Conserved architecture of RAD51 recombinase in ruminants revealed through molecular cloning and characterization

SONIKA AHLAWAT¹, REKHA SHARMA², REENA ARORA³, LATIKA JAISWAL⁴, MEENU CHOPRA⁵, PRIYANKA SHARMA⁶ and SACHINANDAN DE⁷

National Bureau of Animal Genetic Resources, Karnal, Haryana 131 002 India

Received: 21 March 2016; Accepted: 6 June 2016

ABSTRACT

Meiotic recombination is a highly conserved process that ensures accurate segregation of homologous chromosomes and contributes to genetic variability of a species to withstand the pressure of natural selection. RAD51 recombinase plays a pivotal role in double strand break repair during homologous recombination and also ensures that chromosomal integrity is maintained during meiotic cell cycle. The present study was aimed to clone and sequence characterize RAD51 gene and its paralogs from four ruminant species (buffalo, cattle, sheep and goat). The complete open reading frame of RAD51 gene was observed to be 1020 nucleotides in length encoding a putative protein of 339 amino acids. Functional feature prediction by SMART and Pfam revealed highly conserved DNA binding helix-hairpin-helix motif in the N terminal domain and Walker A and Walker B motifs in the catalytic domain which confer ATP binding and hydrolysis activity, respectively. All RAD51 orthologs and paralogs (RAD51B, RAD51C, RAD51D, XRCC2 and XRCC3) were subjected to phylogenetic analysis. Evolutionary analysis between the different orthologs and paralogs suggested that these proteins are conserved from yeast to humans since all the orthologs from different species clustered in their respective clade, clearly demarcating the paralogs. Gene duplication events have expanded the RAD51 family in eukaryotes offering them increased capacity to repair DNA and promote homologous recombination.

Key words: Characterization, Homologous recombination, RAD51, Ruminants

Meiosis is an evolutionarily conserved process in all sexually reproducing living organisms which contributes to genetic diversity through crossing-over. Meiotic recombination is an essential component for successful completion of meiosis which ensures accurate segregation of homologous chromosomes and generates haploid gametes from diploid precursor cells (Keeney 2008). Recombinational interactions promote the pairing of homologs and their end-to-end connection by zipper-like structures called synaptonemal complexes (Lui et al. 2006, Tsubouchi and Roeder 2003). In the recombination event, nuclease processing of double strand break ends generates single-stranded tails, which then assemble into nucleoprotein filaments comprising recA-family proteins and their accessory factors (Zakharyevich et al. 2010, Shinohara and Shinohara 2004). These filaments mediate DNA homology search and strand invasion of a homologous template chromosome to form joint molecule intermediates (Shinohara et al. 1992, Hong et al. 2001).

Bacterial recA was the first recombination gene to be cloned from Escherichia coli (Lusetti and Cox 2002). Meiotic recombination in eukaryotic cells generally requires two homologues of the E. coli recA protein - RAD51 (RAD51 recombinase) and DMC1 (disrupted meiotic cDNA) which were first identified in Saccharomyces cerevisiae (Bishop et al. 1992, Shinohara et al. 1992). RAD51 is required for homologous recombination (HR) pathways in both mitotic and meiotic cells (Krogh and Symington 2004) whereas DMC1 is a meiosis-specific protein (Bishop et al. 1992). Cooperative action of RAD51 and DMC1 during meiotic recombination has been established because of their co-localization in side-by-side foci on meiotic chromatin (Bishop 1994). Although both proteins are essential for meiosis, a recent report on budding yeast, however, suggests that RAD51 does not seem to participate directly in strand exchange, but facilitates DMC1-mediated DNA strand-exchange (Cloud et al. 2012). On the contrary, in the fission yeast, deletion of DMC1 causes only a moderate reduction in crossing over and a slight reduction in fertility. This suggests that RAD51 can partially substitute for DMC1 in inter-homolog recombination (Grishchuk and Kohli 2003). In addition to playing important role in homologous recombination,
**RAD51** also ensures that chromosomal integrity is maintained in meiotic cell cycle (Li et al. 2006). The disruption of **RAD51** gene can lead to cell death, embryo inviability and chromosomal aberrations in mouse (Tsuzuki et al. 1993).

**RAD51** orthologs have been well characterized in mice and humans based upon their identity to the yeast **RAD51** gene (Morita et al. 1993, Shinohara et al. 1993). Mouse **RAD51** cDNA was found to complement *S. cerevisiae **RAD51*** mutation (Morita et al. 1993) and the human homolog was shown to have ATP-dependent strand transfer activity (Baumann et al. 1996). Mammals possess five **RAD51** paralogs: **RAD51B**, **RAD51C**, **RAD51D**, **XRCC2** and **XRCC3** (Suwaki et al. 2011). These paralogs share 20–30% identity at the amino acid level with **RAD51** and with each other. All **RAD51**-like proteins share a highly conserved 230 amino acid long central domain (RecA_**RAD51** domain) which has two highly conserved consensus motifs, Walker A and Walker B, which are characteristic of ATPases and confer ATP binding and hydrolysis activities, respectively. **RAD51** also has a non-specific DNA-binding domain called helix-hairpin-helix (HhH) motif which is absent in other paralogs (Lin et al. 2006).

Although **RAD51** has been shown to have pivotal roles in meiosis, homologous recombination and double strand break (DSB) repair but it has not yet been characterized in mammals other than humans and mice. There are no reports available on characterization of **RAD51** in ruminants. Keeping in view the evolutionary divergence among mammalian species, the present study was aimed to clone and sequence characterize **RAD51** gene in four ruminant species (buffalo, cattle, sheep and goat) and to determine the phylogenetic distribution of **RAD51** and its paralogs across a diverse range of eukaryotic species.

**MATERIALS AND METHODS**

**Sample collection:** Testis tissue samples of *Bubalus bubalis* (buffalo), *Bos indicus* (cattle), *Ovis aries* (sheep), *Capra hircus* (goat) were collected from slaughter house (New Delhi and Kolkata, India) with the help of an on-site veterinary officer. The fresh tissue samples were dispensed in RNA later and stored in deep freeze (−80°C) until further use.

**RNA isolation and cDNA preparation:** Total RNA was isolated from tissue samples using Trizol reagent and treated with DNase I following manufacturer’s instructions. RNA integrity was assessed by loading 2 µl RNA sample on 1.5% agarose gel. The concentration of isolated RNA was determined by measuring optical density at 260 nm using NanoDrop 1000 Spectrophotometer. About 1 µg of total mRNA was reverse transcribed to cDNA with Superscript III cDNA synthesis kit using hexamer primers according to manufacturer’s instructions and stored at −80°C till further use.

**RT-PCR and cloning:** The full length coding sequence of **RAD15** gene was amplified from cDNA of testicular tissue. The primers for this study were designed manually using the conserved sequences from *Bos taurus* mRNA sequence (NM_001046179). Primer sequences used for amplification are as follows: F CTAAGTAATGGCTATGCAAATGCGAGTTGAAA and R TCGTCTTTGGCATTCTCCGACTTCCA. PCR was carried out on a Veriti 96 well thermal cycler in a 25 µl reaction mixture containing 2.5 µl of 10’x Taq reaction buffer, 0.5 µl of 10 mM dNTP mix, 0.5 µl of 10 PM for each primer and 0.25 units of Taq DNA polymerase. The PCR reaction cycle was accomplished by denaturation for 3 min at 94°C; 30 cycles of denaturation at 94°C for 30 sec, annealing at 56°C for 30 sec, extension step at 72°C for 30 sec, with a final extension at 72°C for 5 min. PCR products were electrophoresed alongside DNA molecular weight marker in 1.5% agarose gel and then stained with ethidium bromide. The amplified fragments were gel purified using gel extraction kit and subcloned into pTZ57R/T vector. Recombinant clones were selected and plasmid DNA was extracted. Multiple clones were sequenced using the Sanger sequencing to obtain the coding DNA sequence of **RAD51** in four ruminant species.

**Data extraction and sequence analysis:** **RAD51** nucleotide sequences were used as query in BLAST search against NCBI (http://www.ncbi.nlm.nih.gov/sites/entrez) database to collect all possible orthologous and paralogous sequences in different species of eukaryotes (yeast, plants and animals). BLASTp (Altschul et al. 1990) search was performed using protein sequences corresponding to **RAD51** gene to retrieve similar sequences from other species such as *Bos taurus*, *Bos mutus*, *Bubalus bubalis*, *Capra hircus*, *Ovis aries*, *Gallus gallus*, *Pan troglodytes*, *Homo sapiens*, *Mus musculus* etc. Percentage identity matrix was obtained by pair wise alignment of all the sequences with each other using clustal omega (Sievers et al. 2011). SMART (http://smart.embl-heidelberg.de) and Pfam (pfam.sanger.ac.uk/search) were used to predict the functional domains of the **RAD51** protein.

**Evolutionary analysis:** Multiple protein sequence alignment was performed using MAFFT (http://mafft.cbrc.jp/alignment/server/) and Biodeit. Phylogenetic analysis was performed using Neighbour Joining and Maximum Likelihood algorithm with MEGA 6.0 package (http://www.megasoftware.net) and in each case, branch

![Fig. 1. Gel image of RT-PCR product of **RAD51** gene in ruminants.](image-url)
Confidence value was calculated with bootstrapping with 1000 iterations using pairwise deletion option of amino acid sequences with gamma parameters. Here, the Jones, Taylor and Thornton (JTT) model for amino acid sequences and gamma parameters was used (Jones et al. 1992).

RESULTS AND DISCUSSION

Homologous recombination is an enigmatic process...
which not only creates genetic diversity but also provides an important way of repairing DNA damage without errors. Meiotic recombination is initiated by the introduction of programmed double-strand breaks (DSBs) catalyzed by an evolutionarily conserved, meiosis-specific, topoisomerase-like protein, SPO11 (Keeney 2008). Resection of the 5’ ends of the breaks results in 3’ single stranded tails, which are bound by 2 conserved recA-like proteins, RAD51 and DMC1 in many organisms, including yeast and mammals which are involved in DNA repair (Bishop et al. 1992, Shinohara et al. 1992). *Escherichia coli* recA gene is among the most studied recombination gene which encodes a DNA-dependent ATPase that binds to single-stranded DNA and promotes strand invasion and exchange between homologous DNA molecules. Eukaryotic recA homologs are RAD51 and DMC1, which were first discovered in the budding yeast *Saccharomyces cerevisiae* (Shinohara et al. 1992).

In the present study, the complete open reading frame of RAD51 gene was amplified by RT-PCR from the mRNA isolated from testis tissue of buffalo, cattle, sheep and goat. The amplified products were sequenced in both directions after cloning in pTZ57R/T vector and the DNA sequences from these 4 species were submitted to NCBI and are available under accession numbers: KR935222, KR935223, KR935224 and KR935225 for buffalo, cattle, sheep and goat, respectively. The complete ORF of RAD51 gene in all the four species was observed to be 1020 nucleotides in length (Fig. 1) encoding a putative protein of 339 amino acids. Comparative analysis of cDNA sequence of 4 ruminant orthologs revealed limited sequence divergence at both the nucleotide and the amino acid level indicating a high degree of evolutionary conservation. The sequence identity at nucleotide level between cattle and buffalo was 97.5% whereas between sheep and goat, it was 99.6%, as determined by percentage identity matrix in Clustal Omega. The homology of sheep and goat RAD51 nucleotide sequence was more with that of cattle (98.1 and 98.3%, respectively) as compared to buffalo (97.7 and 97.9% respectively).

![Fig. 3. Phylogenetic tree of RAD51 and its paralogs in eukaryotes constructed by neighbour method (Poission correction with gamma parameters).](75)
Multiple sequence alignment of the ruminant RAD51 protein with other orthologs revealed highly conserved structure of this protein in different species (Fig. 2). Functional feature prediction by SMART and PDBsum also revealed significant identity of cattle, buffalo, sheep and goat RAD51 protein with humans as well as other animal species. RAD51 protein in humans and mice typically has two major domains: N terminal domain (~70 AA) and Catalytic domain (~240 AA). We also observed these domains to be present in the RAD51 protein of ruminants. The N terminal domain revealed the putative double stranded DNA binding helix-hairpin-helix (HhH) motif which is characteristic of the RAD51 protein (Fig. 2). The catalytic domain has Walker A and Walker B motifs which are present in ATPases and confer ATP binding and hydrolysis activity respectively (Walker et al. 1982). These motifs were also found to be highly conserved. Put together, a high degree of evolutionary conservation was evident between ruminant sequences generated by us as well as those retrieved from NCBI.

Numerous gene duplication events have expanded the RAD51 family in vertebrate animals producing 5 RAD51 paralogs (RAD51B, RAD51C, RAD51D, XRCC2 and XRCC3) and all the corresponding proteins have acquired new functionalities (Suwaki et al. 2011). All the RAD51 paralogs have DNA stimulated ATPase activity and are important for homologous recombination and DNA repair. Animal cells defective in any of the RAD51 paralogs are hypersensitive to DNA cross-linking agents, such as Cisplatin and Mitomycin C, and show spontaneous chromosomal aberrations (Takata et al. 2001). Embryonic lethality is observed in mouse if any of the RAD51 paralog is individually knocked out suggesting that these paralogs are imperative for animals’ viability (Deans et al. 2000, Tsuzuki et al. 1996). For instance, disruption of XRCC2 in mice resulted in embryonic lethality occurring from mid-gestation but the embryos that survived until later stages of embryogenesis showed developmental abnormalities and died at birth. Genetic instability was evident in embryonic cells with a high degree of chromosomal aberrations (Deans et al. 2000). Similarly, haploinsufficiency of RAD51B has been reported to cause centrosome fragmentation and aneuploidy in human colon cancer cell line (Date et al. 2006). This suggests that all RAD51 paralogs play a key role in preserving genomic integrity during the early stages of development, as well as during physiologically normal cell proliferation.

It was validated by 2 hybrid and immunoprecipitation experiments that the 5 RAD51 paralogs form 2 complexes: a heterotetramer composed of RAD51B, RAD51C, RAD51D and XRCC2 referred to as the BCXD2 complexes and a heterodimer called CX3 composed of RAD51C and XRCC3 (CX3). These complexes facilitate the binding of RAD51 with single stranded DNA and hence promote homologous recombination (Masson et al. 2001, Schild et al. 2000). Genetic complexity of the eukaryotic organisms dictates the need for functional innovations for each member of this family. RAD51 paralogs, when present in a different complex, function at least in part independently in assisting RAD51 mediated recombination. They have evolved to perform specific functions during meiotic and mitotic recombination but all are integral components of DNA damage signaling pathways which are critical for efficient DNA damage repair, thus conforming to a single RAD51 gene family.

Amongst all the 5 paralogs, RAD51C is thought to play more prominent cellular functions since it is the component of both BCX2D and CX3 complexes. In fact, RAD51C plays an important role in DNA damage response, acting as a transducer of the damage signal to ensure engagement of the homologous recombination pathway of repair (Badie et al. 2009). RAD51C or RAD51C-associated proteins were also proposed to be components of Holliday junction resolvase (Liu et al. 2004) and hence are involved in the processing of the Holliday junction in mammalian cells. RAD51C and XRCC3 play key roles in the homology search together with RAD51, and were shown to act as cofactors facilitating loading of RAD51 onto DNA (Bishop et al. 1998, Badie et al. 2009). XRCC3 is shown to be required for stabilizing RAD51C possibly by forming dimers, and this function requires the ATPase activity, suggesting that XRCC3 may regulate the dimerization through its ATP binding and hydrolysis activity (Yamada et al. 2004).

Biochemical evidence has suggested that RAD51B may play a role in the late phase of homologous recombination by binding to the Holliday junction (Yokoyama et al. 2003). Chicken DT40 cells deficient in RAD51B show hypersensitivity to DNA-damaging agents, impaired damage dependent RAD51 focus formation, increase in chromosome aberrations and decrease in sister chromatid exchange (Takata et al. 2001). While, RAD51D is involved in promoting HR reactions at telomeres to the extent that the cells deficient in it show loss of telomere capping and telomere shortening in the presence of telomerase, consistent with a role for homologous recombination (HR) in telomere protection (Tarsounas et al. 2004).

To analyze the level of conservation and evolutionary relationship of the RAD51 with its paralogs, all the RAD51 orthologs and paralogs from a diverse range of species were subjected to phylogenetic analysis. The phylogenetic tree generated by both neighbour joining and maximum likelihood methods resolved into 4 main branches. The XRCC2, RAD51D, RAD51B members formed 3 separate branches whereas RAD51, RAD51C and XRCC3 clustered into a fourth branch (Fig. 3). The phylogeny separated all the 6 genes into different and distinct clades. XRCC2 is the first to separate amongst all the paralogs followed by RAD51D and then by RAD51B, RAD51, XRCC3 and RAD51C form well supported sister groups in the phylogenetic tree. All the orthologous sequences of other eucaryotic species for all the genes of the RAD51 family were components of their respective clades without any admixing indicating sufficient divergence between the paralogs across all taxa. Plants form a separate cluster in
all the clades being well separated from animal species with good bootstrap support value. Humans and chimpanzees always clustered together for all the genes in their respective clades. The same picture was seen in case of rat, mouse and Chinese hamster indicating sequence similarity between the orthologs for all the paralogs.

Evolutionary analysis of ruminant RAD51 with other animals, humans, fishes, plants and yeast clearly separated Saccharomyces cerevisiae (yeast) gene from all other sequences. Plants such as Arabidopsis thaliana and Solanum lycopersicum formed a separate cluster (with 100% bootstrap support) which was quite distinct from all other animal species. Danio rerio (Zebra fish) is the first to separate from the group followed by Xenopus tropicalis (frog) and Gallus gallus (Jungle fowl) indicating less sequence similarity as compared to homology between different mammals. As is expected, Homo sapiens and Pan troglodytes clustered together with 85% confidence value. All the ruminants (our sequences as well as sequences retrieved from NCBI) clustered together indicating high sequence similarity.

Lin et al. (2006) conducted extensive searches for RAD51-like genes from public databases and performed detailed phylogenetic analyses of the genes identified. They proposed that RAD51 family members can be categorized into three distinct groups, designated as the recA, RAD50 and RADβ subfamilies. The recA subfamily includes members from eubacteria, plants, and protists, whereas the RADα and RADβ subfamilies each constitute genes from archaea and eukaryotes. RADβ subfamily was proposed to contain RAD51 and DMC1 genes, both of which are important enzymes of recombination. RADβ subfamily members form separate monophyletic group which is different from the monophyletic group formed by RADα subfamily members viz. RAD51C, XRCC3, RAD51B, RAD51D, and XRCC2 genes. This research group went on to state that these five eukaryotic RAD51-like genes were derived from a single ancestral RADβ gene by successive duplication events which probably occurred before the divergence of plants from fungi and animals. The phylogenetic tree constructed with RADα and RADβ subfamilies showed that these subfamilies form two separate groups. In the RADβ subfamily, the RAD51C group first separates from the other four eukaryotic RADβ genes, and XRCC3 is next to follow. RAD51D and XRCC2 genes formed two well-supported sister groups indicating that they have emerged most recently among the eukaryotic RADβ genes.

However, our analysis indicated that XRCC2 is the first to separate from the group suggesting most distinct structure and function. RAD51D and subsequently RAD51B separate from other members. In fact, in our study, RAD51, XRCC2 and RAD51C seem to be members of one sub-family by virtue of sharing highest sequence similarity. This discrepancy in the evolutionary analysis could be because Lin et al. (2006) compared sequences of humans, chicken, zebrafish, sea urchin, Arabidopsis and rice which encompass members of different taxa literally. On the contrary, in our study we have analyzed different animal species and kept plants as an out-group for comparison. However, we agree with the observations of Lin et al. (2006) who reported that flowering plants Arabidopsis thaliana and rice (Oryza sativa) possess four conserved recA-like genes that have higher levels of sequence similarity with eubacterial recA genes than with the eukaryotic RAD51-like genes. This observation was also seen in our analysis since all plant RAD51 genes formed separate cluster that was distinct from animal RAD51 genes sequences.

The mechanism of double-strand break repair process can be elucidated by understanding the effect of ATP hydrolysis on the assembly of DNA repair complexes. RAD51 undergoes conformational changes during ATP binding or hydrolysis that regulates its polymerization on single stranded DNA. Similarly, other RAD51-related protein complexes are also controlled by this small effector molecule (Namsaraev and Berg 1998, Kim et al. 2002). ATP stimulates the binding of RAD51D-XRCC2 to single-stranded DNA (Kurumizaka et al. 2002) and ATP hydrolysis is necessary for normal dynamics of the RAD51C-XRCC3 complex (Yamada et al. 2004). Because of the important roles that Walker A and Walker B motifs play in DNA interstrand crosslink repair, their structure and function have been conserved in all RAD51 paralogs. Functional feature prediction by SMART and Pfam in our study in ruminants confirmed the presence of Walker A, Walker B and HhH motifs which are characteristic of the RAD51 protein. These domains were also identified in other orthologs by in silico analysis which reinforces that the structure and function of this protein is conserved across different taxa such as chicken, rabbit, humans, mice and ruminants. Put together, our study first time reports the cloning and characterization of RAD51 in four ruminant species and establishes conserved architecture of this recombinase across diverse species.

ACKNOWLEDGEMENT

This study was supported by the Indian Council of Agricultural Research, New Delhi.

REFERENCES


recombination, synaptonemal complex formation, and cell cycle progression. 


Bishop D K. 1994. RecA homologs DMC1 and RAD51 interact to form multiple nuclear complexes prior to meiotic chromosome synopsis. 

Cell 79: 1081–92.


Science 337: 1222–25.


Homologous pairing and ring and filament structure formation of mutation data matrices from protein sequences. 

Genes and Development 14: 211–23.

Kim H K, Morimatsu K, Norden B, Ardhammar M and Takahashi Hong E L, Shinohara A and Bishop D K. 1994. RecA homologs DMC1 and RAD51 interact to promote renaturation of single-strand DNA (ssDNA) and assimilation of ssDNA into homologous super-coiled duplex DNA. 


Genes and development 16: 402–12.


Computer Applications in Biosciences 8: 275–82.


Genome Dynamics and Stability 2: 81–123.


Genes Cells 7: 1125–34.


Li W and Ma H. 2006. Double-stranded DNA breaks and gene functions in recombination and meiosis. 


Cell 69: 457–70.


Nature Genetics 4: 239–43.


Molecular Systems Biology 7: 539–44.


Cell 117: 337–47.


EMBO Journal 1: 945–51.


