMEY* gene family in bovine, which was derived from the transposition of

a gene block, *ZNF280B-ZNF280A-PRAME*, on BTA17, and duplicated afterwards. The expansion of *PRAME* genes occurred not only in Primates and Rodentia, but also in Artiodactyla. The phylogenetic analysis revealed two distinct clades of *PRAME* that evolved under different selection forces. The largely amplified autosomal *PRAME*s are under positive selection, whereas the *PRAMEY*s are under stronger functional constraints. The *PRAMEY* gene family is expected to be important in spermatogenesis. We anticipate that future research on the roles of *PRAME* and *PRAMEY* in the crosstalk between the spermatogenesis and immunoresponse will facilitate understanding of both spermatogenesis and tumor developments.

Materials and Methods

RNA extraction and cDNA synthesis

Total RNA was extracted from bovine testicular tissue at 4 days, 20 days, 3–4 months, 8 months, and 2 years of age with Trizol® reagent (Invitrogen, Carlsbad, CA, USA). Equal amounts of total RNA from different ages of testes were pooled and treated with DNase I twice (before and after mRNA purification) (Ambion, Austin, Texas, USA). Messenger RNAs were purified from the pooled total RNA (Oligotex; Qiagen, Valencia, CA, USA). First strand cDNAs were synthesized with random hexamers and oligo-T primers using SuperscriptIII reverse transcriptase (Invitrogen, Carlsbad, CA, USA); blunt-ended double-stranded cDNAs were generated as described [45]. Adaptors [phosphorylated oligonucleotides 1 (5′-CTGAGCGGAATTCGTGAGACC-3′) and 2 (5′-CCAGAGTGCTTAAGGCGAGTCAA-3′)] were attached to cDNAs using T4 polynucleotide kinase (NEB, Ipswich, MA, USA)[46]. Adaptor-ligated cDNA products were used for direct testis cDNA selection.

Direct testis cDNA selection and sequencing

The entire BTAY DNA was isolated by a micro-dissection approach [47]. The DNA fragments were PCR amplified and labeled with biotin-16-dUTP (Roche, Indianapolis, IN, USA) by nick translation (Roche, Indianapolis, IN, USA). Direct testis cDNA selection was detailed in Yang *et al.* (2011) [29] and Del Mastro and Lovett (1997) [46]. The selected cDNAs were PCR-amplified using the adaptor oligo 1 as the primer. Selection efficiency was assessed by qPCR with Y-linked genes, *SRY* and *DDX3Y*, as positive controls and *β-Actin* and *CDYL* as negative controls. PCR products were cloned using a TOPO-TA cloning kit (Invitrogen, Carlsbad, CA, USA). Randomly selected clones (n = 2208) were grown overnight at 37°C in 2 ml, 96-deep-well culture plates. All clones were dot-blotted on nylon transfer membranes and hybridized with ^{32}P-dCTP labeled BTAY fragments and PCR fragments of four genes (*HSFY*, *UBE2D3*, *RPL23A*, and *ZNF280B*) that were highly redundant in our test sequencing result. After dot-blot and elimination of the most likely repetitive clones, 753 clones were selected for sequencing. Plasmid DNA was purified by alkaline lysis (Qiagen, Valencia, CA, USA), and sequenced on an ABI-3730XL DNA analyzer at the Pennsylvania State University Genomics Core Facility.

RT-PCR

Total RNAs were extracted from nine different tissues (testis, liver, kidney, spleen, cerebellum, adrenal gland, longissimus muscle, lymph node, and spinal cord) of a 2-years old bull and ovarian tissue from a mature cow, then treated with DNase I (Ambion, Austin, TX, USA) and reverse transcribed using Superscript™ III First-Strand Synthesis System (Invitrogen, Carlsbad, CA, USA). RT-PCR was performed in 20 µl containing 10 ng cDNA, 200 µM dNTPs, 1.5 mM MgCl$_2$, 2.5 µM of each primer, 1 unit Taq DNA polymerase (Bioline, Taunton, MA, USA). The PCR conditions were: 94°C for 7 min followed by 35 cycles each of 95°C for 40 sec, 55°C–65°C for 40 sec, 72°C for 40 sec, with a final extension at 72°C for 7 min. Products were resolved on 1.5% agarose gels with ethidium bromide in 1× TAE buffer.

RACE

Total RNAs from bovine testis (5–11 days, 3 months, 8 months and 2 years of age) were used for 5′ and 3′ rapid amplification of cDNA ends (RACE). The RACE experiment was conducted essentially as described in Yang *et al.* [29].

Short-read sequencing for locus-specific expression

The selected cDNAs were sequenced at the National Center for Genome Resources using an Illumina GAIIx. Library construction and sequence methods were described previously [29]. A total of 6,710,574 high-quality paired end reads of 2×36 bp were generated. These reads were aligned to nine unique *PRAMEY* sequences identified through BlastClust [48] with 100% similarity and 100% coverage as the criteria. For aligning the short-reads, the software GSNAP [49] was used as part of the Alpheus pipeline [50]. Two mismatches were allowed during the alignment step and only the reads that hit the reference uniquely were considered for counting towards locus-specific expression. Since the reads were paired end, only the reads where both ends hit the same reference were considered. These counts were further sub-grouped under two categories: (A) both reads are unique hits with at least one of them being exact match and (B) both reads are unique hits & both are exact matches. The read counts in these two categories were considered a measure of expression pertaining to the specific locus.

Testis tissue section in situ hybridization (ISH)

The bovine testis was fixed [51], embedded in paraffin and sectioned (4 µm). Sense and antisense RNA probes of *PRAMEY* were selected (Table S4) using G-PROBE (Genetyx Co., Tokyo, Japan) and the 120-bp probes were subjected to *in vitro* transcription to produce digoxigenin (DIG)-labeled cRNA with the AmpliScribe T7-Flash Transcription Kit (Epicentre, Madison, WI, USA). Uniform labeling of DIG-labeling was confirmed using the NBT/BCIP detection system (Roche Diagnostics, Indianapolis, IN, USA). ISH was performed as described previously [38,39]. Serial tissue sections were used for sense and antisense probe hybridizations. The spermatid-specific gene *Protamine 1* (*PRM1*) served as the positive control, while *LNE120* staining was used as the negative control.

Strand-specific qPCR

First strand sense and antisense cDNAs were developed with strand-specific reverse transcript primers (Table S3) (Super-Script™ III First-Strand Synthesis System, Invitrogen, Carlsbad, CA, USA) from 5–11 day, 3 month, 8 month and >24-month bovine testis total RNA and used as templates for qPCR with gene specific primer sets (Table S3). All qPCRs were performed in the Power SYBR Green PCR Master Mix (Applied Biosystems, CA, USA) and Applied Biosystems 7500 real-time PCR system following the manufacturer's instructions. Amplification conditions were 2 min at 50°C; 10 min at 95°C; followed by 40 cycles of 20 sec at 95°C, 20 sec at 57°C and 30 sec at 72°C. Cycle threshold acquisition used default parameters with CT values for *PRAMEY* sense/antisense RNAs normalized to 18S rRNA in each sample. RNA samples without a reverse transcript served as the

negative control. Each qPCR was conducted in duplicate on three independent RNA samples (biological replicates). Significance was evaluated by one-way ANOVA using SAS (SAS Institute Inc., NC, USA).

Sequence alignment, gene prediction and phylogenetic tree construction

For the identification of bovine *PRAME* paralogs, we used the two identified transcripts (GenBank acc. no. GU144301 and GU144302) to blast against ~600 bovine Y-BACs that are available in GenBank to retrieve all potential paralogous regions on BTAY. The redundant regions were removed by detecting the overlaps between Y-BACs using purpose-designed scripts. The paralogs with inferred splicing sites/signals and comparable coding regions were considered as active *PRAMEY* genes; in contrast, the others were pseudogenes due to either frameshift mutations or premature stop codons.

Using the human *PRAME* sequences on HSA22 (GenBank acc. no. NM_206956.1) and HSA1 (GenBank acc. no. NM_023013.1) to blast [48] against the nucleotide databases in NCBI [18], we were able to retrieve the annotated *PRAME* homologs in humans, chimpanzees, orangutans, horses, cats and cattle (e-value <1E^{-50} and coverage >40%, Table S1 and S2). For the swine orthologs, the blast search was against the swine HTGS database as the swine genome sequence was not well annotated. The retrieved porcine BAC sequences were annotated for *PRAME* in this study using Splign [19] and the getorf program in EMBOSS [52]. The redundant porcine paralogs were removed. The identified homologs were used to construct the phylogenetic trees using the ML, BI and NJ methods (substitution model: TrNef + I + G) implemented in the TOPALi program [53]. The alignment gaps were trimmed using Gblocks [54,55]. The branches with a bootstrap value <80% were collapsed (Fig. 4). We further investigated the duplication and speciation events by reconciling the *PRAME* gene tree with a species tree obtained from NCBI taxonomy database [56] using Notung 2.6 [22] (Fig. 5). No *PRAME* ortholog was identified from the lineages beyond Eutheria, including opossum (6.8x genome coverage), platypus (6x), chicken (6.6x), frog (7.5x) and zebrafish (6.5x).

Lineage- and site-specific selection test

We conducted a pairwise dN and dS analysis (Dnasp version 5.0) [57] for the orthologs located on the same chromosome across species studied. The sequences with a pairwise dS value of <0.02 were removed, and the resulted 78 sequences were used for lineage-specific positive selection test [23] (Table S1). The median dN/dS ratio was calculated for different clades and compared by the Mann-Whitney test [58]. The 78 sequences were aligned by ClustalW [59] and the gaps were trimmed by Gblocks. The aligned segments included 912 positions of the original 2677 positions. We used the codeml program implemented in PAML package for the selection test. A simple model assuming a single dN/dS ratio for branches was compared with another model assuming free dN/dS ratio for all the branches (branch models). The likelihood ratio test (LRT) indicates that the dN/dS ratios are significantly varied among lineages (p<0.001, $2\Delta l = 309.20 > X^2_{0.001,df=152} = 211.62$). We conducted LRT for each branch using the branch-site models, model A null and model A [24]. The sites under positive selection detected by Bayes Empirical Bayes (BEB) analyses were retrieved when the LRTs were significant.

For the site-specific positive selection test [23,60], we focused on investigating the Clade I and Clade IIa, which were newly identified in this study. We established two datasets, one with the 12 sequences in Clade I and the other with the 12 sequences in

Clade IIa (Table S1). The Clade I dataset included 1290 aligned positions of the original 1677 positions; The Clade IIa dataset included 1065 bases of the original 1902 positions. PAML [23] and HyPhy [60] packages were used to detect the selection. We compared four different model pairs, M0 (one-ratio)/M3 (discrete), M1a (nearly neutral)/M2a (positive selection), M7 (beta)/M8 (beta and ω>1), and M8a (beta and $\omega_s = 1$)/M8 in PAML. Three methods, SLAC (Single Likelihood Ancestor Counting), FEL (Fixed Effects Likelihood) and REL (Random Effects Likelihood), implemented in HyPhy (Hypothesis Testing Using Phylogenies) package [60] were also used to detect the positive selection sites (Table S7). The protein model of the *PRAME* gene on BTA16 (GenBank acc. no. XM_001256020) was built by I-TASSER [61].

Supporting Information

Figure S1 Motif alignment between the bovine PRAMEY and the human PRAME on HSA22. The aliphatic sites of LXXLL motifs observed on the human PRAME on HSA22 [16,12] are conserved in the bovine PRAME(Y). These motif modifications are restricted to the aliphatic group, including the leucine to valine in the third and seventh motifs and leucine to isoleucine in the fourth motif. An exception is that the first leucine in the fifth motif was modified to the non-aliphatic phenylalanine. The colors in the alignment indicated different types of amino acids (White: Aliphatic sites; Red: Acidic sites; Cyan: Basic sites; Purple: Aromatic sites; Yellow: Cystenine). * The aliphatic site positions were annotated based on the PRAME on HSA22.
(TIF)

Figure S2 Alignment of the *ZNF280B/ZNF280A/PRAME* gene block across 17 species. The *ZNF280B/ZNF280A/PRAME* gene blocks are conserved in the syntenic regions in most mammals except the rodents where the block was rearranged in two different chromosomes (MMU4/10 and RNO5/20). This plot was generated based on the HSA22 assembly (hg19, Feb. 2009). The boxes represent ungapped alignments; the lines represent gaps. This plot was generated using blastz alignment from the UCSC genome browser (http://genome.ucsc.edu/).
(TIF)

Table S1 A list of BACs containing homologous *PRAMEY*.
(DOC)

Table S2 *PRAME/PRAMEY* homologs in the phylogenetic tree.
(DOC)

Table S3 Primers for (RT-) PCR and strand-specific qRT-PCR.
(DOC)

Table S4 Probes for *in situ* hybridization.
(DOC)

Table S5 Positively selected sites detected from branch-site model tests.
(DOC)

Table S6 Site-specific selection tests on the homologs in Clade I and Clade IIa.
(DOC)

Table S7 The integrative analysis of positively selected sites in Clade IIa.
(DOC)

Author Contributions

Conceived and designed the experiments: TCC YY WSL. Performed the experiments: TCC YY HY EFR. Analyzed the data: TCC YY WSL. Contributed reagents/materials/analysis tools: TCC YY HY AKB EFR WSL. Wrote the paper: TCC YY WSL.

References

1. Simpson AJG, Caballero OL, Jungbluth A, Chen Y, Old LJ (2005) Cancer/testis antigens, gametogenesis and cancer. Nat. Rev. Cancer 5: 615–625.
2. Ikeda H, Lethé B, Lehmann F, van Baren N, Baurain JF, et al. (1997) Characterization of an antigen that is recognized on a melanoma showing partial HLA loss by CTL expressing an NK inhibitory receptor. Immunity 6: 199–208.
3. Nelson PT, Zhang PJ, Spagnoli GC, Tomaszewski JE, Pasha TL, et al. (2007) Cancer/testis (CT) antigens are expressed in fetal ovary. Cancer Immun 7: 1.
4. Stevenson B, Iseli C, Panji S, Zahn-Zabal M, Hide W, et al. (2007) Rapid evolution of cancer/testis genes on the X chromosome. BMC Genomics 8: 129.
5. Peikert T, Specks U, Farver C, Erzurum SC, Comhair SAA (2006) Melanoma antigen A4 is expressed in non-small cell lung cancers and promotes apoptosis. Cancer Res 66: 4693–4700.
6. Ono T, Kurashige T, Harada N, Noguchi Y, Saika T, et al. (2001) Identification of proacrosin binding protein sp32 precursor as a human cancer/testis antigen. Proc. Natl. Acad. Sci. U.S.A 98: 3282–3287.
7. Duan Z, Duan Y, Lamendola DE, Yusuf RZ, Naeem R, et al. (2003) Overexpression of MAGE/GAGE genes in paclitaxel/doxorubicin-resistant human cancer cell lines. Clin. Cancer Res 9: 2778–2785.
8. Vodicka R, Vrtel R, Dusek L, Singh AR, Krizova K, et al. (2007) TSPY gene copy number as a potential new risk factor for male infertility. Reprod. Biomed. Online 14: 579–587.
9. Jakubiczka S, Schnieders F, Schmidtke J (1993) A bovine homologue of the human TSPY gene. Genomics 17: 732–735.
10. Yin Y, Li Y, Qiao H, Wang H, Yang X, et al. (2005) TSPY is a cancer testis antigen expressed in human hepatocellular carcinoma. Br J Cancer 93: 458–463.
11. McCarthy N (2005) PRAME in the frame. Nat Rev Cancer 5: 839.
12. Epping MT, Wang L, Edel MJ, Carlée L, Hernández M, et al. (2005) The human tumor antigen PRAME is a dominant repressor of retinoic acid receptor signaling. Cell 122: 835–847.
13. Birtle Z, Goodstadt L, Ponting C (2005) Duplication and positive selection among hominin-specific PRAME genes. BMC Genomics 6: 120.
14. Gibbs RA, Rogers J, Katze MG, Bumgarner R, Weinstock GM, et al. (2007) Evolutionary and biomedical insights from the rhesus macaque genome. Science 316: 222–234.
15. Ortmann CA, Eisele L, Nuckel H, Klein-Hitpass L, Führer A, et al. (2008) Aberrant hypomethylation of the cancer-testis antigen PRAME correlates with PRAME expression in acute myeloid leukemia. Ann. Hematol 87: 809–818.
16. Kastner P, Mark M, Leid M, Gansmuller A, Chin W, et al. (1996) Abnormal spermatogenesis in RXR beta mutant mice. Genes Dev 10: 80–92.
17. Kawano R, Karube K, Kikuchi M, Takeshita M, Tamura K, et al. (2009) Oncogene associated cDNA microarray analysis shows PRAME gene expression is a marker for response to anthracycline containing chemotherapy in patients with diffuse large B-cell lymphoma. J Clin Exp Hematop 49: 1–7.
18. NCBI. Available at: http://www.ncbi.nlm.nih.gov/. Accessed 6 October 2009.
19. Kapustin Y, Souvorov A, Tatusova T, Lipman D (2008) Splign: algorithms for computing spliced alignments with identification of paralogs. Biology Direct 3: 20.
20. Tamura K, Dudley J, Nei M, Kumar S (2007) MEGA4: Molecular Evolutionary Genetics Analysis (MEGA) software version 4.0. Mol. Biol. Evol 24: 1596–1599.
21. Durand D, Halldórsson BV, Vernot B (2006) A hybrid micro-macroevolutionary approach to gene tree reconstruction. J. Comput. Biol 13: 320–335.
22. Vernot B, Stolzer M, Goldman A, Durand D (2008) Reconciliation with non-binary species trees. J. Comput. Biol 15: 981–1006.
23. Yang Z (2007) PAML 4: phylogenetic analysis by maximum likelihood. Mol. Biol. Evol 24: 1586–1591.
24. Zhang J, Nielsen R, Yang Z (2005) Evaluation of an Improved Branch-Site Likelihood Method for Detecting Positive Selection at the Molecular Level. Molecular Biology and Evolution 22: 2472–2479.
25. Sawyer S (1999) GENECONV: A computer package for the statistical detection of gene conversion. Distributed by the author, Department of Mathematics, Washington University in St. Louis. Available at: http://www.math.wustl.edu/~sawyer.
26. Church DM, Goodstadt L, Hillier LW, Zody MC, Goldstein S, et al. (2009) Lineage-specific biology revealed by a finished genome assembly of the mouse. PLoS Biol 7: e1000112.
27. Long M, Betran E, Thornton K, Wang W (2003) The origin of new genes: glimpses from the young and old. Nat Rev Genet 4: 865–875.
28. Hughes JF, Skaletsky H, Pyntikova T, Graves TA, van Daalen SKM, et al. (2010) Chimpanzee and human Y chromosomes are remarkably divergent in structure and gene content. Nature Available at: http://www.ncbi.nlm.nih.gov/pubmed/20072128. Accessed 26 January 2010.
29. Yang Y, Chang T, Yasue H, Bharti AK, Retzel EF, et al. (2011) ZNF280BY and ZNF280AY: autosome derived Y-chromosome gene families in Bovidae. BMC Genomics 12: 13.
30. Skaletsky H, Kuroda-Kawaguchi T, Minx PJ, Cordum HS, Hillier L, et al. (2003) The male-specific region of the human Y chromosome is a mosaic of discrete sequence classes. Nature 423: 825–37.
31. Zhang L, Lu HHS, Chung W, Yang J, Li W (2005) Patterns of Segmental Duplication in the Human Genome. Mol Biol Evol 22: 135–141.
32. Kobe B, Deisenhofer J (1995) A structural basis of the interactions between leucine-rich repeats and protein ligands. Nature 374: 183–186.
33. Lindblad-Toh K, Wade CM, Mikkelsen TS, Karlsson EK, Jaffe DB, et al. (2005) Genome sequence, comparative analysis and haplotype structure of the domestic dog. Nature 438: 803–819.
34. Epping MT, Bernards R (2006) A causal role for the human tumor antigen preferentially expressed antigen of melanoma in cancer. Cancer Res 66: 10639–10642.
35. Minami N, Aizawa A, Ihara R, Miyamoto M, Ohashi A, et al. (2003) Oogenesin is a novel mouse protein expressed in oocytes and early cleavage-stage embryos. Biol. Reprod 69: 1736–1742.
36. Wang PJ, McCarrey JR, Yang F, Page DC (2001) An abundance of X-linked genes expressed in spermatogonia. Nat. Genet 27: 422–426.
37. Lahn BT, Page DC (1997) Functional coherence of the human Y chromosome. Science 278: 675–680.
38. Wang A, Yasue H, Li L, Takashima M, de León FAP, et al. (2008) Molecular characterization of the bovine chromodomain Y-like genes. Anim. Genet 39: 207–216.
39. Liu W-S, Wang A, Yang Y, Chang T, Landrito E, et al. (2009) Molecular characterization of the DDX3Y gene and its homologs in cattle. Cytogenet. Genome Res 126: 318–328.
40. Emes RD, Goodstadt L, Winter EE, Ponting CP (2003) Comparison of the genomes of human and mouse lays the foundation of genome zoology. Hum. Mol. Genet. 12: 701–709.
41. Isahakia MA (1988) Characterization of baboon testicular antigens using monoclonal anti-sperm antibodies. Biol. Reprod 39: 889–899.
42. Gunn SR, Bolla AR, Barron LL, Gorre ME, Mohammed MS, et al. (2009) Array CGH analysis of chronic lymphocytic leukemia reveals frequent cryptic monoallelic and biallelic deletions of chromosome 22q11 that include the PRAME gene. Leuk. Res 33: 1276–1281.
43. Bronson RA (1999) Antisperm antibodies: a critical evaluation and clinical guidelines. J. Reprod. Immunol 45: 159–183.
44. Lu J, Huang Y, Lu N (2008) Antisperm immunity and infertility. Expert Review of Clinical Immunology 4: 113–126.
45. Sambrook J, Russell D (2001) Molecular Cloning. Cold Spring Harbor Laboratory Press.
46. Del Mastro RG, Lovett M (1997) Isolation of coding sequences from genomic regions using direct selection. Methods Mol. Biol 68: 183–199.
47. Liu W-S, Mariani P, Beattie CW, Alexander LJ, Ponce De León FA (2002) A radiation hybrid map for the bovine Y Chromosome. Mamm. Genome 13: 320–326.
48. Altschul SF, Madden TL, Schäffer AA, Zhang J, Zhang Z, et al. (1997) Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic Acids Res 25: 3389–3402.
49. Wu TD, Nacu S (2010) Fast and SNP-tolerant detection of complex variants and splicing in short reads. Bioinformatics 26: 873–881.
50. Miller NA, Kingsmore SF, Farmer A, Langley RJ, Mudge J, et al. (2008) Management of High-Throughput DNA Sequencing Projects: Alpheus. J Comput Sci Syst Biol 1: 132.
51. Kiuchi S, Yamada T, Kiyokawa N, Saito T, Fujimoto J, et al. (2006) Genomic structure of swine taste receptor family 1 member 3, TAS1R3, and its expression in tissues. Cytogenet. Genome Res 115: 51–61.

52. Rice P, Longden I, Bleasby A (2000) EMBOSS: the European Molecular Biology Open Software Suite. Trends Genet 16: 276–277.
53. Milne I, Lindner D, Bayer M, Husmeier D, McGuire G, et al. (2009) TOPALi v2: a rich graphical interface for evolutionary analyses of multiple alignments on HPC clusters and multi-core desktops. Bioinformatics 25: 126–127.
54. Castresana J (2000) Selection of conserved blocks from multiple alignments for their use in phylogenetic analysis. Mol. Biol. Evol 17: 540–552.
55. Talavera G, Castresana J (2007) Improvement of phylogenies after removing divergent and ambiguously aligned blocks from protein sequence alignments. Syst. Biol 56: 564–577.
56. Taxonomy Common Tree. Available at: http://www.ncbi.nlm.nih.gov/Taxonomy/CommonTree/wwwcmt.cgi.
57. Librado P, Rozas J (2009) DnaSP v5: a software for comprehensive analysis of DNA polymorphism data. Bioinformatics 25: 1451–1452.
58. Mann HB, Whitney DR (1947) On a Test of Whether one of Two Random Variables is Stochastically Larger than the Other. The Annals of Mathematical Statistics 18: 50–60.
59. Larkin MA, Blackshields G, Brown NP, Chenna R, McGettigan PA, et al. (2007) Clustal W and Clustal X version 2.0. Bioinformatics 23: 2947–2948.
60. Pond SLK, Frost SDW, Muse SV (2005) HyPhy: hypothesis testing using phylogenies. Bioinformatics 21: 676–679.
61. Zhang Y (2008) I-TASSER server for protein 3D structure prediction. BMC Bioinformatics 9: 40.