RESEARCH ARTICLE   **Open Access**

Nature of selection varies on different domains of IFI16-like PYHIN genes in ruminants

Sushil Kumar, Jatinder Singh Chera, Ashutosh Vats and Sachinandan De[*]

Abstract

Background: ALRs (AIM2-like Receptors) are germline encoded PRRs that belong to *PYHIN* gene family of cytokines, which are having signature N-terminal PYD (Pyrin, PAAD or DAPIN) domain and C-terminal HIN-200 (hematopoietic, interferon-inducible nuclear protein with 200 amino acid repeat) domain joined by a linker region. The positively charged HIN-200 domain senses and binds with negatively charged phosphate groups of single-stranded DNA (ssDNA) and double-stranded DNA (dsDNA) purely through electrostatic attractions. On the other hand, PYD domain interacts homotypically with a PYD domain of other mediators to pass the signals to effector molecules downwards the pathways for inflammatory responses. There is remarkable inter-specific diversity in the numbers of functional *PYHIN* genes e.g. one in cow, five in human, thirteen in mice etc., while there is a unique loss of *PYHIN* genes in the bat genomes which was revealed by Ahn et al. (2016) by studying genomes of ten different bat species belonging to sub-orders *yinpterochiroptera* and *yangochiroptera*. The conflicts between host and pathogen interfaces are compared with "Red queen's arms race" which is also described as binding seeking dynamics and binding avoidance dynamics. As a result of this never-ending rivalry, eukaryotes developed PRRs as antiviral mechanism while viruses developed counter mechanisms to evade host immune defense. The *PYHIN* receptors are directly engaged with pathogenic molecules, so these should have evolved under the influence of selection pressures. In the current study, we investigated the nature of selection pressure on different domain types of *IFI16-like* (*IFI16-L*) *PYHIN* genes in ruminants.

Results: Three transcript variants of the *IFI16-like* gene were found in PBMCs of ruminant animals-water buffalo, zebu cattle, goat, and sheep. The *IFI16-like* gene has one N-terminal PYD domain and one C-terminal HIN-200 domain, separated by an inter-domain linker region. HIN domain and inter-domain region are positively selected while the PYD domain is under the influence of purifying selection.

Conclusion: Herein, we conclude that the nature of selection pressure varies on different parts (PYD domain, HIN domain, and inter-domain linker region) of *IFI16-like PYHIN* genes in the ruminants. This data can be useful to predict the molecular determinants of pathogen interactions.

Keywords: PAMP, DAMP, PRR, IFI16, IFI16-like, PYHIN, PYD, HIN, Inter-domain, Selection

Background

Immune defense mechanism in mammalian physiology is a highly coordinated process, which activates various mechanisms for local and systemic immunity. The first line of defense, the innate immune system, relies on Pattern Recognition Receptors (PRRs) to identify Pathogen-Associated Molecular Patterns (PAMPs) and Damage-Associated Molecular Patterns (DAMPs) for immune response generation [1–4]. PAMPs and DAMPs are the signature molecules of pathogens and damaged host cells respectively. Five categories of PRRs have been characterized in the mammalian cells which can be divided into two groups based on whether they are membrane-bound or free. Membrane-bound PRRs are *Toll-like Receptors (TLRs)* and *C-type Lectin Receptors (CLRs)*. PRRs which reside and function inside the cellular compartments include *NOD-like Receptors (NLRs), RIG1-like receptors (RLRs)* and *AIM2-like receptors*

* Correspondence: sachinandan@gmail.com
Animal Genomics Laboratory, Animal Biotechnology Centre, National Dairy Research Institute, Karnal, Haryana 132001, India

(ALRs) [5–7]. These PRRs work in harmony to identify various forms of PAMPs and DAMPs (nucleic acids, proteins, lipids, and carbohydrates) inside the cell and initiate signaling cascade which leads to the expression of interferons and proinflammatory cytokines [8]. Any impairment of this molecular harmony can disturb physiological homeostasis and lead to dire consequences in the host.

A DNA molecule can be categorized in either DAMPs (if it comes from the damaged host cell or mitochondria) or PAMPs (if it originates from intracellular or extracellular pathogens). A number of proteins have been proposed as putative PRRs for DNA. These PRRs include- *Absent in Melanoma 2* (*AIM2*), *Interferon-γ inducible protein 16* (*IFI16*), cyclic *GMP- AMP synthase* (*cGAS*), *Z-DNA binding protein1* (*DAI*), *Toll-like Receptor 9* (*TLR9*) and *RNA pol III*. The presence of so many DNA sensors is still an unraveled mystery [9]. Among all these, *IFI16* and *AIM2*, which come under the category of *ALRs*, function in the recognition of viral DNA inside the cellular compartments [5, 10, 11]. *ALRs* are germline encoded PRRs that belong to the *PYHIN* gene family of cytokines, which have signature N-terminal PYD (Pyrin, PAAD or DAPIN) domain and a C-terminal HIN-200 (hematopoietic, interferon-inducible nuclear protein with 200 amino acid repeat) domain joined by a linker region [12, 13]. The positively charged HIN-200 domain senses and binds with the negatively charged phosphate groups of single-stranded DNA (ssDNA) and double-stranded DNA (dsDNA) purely through electrostatic attractions. On the other hand, the PYD domain interacts homotypically with PYD domains of other mediators to pass the signals to the effector molecules downstream of the pathways for inflammatory responses [14]. The role of inter-domain linker region is proposed in the alignment of PYD and HIN-200 domains for properly achieving their respective 3D structures and functions.

The syntenic *PYHIN* gene cluster region among mammalian species is variable in its size, number and gene arrangements. In addition, it also possesses signature border markers of *cell adhesion molecule 3 (CADM3)* genes on one end, and *olfactory receptors* and *spectrin alpha chain (SPTA1)* genes on the other end in all mammals [15]. This genomic region is believed to have evolved through multiple events of gene duplications, recombination, and species-specific family expansions/losses. There is also a remarkable inter-specific diversity in the numbers of functional *PYHIN* genes which varies with one in *cow*, five in *human*, thirteen in *mouse* (Additional file 1), and a unique loss of *PYHIN* genes in the genomes of ten different *bat* species belonging to sub-orders yinpterochiroptera and yangochiroptera [15–17].

The *PYHIN* receptors directly engage with pathogenic molecules, they may have evolved under selective pressure from the pathogens [18, 19]. The conflicts between host and pathogen interfaces are compared with "Red queen's arms race" which is also described as binding seeking dynamics and binding avoidance dynamics [20]. According to this hypothesis, host restriction factors adapt positively to binding to viral factors and on the other hand, due to the course of evolution, viruses also develop mechanisms to avoid host immune responses. As a result of this never-ending rivalry, eukaryotes developed PRRs as antiviral mechanism while viruses developed counter mechanisms to evade host immune defense. For example, *HSV-1* encoded immediate-early protein *ICP0*, an *E3 ubiquitin ligase*, binds and directs the proteasomal degradation of *IFI16*. Another protein *pUL83* encoded by *HCMV* binds with PYD domain of *IFI16* and blocks its multimerization.

In the current study, we investigated the nature of selection pressure on different domain types of *IFI16 like* (*IFI16-L*) *PYHIN* genes in ruminants of the order Artiodactyla; all members of which are predicted to possess single functional *PYHIN* gene in their genomes. The inflammasome forming *AIM2* is present as the pseudogene in the genomes of the ruminant animals. This gene became inactive due to accumulations of deletion, frame-shift and stop codons [16]. The *ALR* family member *IFI16* is thought to act as an important innate immune PRR for viral DNA inside cytoplasmic and nuclear compartments of the cell. It detects and binds with viral dsDNA as well as viral ssDNA stem-loop structures [21]. Various studies have shown direct engagements of *IFI16* with *herpesviruses* in cellular compartments of a variety of human cell types [22]. *Bovine herpesviruses* (DNA viruses) are responsible for a variety of disease conditions in both domestic and wild ruminants and cause a severe economic loss in the form of abortion, reduced production and impaired work abilities [23, 24]. *IFI16* induces inflammasome and *interferon-β* through *PYCARD* and *STING* mediated pathways respectively [4, 25–27]. It is also reported to acts as viral restriction factor through epigenetic modifications [28–30]. *Herpesviruses* cause both lytic infections and latent infections. *IFI16* has a critical role in herpesviral latency, which makes it a potential therapeutic target against herpesviral infections [31]. The characterization and evolutionary analysis of *IFI16-L* gene in ruminant animals will help to identify residues which directly engage with viruses and will give insights into molecular markers of disease resistant/susceptibility against DNA viruses including *herpesviruses* [32]. This will further be helpful in designing of innate immunologicals and vaccine strategies [33].

Herein, we conclude that the nature of selection pressure varies on different parts (PYD domain, HIN domain, and the inter-domain linker region) of *IFI16-like PYHIN* genes in the ruminants.

Kumar *et al. BMC Evolutionary Biology*   (2019) 19:26

Page 3 of 11

Results

IFI16-like PYHIN genes have multiple alternatively spliced transcript variants

In the present study, the *IFI16-like ALR* gene sequences were selectively amplified, cloned and sequenced from Indian cattle (*Bos indicus*), water buffalo (*Bubalus bubalis*), goat (*Capra hircus*) and sheep (*Ovis aries*). The transcript variant 'C' in *buffalo* and *sheep*; and 'B' transcript of *goat* lacks HIN domain while the cattle transcript variant 'C' lacks PYD domain and a putative nuclear localization signal (Fig. 1).

IFI16-like ALR is a shuttle protein like human ALR IFI16

Through Immuno-cytochemistry in buffalo fibroblasts, we found that *IFI16-like PYHIN* protein was localized in both nuclear and cytoplasmic compartments (Fig. 2). For shuttling between nucleus and cytoplasm, it contains a nuclear localization signal (NLS) at its N-terminal which is partially fused with PYD domain (Fig. 1). Single putative NLS motif found in Buffalo (82 - LRKEKLKVLKKS-KAK – 96), sheep (83 - RKEKLKAIRRKKVK – 96) and goat (83 - RKEKLKAIKKNKAK – 96) while two putative NLS motif sequences (82 - LRKEKLKVIKKNKAK – 96, 131 - KKKRTTKTNESKRR – 144) were identified in the *IFI16-L* gene of zebu cattle.

Phylogenetic analysis of IFI16-L gene domains reveals their evolutionary relatedness to human IFI16 domains

The sequences of ALRs from different species which represented different mammalian orders were collected from NCBI for phylogenetic analysis. The inter-domain linker region between PYD and HIN domains is highly variable, therefore, phylogenetic analysis of whole genes and of the inter-domain (ID) linker region alone in unrelated species might produce misleading results [15, 16]. So, all the domain sequences and inter-domain sequences were used separately for the alignment and analysis. Three separate phylograms were generated based on their amino acid alignments by maximum likelihood (ML) approaches (Figs. 3, 4 and 5). Phylogenetic analyses were conducted by "one click" mode at *Phylogeny.fr* suite of programs [34, 35] using PhyML 3.0 with 100 bootstrap replicates and the trees were drawn using TreeDyn 198.3 software. Trees of domains i.e. PYD and HIN were rooted with the marsupial Tasmanian devil (*Sarcophilus harrisii*) and their topologies were found to be congruent
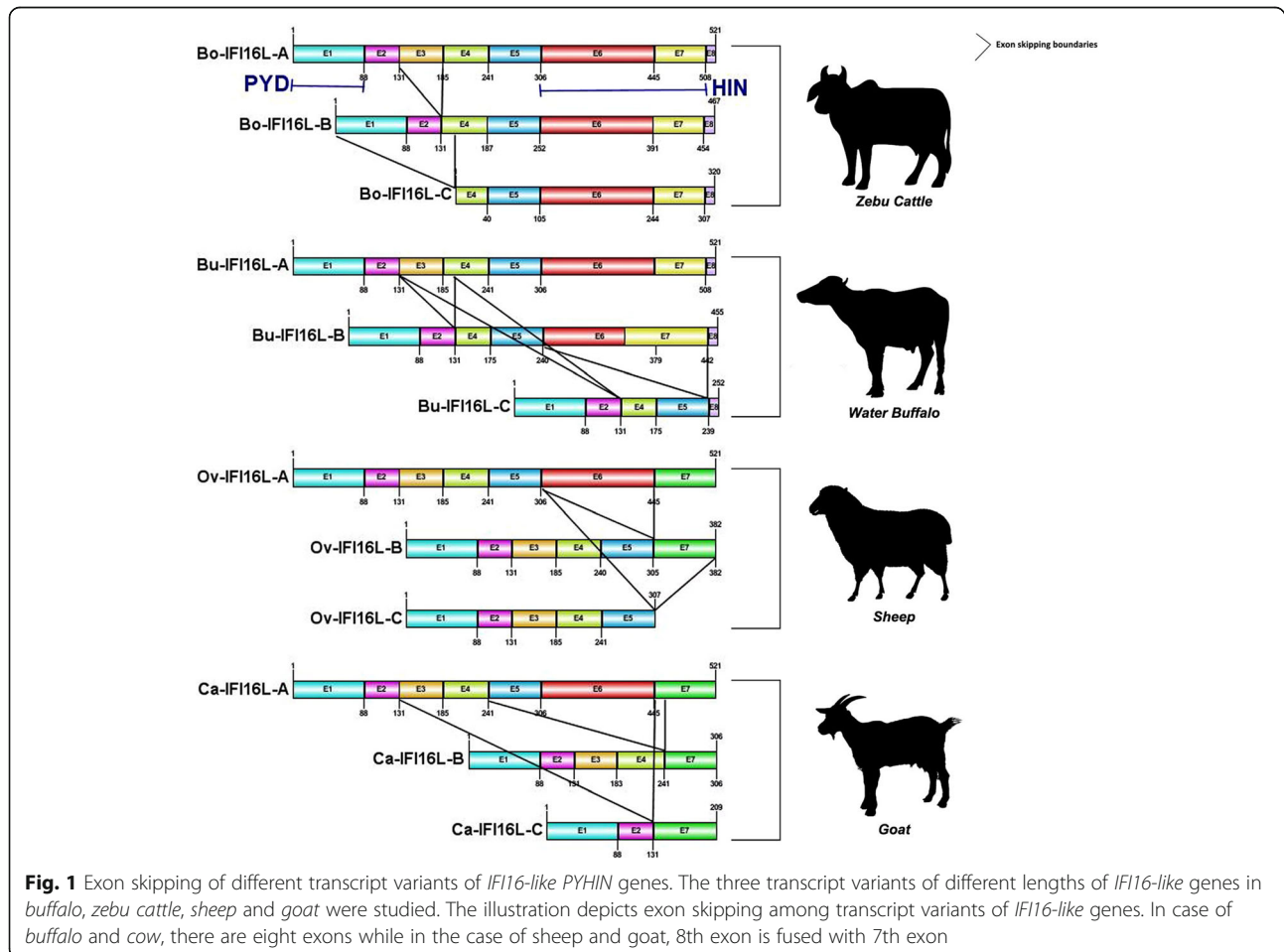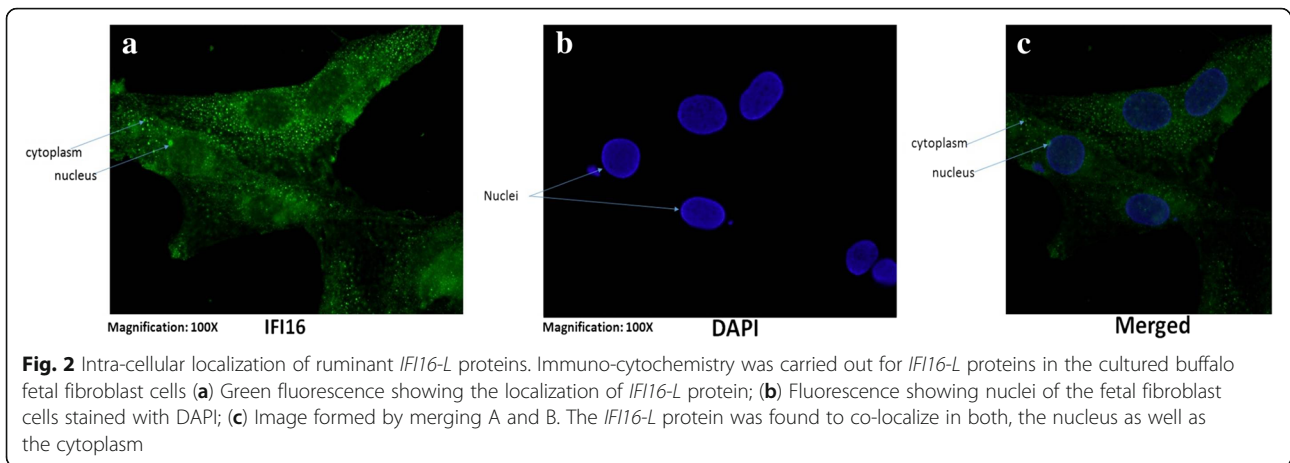

Fig. 1 Exon skipping of different transcript variants of *IFI16-like PYHIN* genes. The three transcript variants of different lengths of *IFI16-like* genes in *buffalo*, *zebu cattle*, *sheep* and *goat* were studied. The illustration depicts exon skipping among transcript variants of *IFI16-like* genes. In case of *buffalo* and *cow*, there are eight exons while in the case of sheep and goat, 8th exon is fused with 7th exon
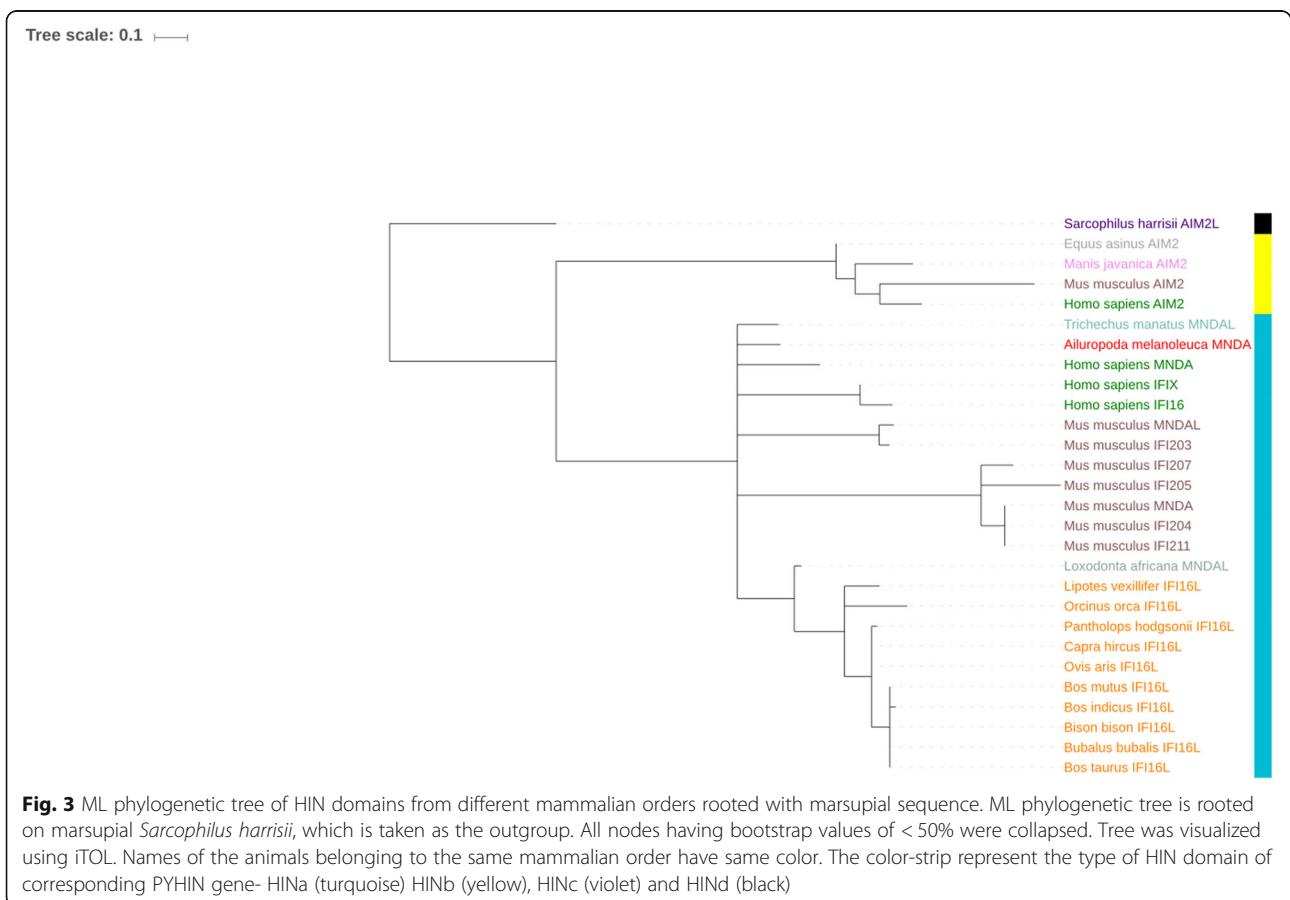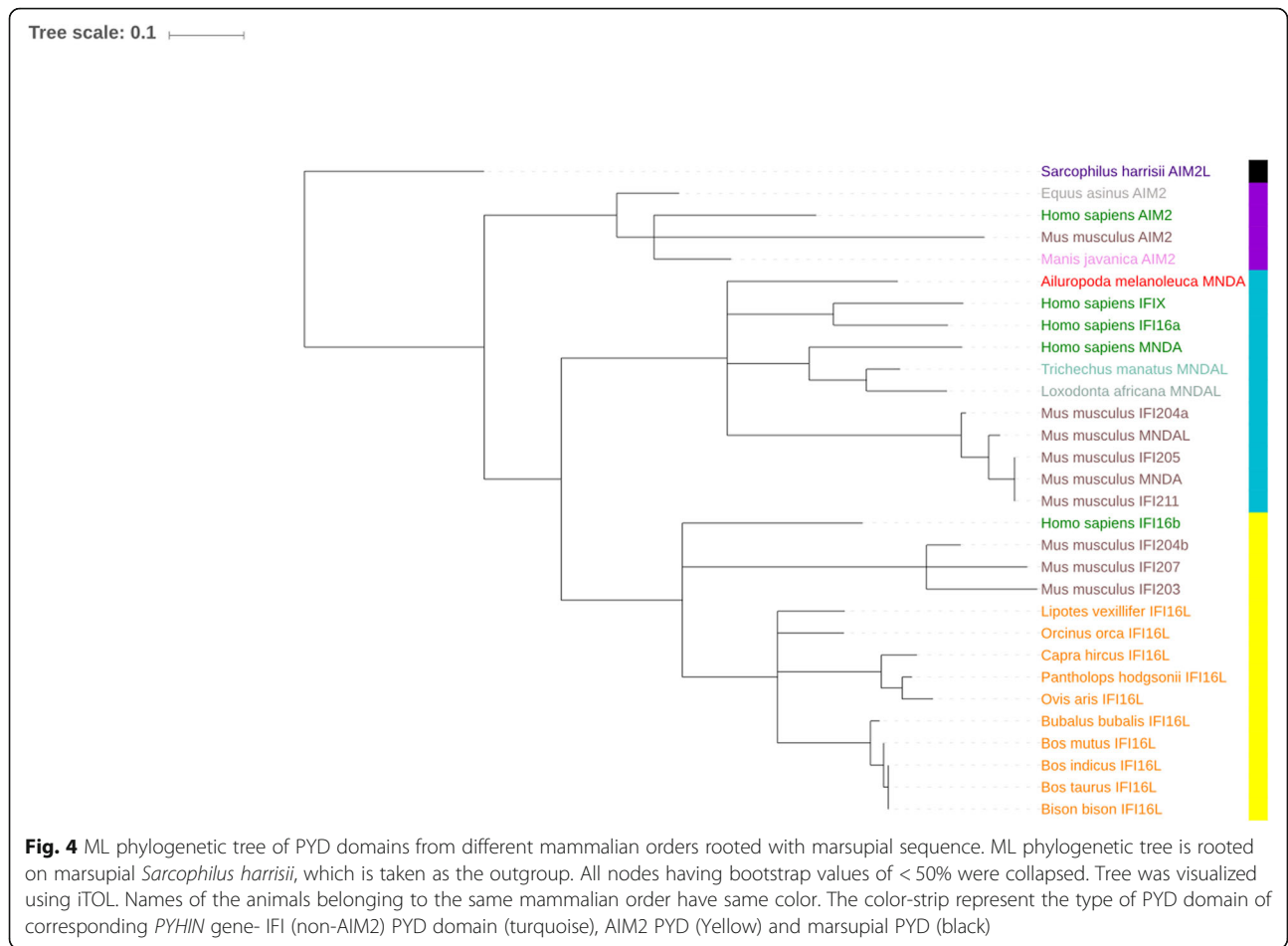
Fig. 2 Intra-cellular localization of ruminant *IFI16-L* proteins. Immuno-cytochemistry was carried out for *IFI16-L* proteins in the cultured buffalo fetal fibroblast cells (**a**) Green fluorescence showing the localization of *IFI16-L* protein; (**b**) Fluorescence showing nuclei of the fetal fibroblast cells stained with DAPI; (**c**) Image formed by merging A and B. The *IFI16-L* protein was found to co-localize in both, the nucleus as well as the cytoplasm

with the previously reported phylograms of HIN and PYD domains. The phylogenetic tree of inter-domain linker region was rooted on the whale *Orcinus orca*.

The HIN domain ML phylogram produced three clades of placental mammal sequences based on their HIN domain types i.e. HINa, HINb and HINc while the fourth clade of the marsupial sequence has HINd domain which aligns separately. The HIN domains of

IFI16-L ALRs among species members of the order Artiodactyla (which also include ruminants), align in the clade of HINb domain along with the HINb domain of human *IFI16*.

On the other hand, the PYD domain phylogram shows different tree topology compared to the HIN domain tree topology and it produced two separate clades among placental mammals while the third clade is made of the
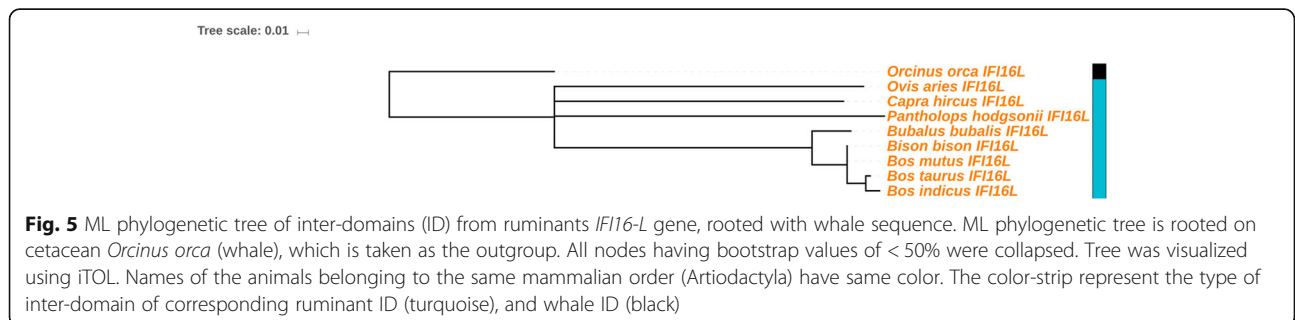

Fig. 3 ML phylogenetic tree of HIN domains from different mammalian orders rooted with marsupial sequence. ML phylogenetic tree is rooted on marsupial *Sarcophilus harrisii*, which is taken as the outgroup. All nodes having bootstrap values of < 50% were collapsed. Tree was visualized using iTOL. Names of the animals belonging to the same mammalian order have same color. The color-strip represent the type of HIN domain of corresponding PYHIN gene- HINa (turquoise) HINb (yellow), HINc (violet) and HINd (black)

Fig. 4 ML phylogenetic tree of PYD domains from different mammalian orders rooted with marsupial sequence. ML phylogenetic tree is rooted on marsupial *Sarcophilus harrisii*, which is taken as the outgroup. All nodes having bootstrap values of < 50% were collapsed. Tree was visualized using iTOL. Names of the animals belonging to the same mammalian order have same color. The color-strip represent the type of PYD domain of corresponding *PYHIN* gene- IFI (non-AIM2) PYD domain (turquoise), AIM2 PYD (Yellow) and marsupial PYD (black)

marsupial sequence. One clade of placental mammals formed by *AIM2-like PYHIN* genes while another clade is formed by the rest of the interferon-inducible *PYHIN* genes including ruminant *IFI16-L*. The phylogram of inter-domain sequences formed a single clade of ruminant animals.

Nature of selection varies along the different domain regions of ruminant *IFI16-like PYHIN* gene

To explore variations in selective pressures among the PYD domain, HIN domain, and inter-domain, various statistical models as implemented in Datamonkey web

server, were employed. The full length nucleic acid ORF sequences of selected ruminants were divided into three parts, PYD domain, HIN domain and the inter-domain. Since recombining sequences can give misleading results in phylogenetic analysis [15, 16], the DSS method [36] was used as implemented in the TOPALi program [37]. No recombination breakpoints were found using this method. Recombination breakpoints were further checked in multiple sequence alignments by GARD (Genetic Analysis for Recombination Detection) algorithm [38]. The GARD analyses revealed the presence of one putative recombination breakpoint in the HIN domain alignment,

Fig. 5 ML phylogenetic tree of inter-domains (ID) from ruminants *IFI16-L* gene, rooted with whale sequence. ML phylogenetic tree is rooted on cetacean *Orcinus orca* (whale), which is taken as the outgroup. All nodes having bootstrap values of < 50% were collapsed. Tree was visualized using iTOL. Names of the animals belonging to the same mammalian order (Artiodactyla) have same color. The color-strip represent the type of inter-domain of corresponding ruminant ID (turquoise), and whale ID (black)

which was found non-significant to produce a topological incongruence by the KH (Kishino–Hasegawa test, at $P = 0.1$) test.

The HyPhy software package at datamonkey web server [39] was subsequently used to analyze the sequences. With the user-friendly graphical interfaces, new fast methods and multi-tasking abilities, HyPhy provides better analysis power to users in comparison of PAML software package [40–46]. The web-server version of HyPhy i.e. Datamonkey provides additional advantages. With the use of the codon model selection (CMS) module, HKY85 was found as the best model for the statistical analysis of all the aligned sequences. PARRIS (a PARtitioning approach for Robust Inference of Selection) test [47] from HyPhy package (implemented in Datamonkey web server) was used to find the positive selection pressure. This test is similar to the M2a versus M1a test (PAML) but it (PARRIS) also considers synonymous substitution rate variations (dS). We found the evidences of positive selection (at $p < 0.1$ threshold) in HIN domain (p-value = 0.047) and the inter-domain linker region (p-value = 0.080) while PARRIS test inferred no evidence of positive selection in the PYD domain (p-value = 0.999) at $p < 0.1$ threshold (see Additional file 2). A second-round analysis was performed which was aimed at the detection of selection at individual codons. The nature of selection at individual codon can highlight the residues interacting with pathogenic molecules. As the aim was to find out the individual sites under selections, so models like aBSREL, which provides information about lineage specific selection were not taken into considerations. The fixed effect likelihood (FEL) method and fast unconstrained Bayesian approximation method (FUBAR) [48] were applied to all the alignments. FUBAR use a Bayesian approach and produces similar information to FEL, but it is more powerful and fast in comparison to FEL [48–50]. Both FEL and FUBAR analysis (p-value < 0.1 and posterior probability > 0.9 respectively) identified codon sites that evolved under the influence of pervasive diversifying selection and pervasive purifying selection. A total of ten codons in the inter-domain region and five codons in the HIN domain

were identified to be evolved under pervasive diversifying selection (Table 1) but no codon was found to be positively selected in the PYD domain. Codons evolved under pervasive purifying selection were six in PYD domain, seven in HIN domain and three in inter-domain linker region (Table 1). Mixed Effect Likelihood Model (MEME) [40] was also applied to all the aligned sequences in order to explore the episodic diversifying selection on the individual codons. Three codons in the HIN domain and only one codon in the inter-domain region were observed through this analysis. On the contrary, no episodic diversifying selection was observed in the PYD domain (at the $p < 0.05$ level). These results make clear demarcation about the type of selection pressure on different parts of *IFI16-L* sequences.

Discussion

Different types of transcript variants of a gene may be produced due to exon-skipping events at transcription level. There are found three transcript variants of different lengths of *IFI16-L* gene in each *ruminant* under study. These transcript variants can translate to form PYD-only proteins (POP) and HIN-only proteins (HOP). The presence of these POPs and HOPs has also been mentioned in the earlier studies [15, 16]. It appears that the main transcript is maintained for common pathogen threats while variable length transcript variants may be required in the case of exposure to new pathogens. The alternative splicing of immune genes has also been reported to occur in a tissue-specific manner [51]. These splice forms increase the diversity of particular protein and this may be a possible adaptive mechanism by which the pathogenic counteraction can be neutralized [52–54]. It is also possible that some of the transcript variants (most probably POPs and HOPs) may act as negative regulators of *IFI16-L* signaling [55].

Nuclear Localization Signal (NLS) is the positively charged, lysine or arginine-rich peptide residues at either N-terminal or C-terminal of the protein, which helps in importing protein into the nucleus through the nuclear pore complex. The presence of N-terminal NLS is a feature of nuclear *PYHIN* proteins. This feature of *PYHINs*

Table 1 Positively and negatively selected codon sites in different domains of *IFI16-like PYHIN* gene, obtained through FEL, FUBAR and MEME analysis

Domain	Total codons	FEL		FUBAR		MEME
		Codons under pervasive diversifying selection	Codons under pervasive purifying selection	Codons under pervasive diversifying selection	Codons under pervasive purifying selection	Codons under episodic diversifying selection
PYD	88 (1–88)	–	3, 15, 19, 26, 29, 63	–	3, 15, 26, 63	–
ID	218 (89–306)	234	113, 198, 301	91, 93, 105, 120, 218, 226, 233, 234, 250, 285	113, 198	234
HIN	203 (307–509)	344	320, 338, 357, 391, 420, 459	344, 398, 451, 458, 496	320, 357, 459	344, 451, 496

can help to sense DNA viruses (like *herpesviruses*) in the nucleus as their DNA remains covered by an envelope in the cytoplasm. *IFI16* sense viral DNA in the nucleus and the nuclear shuttling of *IFI16* in and out is modulated by acetylation/de-acetylation status of its (*IFI16*) NLS [56]. This indicates that *IFI16-L* gene in the ruminants might be playing a role in nuclear viral DNA sensing in localization dependent manner similar to *IFI16* in humans.

The HIN domain ML phylogram produced three clades of placental mammal sequences based on their HIN domain types i.e. HINa, HINb and HINc while the fourth clade of the marsupial sequence has HINd domain which aligns separately. The HIN-C type domain is the exclusive property of *AIM2-like* receptors [16]. The different types of HIN domains are previously reported for their different DNA binding abilities. These different types of HIN domains in mammalian lineages may have evolved for different types of DNA structures from different pathogens present in the niche of a particular lineage [10, 16].

On the other hand, the PYD domain phylogram produced two separate clades among placental mammals while the third clade is made of the marsupial sequence. Separate evolutionary history of *AIM2-like PYHIN* genes indicates a possible different role of *AIM2-like* PYD domain in the cellular physiology to the other types of PYD domains. *IFI16-L PYHIN* gene in the ruminants and cetaceans are quite similar, despite of differences in their habitats (land and marine respectively). In the phylogenetic tree of inter-domain linker region, all the sequences were aligned into single clade; which shows that inter-domains of *IFI16-L* genes are conserved in related species.

Eukaryotic cells are equipped with different types of DNA sensors to sense and respond against the different types of invading foreign DNA. In nature, hosts and pathogens have been co-evolving. Their interactions put selection pressure on each other which results in a diversifying selection. *IFI16-L PYHIN* genes have three distinct parts with totally different functions which are, the N-terminal PYD domain that homotypically binds and interacts with the PYD domain of adaptor protein *PYCARD* for downstream signaling, the C-terminal HIN domain which binds with pathogenic dsDNA and the inter-domain linker region which may play a role in localization of PYD and HIN domains for their proper functioning. Based on their different functions the nature of selection might be acting differently on these distinct parts.

The distribution of positively selected sites on HIN domain and inter-domain linker region is congruent with the previous study on primate *IFI16*. But in the case of PYD domain in ruminant *IFI16-L*, no positively selected

sites were found however, three positively selected sites were reported on the PYD domain of primate *IFI16* [19]. At the interacting sites of the host and the pathogen proteins, there are more non-synonymous substitutions (dN) which drives the positive selection on these interacting residues. To evade the host sensing mechanism, the invading viral proteins target the ALRs as a counteract measure. The *pUL83* protein which is encoded by herpesvirus HCMV binds to the PYD domain of *IFI16* and blocks its oligomerization hence, inhibiting the downstream signaling process [57]. This interaction of viral *pUL83* protein might be the driving force for positive selection in the PYD domain of primate *IFI16*.

However, similar selection pressure in PYD domain appears absent in ruminant species. *HSV-1* encoded protein immediate early protein *ICP0*, in a way to evade host defense recognition mechanism, targets and binds to *IFI16*. It leads to its (*IFI16*) proteasome mediated degradation process which ultimately results in the loss of *IFI16* protein [58]. Evidence of positive selection pressure might help to evolve new host residues to prevent viral antagonists and subsequent host target modifications. The HIN domain is primarily involved in the binding of foreign DNA directly thus, it should have been under the influence of selective pressure, but there are also positively selected residues were found in the inter-domain region of *IFI16* and *IFI16-L PYHIN* genes of primates and ruminants respectively.

As we know that viral DNA sensing by *IFI16* is localization dependent, so the nuclear localization signal (NLS) might be the prime target of viral factors to interfere in the translocation of *IFI16-L*. Some part of NLS lies in the PYD domain and the rest resides in the inter-domain linker. The part of NLS that resides in the inter-domain has two positively selected sites which show that this NLS part is engaged in the conflicts with viral defense proteins. In this regard, further studies are required to fully understand the role of *IFI16-like PYHIN* genes in context to their foreign DNA binding capabilities. Pathogens are one of the strongest selective forces. The host PRRs is evolved to recognize special featured structures of pathogen called PAMPs. The purifying selection on the PYD domain may be compared to that of house-keeping genes that are essential for proper functioning of vital molecular machinery. Any mutation in these genes will be deleterious and will be eliminated. The purifying selection has important role in long term stability of biological structures by removing deleterious mutations [25, 59–61].

Conclusion
ALRs of *PYHIN* family genes are primitive class of innate immune PRRs of pathogenic double stranded (dsDNA) found in (but not limited to) mammals. A single

Kumar *et al. BMC Evolutionary Biology*   (2019) 19:26

Page 8 of 11

functional *PYHIN* gene i.e. *IFI16-like (IFI16-L)* gene is reported in the clade Ruminantia. *IFI16-L* has evolutionarily conserved orthology in Ruminantia and Cetacea which are quite similar clades. Because of its activity as an immune sensor just like human *IFI16*, it might be engaged in direct conflicts with pathogenic proteins. This phenomenon could put pressure on *IFI16-L* gene to evolve. *IFI16-L* protein has three distinct parts- the PYD domain, the HIN domain and the inter-domain linker region; to perform separate activities. Due to this, the nature of selection pressure also varies on these different parts.

In the current study, the selective pressure that acts on *IFI16-like PYHIN* genes in ruminants was investigated. The inter-domain region is influenced by maximum diversifying selection pressure which indicates its involvement in the interaction with pathogenic defense factors. The DNA binding HIN domain is also under the influence of diversifying selection pressure. This *IFI16-L* ALR encodes three transcript variants with complete ORFs. This may be an adaptive strategy to lame the viral defense. Also, some of these transcript variants may play a role in the regulation of its *(IFI16-L)* activities. Furthermore, research is needed to shed more light on the role of the *IFI16-like PYHIN* genes, particularly its transcript variants, in immune DNA sensing and the interactions of inter-domain linker region with various pathogenic proteins.

Materials and methods

Sample collection

We assembled blood [into heparinized vacuutainers (BD)] of four ruminant species from National Dairy Research Institute (NDRI) animal herd and local slaughterhouse, as follows: *water buffalo (Bubalus bubalis), zebu cattle (Bos indicus), goat (Capra hircus), sheep (Ovis aries).* Three sets of blood samples were collected as biological replicates, immediately PBMCs were separated; mixed with Trizol Reagent (Invitrogen, USA) and frozen into − 80 °C for further use.

PCR, cloning and sequencing

Total RNA was extracted from 0.5 mL of collected blood samples (TRIzol reagent Method). The quality and quantity of RNA samples were determined using a NanoDrop 2000 spectrophotometer (Thermo Scientific, USA) and were further ascertained by agarose gel electrophoresis. One μg of each RNA sample was reverse transcribed using the RevertAid First Strand cDNA synthesis kit (Thermo Scientific, USA), by using random hexamer primers according to the manufacturer's instructions. Oligonucleotide primers (Table 2) for full coding regions of *IFI16-L* genes were designed from the conserved regions of the predicted *IFI16-L* gene sequences of

ruminants   (XM_006060336,   XM_863928, XM_012172466,   XM_005701415)   from   GenBank. Around 100 ng of cDNA was used as the template for each amplification reaction containing 0.5 μL Advantage Taq Polymerase (Clontech, USA), 10 pmol of each *IFI16-L* primers in a total of 25 μL reaction mix. The amplified PCR products were cloned using InsTAclone PCR cloning kit (Thermo Scientific, USA) and transformed into XL1-Blue competent *E.coli* cells (Stratagene). The recombinant clones were subsequently screened through PCR. Sanger's sequencing was performed with DTCS quick start cycle sequencing kit (Beckman Coulter) and run on GenomeLab™ GeXP Genetic Analysis System (Beckman Coulter).

Immunofluorescence assay

Buffalo fetal fibroblasts were cultured for 24 h at 37 °C with 5% CO_2 and the next day the cells were fixed with 4% paraformaldehyde for 30 min at 4 °C. Cells were incubated with the Anti-*IFI16* polyclonal antibody (SIGMA) at 37 °C for 1 h and afterward incubated with FITC-conjugated Anti-rabbit secondary antibodies under the same conditions. Dilution of both primary and secondary antibodies used was 1:50. The cells were further co-stained with 1 μg/mL of DAPI and examined under fluorescent microscope.

Bioinformatics analysis

Homologs of *IFI16-L* gene were searched with NCBI BLASTn tool and UCSC-BLAT. Additional mammalian *PYHIN* nucleotide sequences were retrieved from NCBI Database for phylogenetic analysis (see Additional file 3). Amino acid sequences of these *ALRs* were obtained by translating the respective nucleotide sequences using the ExPASy Translate tool. Molecular weights and isoelectric points were calculated using the Compute pI/Mw tool [62]. Conserved domain architectures of *IFI16-L* protein sequences were determined with InterPro [63] and Conserved Domain Search Service [64]. We predicted a nuclear localization signal (NLS) in the N-terminal part of *IFI16-L ALRs* using the NLStradamus program [65], which is based on Hidden Markov models (HMMs) statistical approaches.

Phylogenetic analysis

Nucleotide sequences of PYD domain, HIN-200 domain and inter-domain (Additional file 3) were aligned separately using the RevTrans 2.0 web server [34], which uses a scaffold of protein sequence alignment to construct the DNA multiple alignments of the coding DNA in order to reduce the signal to noise ratio. Divergent sequences which created alignments errors; were identified and removed manually. Phylogenetic analyses of PYD and HIN domains were conducted using "one-click"

Table 2 Primers used for gene amplification and sequencing

Gene	Primer name	Sequence (5′ – 3′)	T_m (°C)	Length (bp)
IFI16-L	IFI16FW (Full ORF)	AGCCAGCACTAAATCAAGGGTTTTGACTAGTC	62	32
	IFI16RV (Full ORF)	GCCAAGTTATAAAAGTTATAGATGCAAAATGGGACA	60	36
	IFI16FW2	GTGACACAGCCACCAAACCTAAGGAC	61	26
	IFI16FW3	GAGGCACATCCCTAATTTGAGAGATGGGG	63	29
	IFI16RV1	ATCTGGCCACTGCTCACAGTTGGG	61	24
M13	M13/pUC	GTAAAACGACGGCCAGT	47	17
	M13/pUCRV	CAGGAAACAGCTATGAC	45	17

mode at *Phylogeny.fr* suite of programs [35, 66]. Translated amino acid sequences were given as input which was aligned using MUSCLE 3.8.31 and used Gblocks 0.91b filter to check ambiguous (poorly aligned and divergent) sequence regions. No ambiguous sequence was found in any of the alignments under study. A maximum likelihood tree was created using PhyML 3.0 with 100 bootstrap replicates and the tree was rendered using TreeDyn 198.3 software. Trees of the functional domains i.e. PYD and HIN, were re-rooted on *Sarcophilus harrisii* (marsupial), which is taken as an outgroup; while in the case of inter-domain phylogenetics tree, *Orcinus orca* (cetacean), is taken as an outgroup. All the nodes having bootstrap values of < 50% were collapsed. Trees were exported in the Newick format and visualized using iTOL online software [67].

To determine the selective force on *IFI16-L* gene in ruminants, nucleotide sequences of eight ruminants (*water buffalo, zebu cattle, goat, sheep, Taurus cattle, domestic yak, bison and Tibetan antelope*) were selected. The PYD domain, HIN domain and the inter-domain linker region nucleotide sequences were aligned separately with the RevTrans 2.0 utility and the codon sequences were subsequently used for selection pressure analysis. Amino acid sequence alignments of PYD domain, HIN domain and inter-domain region are provided in Additional file 4.

Selective pressure analysis
Prior to selective pressure analysis, we checked all of the groups of nucleotide sequences for intragenic recombination. Recombination in sequences was tested with a difference of the sum of squares (DSS) method using a step size of 10 bp and a sliding window of 80 bp [36] as implemented in the software package TOPALi v2.5 [37]. Intragenic recombination was also assessed through GARD (Genetic Analysis for Recombination Detection) analyses [38, 39]. The significant topological incongruence of recombination breakpoint was tested using the Kishino–Hasegawa test (KH test) [68].

We used statistical tools from the HyPhy package [69] implemented in the Datamonkey web server [38, 70] to characterize the patterns of selection on the alignments. The Neighbour-Joining (NJ) tree, with HKY85 [71] nucleotide substitution model, was selected with the HyPhy model selection tool for each alignment. For the analysis with HyPhy softwares, the default set parameters of *p*-values/posterior probabilities were taken into account for each analysis. Maximum likelihood tests were performed utilizing the partitioning approach for robust inference of selection (PARRIS) method [47], which is an extension of the standard likelihood ratio test and can also better deal with recombining sequences. FEL (Fixed Effect Likelihood) and FUBAR (Fast Unbiased Bayesian AppRoximation) [42, 48] were utilized to identify sites evolving under pervasive diversifying and pervasive purifying selection pressures. MEME (Mixed Effects Model of Evolution) [40] was used to identify episodically diversifying sites. FEL is an ML based model which assumes constant selection pressure on each site. The FUBAR model estimates site-by-site ω values while MEME allows the distribution of ω to vary over sites and additionally from branch to branch. Significance was assessed by posterior probability > 0.9 (FUBAR) and *p*-value < 0.05 (MEME), p-value < 0.1 (FEL) and *p*-value < 0.1 (PARRIS).

Additional files

Additional file 1: Diversity of *PYHIN* genes. Orthology in *PYHIN* genes is not conserved. The number and type of *PYHIN* genes varies in different mammalian orders. Here, the *PYHIN* genes and their associated domains from selected orders of classes Mammalia, Reptilia and Sarcopterygii (fish) have been presented. The residue lengths of PYD and HIN domains of all the animals are conserved. Drastic variation in the length of the inter-domain linker regions (shown in grey colour) can be observed among the various animals as well as among the different *PYHIN* genes within the same animal. The small HIN region within the *IFI16-L* from the elephant represents a rudimentary form of HINa domain. (JPG 345 kb)

Additional file 2: Tables containing PARRIS results of (a) PYD domain, (b) inter-domain and (c) HIN domain. (DOCX 17 kb)

Additional file 3: *PYHIN* sequences used for phylogenetic analysis. (DOCX 13 kb)

Additional file 4: Multiple sequence alignments of (a) PYD domain, (b) inter-domain and (c) HIN domain. (DOCX 101 kb)

Publisher's Note
Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

References
1. Janeway CA Jr. Pillars article: approaching the asymptote? Evolution and revolution in immunology. *Cold spring harb symp quant biol.* 1989. 54: 1-13. J Immunol. 2013;191(9):4475–87.
2. Bianchi ME. DAMPs, PAMPs and alarmins: all we need to know about danger. J Leukoc Biol. 2007;81(1):1–5.
3. Lamkanfi M, Dixit VM. Mechanisms and functions of inflammasomes. Cell. 2014;157(5):1013–22.
4. Orzalli MH, Knipe DM. Cellular sensing of viral DNA and viral evasion mechanisms. Annu Rev Microbiol. 2014;68:477–92.
5. Connolly DJ, Bowie AG. The emerging role of human PYHIN proteins in innate immunity: implications for health and disease. Biochem Pharmacol. 2014;92(3):405–14.
6. Kanneganti TD. The inflammasome: firing up innate immunity. Immunol Rev. 2015;265(1):1–5.
7. Tavares AH, Bürgel PH, Bocca AL. Turning up the heat: inflammasome activation by fungal pathogens. PLoS Pathog. 2015;11(7):e1004948.
8. Guo H, Callaway JB, Ting JP. Inflammasomes: mechanism of action, role in disease, and therapeutics. Nat Med. 2015;21(7):677–87.
9. Unterholzner L. The interferon response to intracellular DNA: why so many receptors? Immunobiology Immunobiology. 2013;218(11):1312–21.
10. Unterholzner L, Keating SE, Baran M, Horan KA, Jensen SB, Sharma S, Sirois CM, Jin T, Latz E, Xiao TS, Fitzgerald KA, Paludan SR, Bowie AG. IFI16 is an innate immune sensor for intracellular DNA. Nat Immunol. 2010;11(11):997–1004.
11. Fernandes-Alnemri T, Yu JW, Datta P, Wu J, Alnemri ES. AIM2 activates the inflammasome and cell death in response to cytoplasmic DNA. Nature. 2009;458(7237):509–13.
12. Gariglio M, Mondini M, De Andrea M, Landolfo S. The multifaceted interferon-inducible p200 family proteins: from cell biology to human pathology. J Interf Cytokine Res. 2011;31(1):159–72.
13. Shaw N, Liu ZJ. Role of the HIN domain in regulation of innate immune responses. Mol Cell Biol. 2014;34(1):2–15.
14. Kerur N, Veettil MV, Sharma-Walia N, Bottero V, Sadagopan S, Otageri P, Chandran B. IFI16 acts as a nuclear pathogen sensor to induce the inflammasome in response to Kaposi sarcoma-associated herpesvirus infection. Cell Host Microbe. 2011;9(5):363–75.
15. Brunette RL, Young JM, Whitley DG, Brodsky IE, Malik HS, Stetson DB. Extensive evolutionary and functional diversity among mammalian AIM2-like receptors. J Exp Med. 2012;209(11):1969–83.
16. Cridland JA, Curley EZ, Wykes MN, Schroder K, Sweet MJ, Roberts TL, Ragan MA, Kassahn KS, Stacey KJ. The mammalian PYHIN gene family: phylogeny, evolution and expression. BMC Evol Biol. 2012;12:140.
17. Ahn M, Cui J, Irving AT, Wang LF. Unique loss of the PYHIN gene family in bats amongst mammals: implications for inflammasome sensing. Sci Rep. 2016;6:21722.
18. Schattgen SA, Fitzgerald KA. The PYHIN protein family as mediators of host defenses. Immunol Rev. 2011;243(1):109–18.
19. Cagliani R, Forni D, Biasin M, Comabella M, Guerini FR, Riva S, Pozzoli U, Agliardi C, Caputo D, Malhotra S, Montalban X, Bresolin N, Clerici M, Sironi M. Ancient and recent selective pressures shaped genetic diversity at AIM2-like nucleic acid sensors. Genome Biol Evol. 2014;6(4):830–45.
20. Levin TC, Malik HS. Rapidly evolving *Toll-3/4* genes encode male-specific toll-like receptors in *Drosophila*. Mol Biol Evol. 2017;34(9):2307–23.
21. Jakobsen MR, Bak RO, Andersen A, Berg RK, Jensen SB, Tengchuan J, Laustsen A, Hansen K, Ostergaard L, Fitzgerald KA, Xiao TS, Mikkelsen JG, Mogensen TH, Paludan SR. IFI16 senses DNA forms of the lentiviral replication cycle and controls HIV-1 replication. Proc Natl Acad Sci U S A. 2013;110(48):E4571–80.
22. Dell'Oste V, Gatti D, Giorgio AG, Gariglio M, Landolfo S, De Andrea M. The interferon-inducible DNA-sensor protein IFI16: a key player in the antiviral response. NewMicrobiol 2015;38(1):5–20.
23. Rypuła K, Krasińska M, Kita J. The occurrence of alpha-herpesvirus and pestivirus infections in European bison (Bison bonasus) in the Bialowieza primeval Forest. European Bison Conservation Newsletter. 2011;4:89–94.
24. Ravishankar C, Nandi S, Chander V, Mohapatra TK. Concurrent testing of breeding bulls for bovineherpes virus 1 infection (BHV-1) in India. Vet Ital. 2013;49(2):145–50.
25. Orzalli MH, DeLuca NA, Knipe DM. Nuclear IFI16 induction of IRF-3 signaling during herpesviral infection and degradation of IFI16 by the viral ICP0 protein. Proc Natl Acad Sci U S A. 2012;109(44):E3008–17.
26. Lupfer C, Malik A, Kanneganti TD. Inflammasome control of viral infection. Curr Opin Virol. 2015;12:38–46.
27. Roy A, Dutta D, Iqbal J, Pisano G, Gjyshi O, Ansari MA, Kumar B, Chandran B. Nuclear innate immune DNA sensor IFI16 is degraded during lytic reactivation of Kaposi's sarcoma-associated herpesvirus (KSHV): role of IFI16 in maintenance of KSHV latency. J Virol. 2016;90(19):8822–41.
28. Gariano GR, Dell'Oste V, Bronzini M, Gatti D, Luganini A, De Andrea M, Gribaudo G, Gariglio M, Landolfo S. The intracellular DNA sensor IFI16 gene acts as restriction factor for human cytomegalovirus replication. PLoS Pathog. 2012;8:e1002498.
29. Johnson KE, Chikoti L, Chandran B. Herpes simplex virus 1 infection induces activation and subsequent inhibition of the IFI16 and NLRP3 inflammasomes. J Virol. 2013;87:5005–18.
30. Orzalli MH, Conwell SE, Berrios C, DeCaprio JA, Knipe DM. Nuclear interferon-inducible protein 16 promotes silencing of herpesviral and transfected DNA. Proc Natl Acad Sci U S A. 2013;110:E4492–501.

31. Pisano G, Roy A, Ahmed Ansari M, Kumar B, Chikoti L, Chandran B. Interferon-γ-inducible protein 16 (IFI16) is required for the maintenance of Epstein-Barr virus latency. Virol J. 2017;14(1):221.
32. Mitchell PS, Patzina C, Emerman M, Haller O, Malik HS, Kochs G. Evolution-guided identification of antiviral specificity determinants in the broadly acting interferon-induced innate immunity factor MxA. Cell Host Microbe. 2012;12(4):598–604.
33. Fisher CA, Bhattarai EK, Osterstock JB, Dowd SE, Seabury PM, Vikram M, Whitlock RH, Schukken YH, Schnabel RD, Taylor JF, Womack JE, Seabury CM. Evolution of the bovine TLR gene family and member associations with Mycobacterium avium subspecies paratuberculosis infection. PLoS One. 2011;6(11):e27744.
34. Wernersson R, Pedersen AG. RevTrans - constructing alignments of coding DNA from aligned amino acid sequences. Nucl Acids Res. 2003;31(13):3537–9.
35. Dereeper A, Guignon V, Blanc G, Audic S, Buffet S, Chevenet F, Dufayard JF, Guindon S, Lefort V, Lescot M, Claverie JM, Gascuel O. Phylogeny.fr: robust phylogenetic analysis for the non-specialist. Nucl Acids Res. 2008;36(Web Server):W465–9.
36. McGuire G, Wright F. TOPAL 2.0: improved detection of mosaic sequences within multiple alignments. Bioinformatics. 2000;16(2):130–4.
37. Milne I, Lindner D, Bayer M, Husmeier D, McGuire G, Marshall DF, Wright F. TOPALi v2: a rich graphical interface for evolutionary analyses of multiple alignments on HPC clusters and multi-core desktops. Bioinformatics. 2009;25(1):126–7.
38. Kosakovsky Pond SL, Posada D, Gravenor MB, Woelk CH, Frost SD. Automated phylogenetic detection of recombination using a genetic algorithm. Mol Biol Evol. 2006;23(10):1891–901.
39. Datamonkey. https://www.datamonkey.org. Accessed December 2017.
40. Murrell B, Wertheim JO, Moola S, Weighill T, Scheffler K, Kosakovsky Pond SL. Detecting individual sites subject to episodic diversifying selection. PLoS Genet. 2012;8(7):e1002764.
41. Pond SL, Frost SDW. Datamonkey: rapid detection of selective pressure on individual sites of codon alignments. Bioinformatics. 2005a;21(10):2531–5 https://doi.org/10.1093/bioinformatics/bti320.
42. Pond SL, Frost SDW. Not so different after all: a comparison of methods for detecting amino acid sites under selection. Mol Biol Evol. 2005b;22(5):1208–22 https://doi.org/10.1093/molbev/msi105.
43. Bulmer MS, Crozier RH. Variation in positive selection in termite GNBPs and relish. Mol Biol Evol. 2006;23(2):317–26.
44. Cavatorta JR, Savage AE, Yeam I, Gray SM, Jahn MM. Positive Darwinian selection at single amino acid sites conferring plant virus resistance. J Mol Evol. 2008;67(5):551–9.
45. Parmakelis A, Moustaka M, Poulakakis N, Louis C, Slotman MA, Marshall JC, Awono-Ambene PH, Antonio-Nkondjio C, Simard F, Caccone A, Powell JR. Anopheles immune genes and amino acid sites evolving under the effect of positive selection. PLoS One. 2010;5(1):e8885.
46. Wang Y, et al. Mitogenomic perspectives on the origin of Tibetan loaches and their adaptation to high altitude. Sci Rep. 2016;6:29690.
47. Scheffler K, Martin DP, Seoighe C. Robust inference of positive selection from recombining coding sequences. Bioinformatics. 2006;22(20):2493–9.
48. Murrell B, Moola S, Mabona A, Weighill T, Sheward D, Kosakovsky Pond SL, Scheffler K. FUBAR: a fast, unconstrained bayesian approximation for inferring selection. Mol Biol Evol. 2013;30(5):1196–205.
49. Felsenstein J. Taking variation of evolutionary rates between sites into account in inferring phylogenies. J Mol Evol. 2001;53:447–55.
50. Maldonado E, Sunagar K, Almeida D, Vasconcelos V, Antunes A. IMPACT_S: integrated multiprogram platform to analyze and combine tests of selection. PLoS One. 2014;9(10):e96243 https://doi.org/10.1371/journal.pone.0096243.
51. Johnstone RW, Kershaw MH, Trapani JA. Isotypic variants of the interferon-inducible transcriptional repressor IFI16 arise through differential mRNA splicing. Biochemistry. 1998;37(34):11924–31.
52. Vudattu NK, Magalhaes I, Hoehn H, Pan D, Maeurer MJ. Expression analysis and functional activity of interleukin-7 splice variants. Genes Immun. 2009;10(2):132–40.
53. Ergun A, Doran G, Costello JC, Paik HH, Collins JJ, Mathis D, Benoist C, ImmGen Consortium. Differential splicing across immune system lineages. Proc Natl Acad Sci U S A 2013;110(35):14324–14329.
54. Hancks DC, Hartley MK, Hagan C, Clark NL, Elde NC. Overlapping patterns of rapid evolution in the nucleic acid sensors cGAS and OAS1 suggest a common mechanism of pathogen antagonism and escape. PLoS Genet. 2015;11(5):e1005203.
55. Yabas M, Elliott H, Hoyne GF. The role of alternative splicing in the control of immune homeostasis and cellular differentiation. Int J Mol Sci 2015;17(1). pii: E3.
56. Gack MU, Kirchhofer A, Shin YC, Inn KS, Liang C, Cui S, Myong S, Ha T, Hopfner KP, Jung JU. Roles of RIG-I N-terminal tandem CARD and splice variant in TRIM25-mediated antiviral signal transduction. Proc Natl Acad Sci U S A. 2008;105(43):16743–8.
57. Li T, Diner BA, Chen J, Cristea IM. Acetylation modulates cellular distribution and DNA sensing ability of interferon-inducible protein IFI16. Proc Natl Acad Sci U S A. 2012;109(26):10558–63.
58. Ansari MA, Dutta S, Veettil MV, Dutta D, Iqbal J, Kumar B, Roy A, Chikoti L, Singh VV, Chandran B. Herpesvirus genome recognition induced acetylation of nuclear IFI16 is essential for its cytoplasmic translocation, inflammasome and IFN-β responses. PLoS Pathog. 2015;11(7):e1005019.
59. Charlesworth B, Lande R, Slatkin M. A neo-Darwinian commentary on macroevolution. Evolution. 1982;36(3):474–98.
60. Loewe L. Negative selection. Nature education. 2008;1(1):59.
61. Carlsson EK, Kwiatkowski DP, Sabeti PC. Natural selection and infectious disease in human populations. Nat Rev Genet. 2014;15(6):379–93.
62. Compute pl/Mw. http://web.expasy.org/compute_pi. Accessed December 2017.
63. InterPro: protein sequence analysis & classification. https://www.ebi.ac.uk/interpro. Accessed December 2017.
64. Conserved Domains. https://www.ncbi.nlm.nih.gov/Structure/cdd/wrpsb.cgi. Accessed December 2017.
65. Nguyen Ba AN, Pogoutse A, Provart N, Moses AM. NLStradamus: a simple hidden Markov model for nuclear localization signal prediction. BMC Bioinformatics. 2009;10(1):202.
66. Dereeper A, Audic S, Claverie JM, Blanc G. BLAST-EXPLORER helps you building datasets for phylogenetic analysis. BMC Evol Biol. 2010;10:8.
67. Letunic I, Bork P. Interactive tree of life (iTOL) v3: an online tool for the display and annotation of phylogenetic and other trees. Nucl Acids Res. 2016;44(W1):W242–5.
68. Kishino H, Hasegawa M. Evaluation of the maximum likelihood estimate of the evolutionary tree topologies from DNA sequence data, and the branching order in Hominoidea. J Mol Evol. 1989;29(2):170–9.
69. Kosakovsky Pond SL, Frost SD, Muse SV. HyPhy: hypothesis testing using phylogenies. Bioinformatics. 2005;21(5):676–9.
70. Delport W, Poon AF, Frost SD, Kosakovsky Pond SL. Datamonkey 2010: a suite of phylogenetic analysis tools for evolutionary biology. Bioinformatics. 2010;26(19):2455–7.
71. Hasegawa M, Kishino H, Yano T. Dating the human-ape splitting by a molecular clock of mitochondrial DNA. J Mol Evol. 1985;22(2):160–74.