

RESEARCH

Open Access



# Sequence variation of necdin gene in Bovidae

Sunday O. Peters<sup>1\*†</sup>, Marcos De Donato<sup>2\*†</sup>, Tanveer Hussain<sup>3</sup>, Hectorina Rodulfo<sup>2</sup>, Masroor E. Babar<sup>3</sup> and Ikhide G. Imumorin<sup>4,5</sup>

## Abstract

**Background:** Necdin (NDN), a member of the melanoma antigen family showing imprinted pattern of expression, has been implicated as causing Prader-Willi symptoms, and known to participate in cellular growth, cellular migration and differentiation. The region where NDN is located has been associated to QTLs affecting reproduction and early growth in cattle, but location and functional analysis of the molecular mechanisms have not been established.

**Methods:** Here we report the sequence variation of the entire coding sequence from 72 samples of cattle, yak, buffalo, goat and sheep, and discuss its variation in Bovidae. Median-joining network analysis was used to analyze the variation found in the species. Synonymous and non-synonymous substitution rates were determined for the analysis of all the polymorphic sites. Phylogenetic analysis were carried out among the species of Bovidae to reconstruct their relationships.

**Results:** From the phylogenetic analysis with the consensus sequences of the studied Bovidae species, we found that only 11 of the 26 nucleotide changes that differentiate them produced amino acid changes. All the SNPs found in the cattle breeds were novel and showed similar percentages of nucleotides with non-synonymous substitutions at the N-terminal, MHD and C-terminal (12.3, 12.8 and 12.5%, respectively), and were much higher than the percentage of synonymous substitutions (2.5, 2.6 and 4.9%, respectively). Three mutations in cattle and one in sheep, detected in heterozygous individuals were predicted to be deleterious. Additionally, the analysis of the biochemical characteristics in the most common form of the proteins in each species show very little difference in molecular weight, pI, net charge, instability index, aliphatic index and GRAVY (Table 4) in the Bovidae species, except for sheep, which had a higher molecular weight, instability index and GRAVY.

**Conclusions:** There is sufficient variation in this gene within and among the studied species, and because NDN carry key functions in the organism, it can have effects in economically important traits in the production of these species. NDN sequence is phylogenetically informative in this group, thus we propose this gene as a phylogenetic marker to study the evolution and conservation in Bovidae.

**Keywords:** MAGE, Bovini, Caprini, Phylogenetics

\* Correspondence: [speters@berry.edu](mailto:speters@berry.edu); [mddonate@itesm.mx](mailto:mddonate@itesm.mx)

†Sunday O. Peters and Marcos De Donato contributed equally to this work.

<sup>1</sup>Department of Animal Science, Berry College, Mount Berry, GA 30149, USA

<sup>2</sup>Tecnologico de Monterrey, Escuela de Ingeniería y Ciencias, Ave. Epigmenio Gonzalez 500, 76130 Queretaro, CP, Mexico

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



## Background

Cetartiodactyla is one of the most diversified mammalian orders, with 330 living species group into 128 genera [1]. This group contains whales, dolphins, hippopotamus, antelopes, deer, cattle, buffalo, sheep, goat, giraffes, camels, pigs, among many others [2]. Both morphological and molecular studies have supported the monophyletic evolution of five groups within Cetartiodactyla: (1) Cetacea, (2), Hippopotamidae, (3) Tylopoda, (4) Suina (containing Suidae and Tayassuidae) and (5) Ruminantia, which includes the infraorders Tragulina and Pecora, which in turn includes the 5 families: Giraffidae, Cervidae, Moschidae, Antilocaprinae and Bovidae [3–6]. Of these, members of the Bovidae family are the most diversified, with 142 species, including cattle, buffalo, sheep and goat [7].

Necdin (NDN), a member of the melanoma antigen (MAGE) family which is comprised of over 60 genes that share the highly-conserved MAGE homology domain (MHD) [8], is one of five genes located in the deletion region of the Prader-Willi syndrome in human, a rare genetic neurodevelopmental disease characterized by a variety of physical, cognitive, and behavioral defects [9], including short stature, early-childhood onset obesity, developmental delay/mild intellectual disability; growth hormone insufficiency, low muscle tone, increased food intake, low levels of insulin and insulin-like growth factor 1 (IGF1), incomplete sexual development, hypogonadism, and male infertility [10]. Gene inactivation studies in mouse suggest that NDN is responsible for the specific Prader-Willi symptoms [11]. It is the best characterized MAGE gene implicated as a negative growth regulator, and proposed to participate in a broad range of biological activities including cell growth, migration, differentiation and cell death/survival, but the precise molecular function is largely unknown [12].

NDN, as well as MAGEL2, shows preferential paternal expression (imprinting) being highly expressed in whole brain, hypothalamus, amygdala and pituitary gland, moderately expressed in adipocytes, uterus, lung, adrenal gland, placenta and smooth muscle [13]. NDN has been implicated as a transcription factor which acts activating the gonadotropin releasing hormone (GNRH1) in immature migratory cells, binding to the MSX1 transcription factor, which is a repressor of GNRH1 [14]. It has been shown as an imprinted tumor suppressor gene which affects cancer cell motility, invasion and growth in ovarian, prostate, urothelial, neck and head cancers [15]. Furthermore, NDN has shown to interact with E2F4 and E2F1 in the control of gene transcription [16] as well as to inhibit PPARy1 gene expression providing a potential molecular mechanism through which it regulates adipogenesis [17].

The imprinting pattern of expression, a type of monoallelic pattern of expression, has been shown to increase

evolvability, by facilitating positive Darwinian selection on heterozygous individuals, and by simultaneously allowing a relaxation of purifying selection (also known as negative selection) on heterozygous individuals [18]. In addition, monoallelic expression has shown to be an important evolutionary mechanism for the maintenance of genetic diversity and gene diversification [19]. Thus, NDN could be potentially informative for phylogenetic analysis to study groups that have shown to be difficult to resolve.

In addition, in the proximal region of bovine chromosome 21, where NDN is located, several QTLs have been associated, including calving ease (the percentage of unassisted births in first-calf heifers) [20–22], gestation length [21, 23], scrotal circumference (used as a major selection criterion to improve precocity and fertility) [24] and birth weight [22], among others.

In this study, we sequenced the coding region of the NDN gene in members of the Bovidae family, as potentially associated to traits with economic impact in the domestic species used for food production, clothing materials (hear, wool, leather) and to carry out heavy labor, due to its role in growth and neuronal development, and SNPs found here can be used for association studies to economically important traits in cattle, yak, sheep, goat and buffalo. This gene can also be used as a marker to study the evolution and conservation of this family, which have shown a rapid diversification.

## Methods

The program LAGAN of the mVISTA suite (<http://genome.lbl.gov/vista/index.shtml>) [25] was used for the multiple comparison of the published genomic sequences in cattle (AC\_000178:735683–738,292, Taxon ID: 9913), sheep (NW\_011943091:477276–478,942, Taxon ID: 9938), yak (NW\_005392936:256281–257,942, Taxon ID: 72004), goat (NC\_030828:463679–465,342, Taxon ID: 9925), water buffalo (NW\_005785325:305827–307,494, Taxon ID: 89462), bison (NW\_011494393, Taxon ID: 43346), as well as pig (NC\_010443.4:212656–216,850, Taxon ID: 9823) and Arabian camel (NW\_011591329:665988–667,365, Taxon ID: 9838) as outgroups. To determine the identity of the sequences, a mobile window of 100 nucleotides was used. The resulting alignment was used to select the most conserved sequences within the gene and 500 base pairs upstream and downstream for designing a set of primers that can amplify in multiple species. The online program Primer3Plus (version 2, <http://www.bioinformatics.nl/cgi-bin/primer3plus/primer3plus.cgi/>) [26] was used to design the primers with the reference sequence for cattle. The designed primers were tested with genomic DNA from cattle, sheep, goat, yak, buffalo and pig, to determine which set would produce bright bands without unspecific products.

Genomic DNA samples from the collection of the Animal Breeding, Genetics and Genomics Laboratory at the International Programs, College of Agriculture and Life Sciences, Cornell University, were used to amplify a DNA fragment of 1283 bp in 49 samples of 18 different breeds of cattle, as well as 3 individual yak, 4 samples of 3 breeds of river buffalo, 10 of 4 breeds of sheep and 6 samples of 4 breeds of goat (Table 1). For PCR, we amplify a final volume of 20  $\mu$ L, with 1.5 mM MgCl<sub>2</sub>, 100  $\mu$ M of each dNTP, 0.2  $\mu$ M of each oligonucleotide and 1 U of Taq DNA polymerase (Syd Labs Inc., Malden, MA). The amplification was carried out as follows: an initial denaturation at 94 °C (4 min), followed by 35 cycles of denaturation at 94 °C (30 s), annealing at 55 °C

(45 s), and extensions at 72 °C (1 min), finalizing with an extension at 72 °C (10 min).

PCR products were detected on 2.0% agarose gel including a dilution of 1:10,000 of GelRed Nucleic Acid Stain (Biotium, CA, USA) and compared to GENEMate Quanti-Marker 100 bp DNA ladder (BioExpress, UT, USA) for size estimation. Cycle-sequencing of the amplified fragments was carried out on the Applied Biosystems Automated 3730XL DNA Analyzer using Big Dye Terminator (Applied Biosystems, CA, USA) chemistry and AmpliTaq-FS DNA Polymerase. The sequence was visualized using CodonCode Aligner (V 3.5, CodonCode Corporation, MA, USA) to assess the quality and identify their differences. We did a BLAST search to

**Table 1** Species and breeds of Bovidae with the sequence of NDN gene studied and the accession numbers of all the sequences published in the GenBank

Breed	Species	Number	Country	Accession Numbers
Achai	Indicine cattle	3	Pakistan	JX196877-JX196879
Angus	Taurine cattle	3	USA	JX196880-JX196882
Bhagnari	Indicine cattle	3	Pakistan	JX196883-JX196885
Brangus	Indicine x taurine cattle	3	USA	JX196886-JX196888
Cholistani	Indicine cattle	3	Pakistan	JX196889-JX196891
Dajal	Indicine cattle	2	Pakistan	JX196892-JX196893
Dhanni	Indicine cattle	3	Pakistan	JX196894-JX196896
Hereford	Taurine cattle	3	USA	JX196897-JX196899
Holstein	Taurine cattle	3	USA	JX196900-JX196902
Lohani	Indicine cattle	2	Pakistan	JX196903-JX196904
Nari Master	Indicine x taurine cattle	3	Pakistan	JX196905-JX196907
N'Dama	African cattle	2	Nigeria	JX196908-JX196909
Muturu	African cattle	2	Nigeria	JX196910-JX196911
Red Sindhi	Indicine cattle	3	Pakistan	JX196912-JX196914
Sahiwal	Indicine cattle	3	Pakistan	JX196915-JX196917
Sokoto Gudali	Indicine cattle	2	Nigeria	JX196918-JX196919
Tharparker	Indicine cattle	3	Pakistan	JX196920-JX196922
White Fulani	Indicine cattle	3	Nigeria	JX196923-JX196925
Yak	<i>B. gruniens</i>	3	Pakistan	JX196926-JX196928
Nili-Ravi	<i>B. bubalis</i>	1	Pakistan	JX196873
Ravi	<i>B. bubalis</i>	2	Pakistan	JX196874 JX196875
Nili	<i>B. bubalis</i>	1	Pakistan	JX196876
Kajli	<i>O. aries</i>	1	Pakistan	JX196929
Finn	<i>O. aries</i>	6	USA	JX196930-JX196932 JX196935-JX196937
Dorset	<i>O. aries</i>	2	USA	JX196933 JX196934
WAD	<i>O. aries</i>	1	Nigeria	JX196938
Beetal	<i>C. hircus</i>	1	Pakistan	JX196939
White	<i>C. hircus</i>	1	Pakistan	JX196940
WAD	<i>C. hircus</i>	2	Nigeria	JX196941 JX196942
Red Sokoto	<i>C. hircus</i>	2	Nigeria	JX196943 JX196944

find homologous sequences in other mammalian species. We use genomic sequences from cattle (AC\_000178:736183–737,792) as reference to compared to the NDN sequences obtained in this study from the Bovidae species.

Median-joining network algorithm, which allows multi-state data, was used for the genomic sequences of the NDN genes by the Network software (version 4.6.1.0, [www.fluxus-engineering.com](http://www.fluxus-engineering.com)). Synonymous and non-synonymous substitution rates corrected for multiple substitutions were determined using the SNAP program [27]. The deleterious effect of the non-synonymous SNPs found were estimated using the web server PROVEAN (Protein Variation Effect Analyzer, <http://provean.jcvi.org>), which is a software tool that uses a sequence-based prediction algorithm to determine whether an amino acid substitution has an impact on the biological function of a protein [28]. This algorithm allows for the best-balanced separation between the deleterious and neutral amino acids, based on a threshold. The score  $< -2.5$  indicates that the variant is deleterious and  $> -2.5$  score is considered as a neutral variant. ProtParam (<https://web.expasy.org/protparam>), which is a tool use for the computation of the molecular weight, theoretical pI, protein net charge, instability index, aliphatic index and grand average of hydropathicity (GRAVY) [29].

The evolutionary history of the Bovidae species was inferred using the Maximum Likelihood method based on the Tamura-Nei model [30], conducted in MEGA7 [31]. A bootstrap test [32] of 1000 replicates was used to determine the statistical support of the branches in the most likely tree. Initial tree(s) for the heuristic search were obtained automatically by applying Neighbor-Join and BioNJ algorithms to a matrix of pairwise distances estimated using the Maximum Composite Likelihood (MCL) approach, and then selecting the topology with

superior log likelihood value. A discrete Gamma distribution was used to model evolutionary rate differences among sites (5 categories (+G, parameter = 0.5094)).

In addition, a Bayesian phylogenetic analysis was conducted using Mr.Bayes, v 3.2.1 [33], implementing the general time-reversible (GTR) model with the rate at each site as random variable with a gamma distribution (G) and a proportion of invariable sites. Markov chain Monte Carlo (MCMC) chains were carried out for 10,000,000 generations.

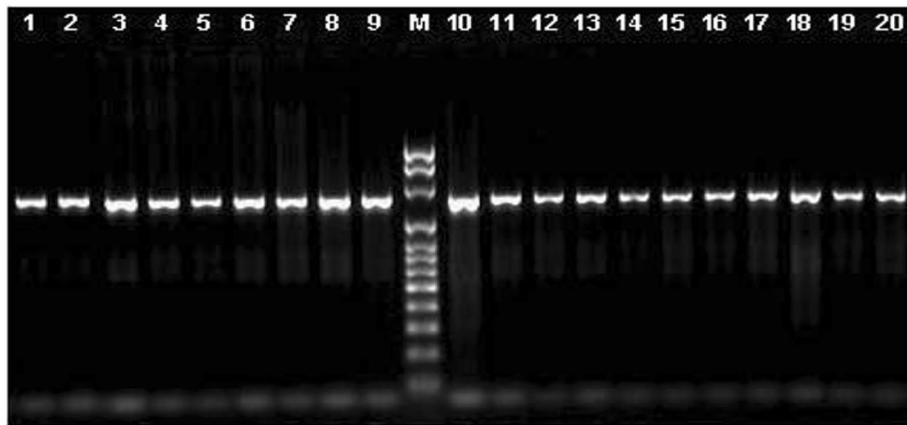
## Results

We used several sets of primers but only the pair FCd/RCd (Table 2) produced a single fragment of 1283 bp in all the species tested, except for pig, (Fig. 1). The amplified fragment contained the entire coding region, and those primers were located in the most conserved regions in the 5' and 3' UTRs (Fig. 2). The primers designed on the promoter regions produced unspecific fragments, even though this region is highly conserved. The primers selected here for the amplification of the coding region and its sequencing (Table 2) seem to be useful to study this gene sequence in all species from the five families of Ruminantia. In fact, primers specific for the other groups of Cetartiodactyla can be obtained to study the gene sequence in other species. Using the three primers we obtained good quality sequence for 1194 bp in all the samples analyzed, which included the entire coding region, plus 15 bp upstream and 200 bp downstream.

Looking into sequence variation in cattle, nine SNPs were detected, eight in the coding region and one in the 3'UTR (Table 3), corresponding to 0.18 SNP per individual. Only four of these SNP produced amino acid changes (non-synonymous substitutions), three in the MHD region and the other in the C-terminal region. Of these amino acid changes in the MHD, three are predicted to be deleterious,

**Table 2** Nucleotide changes in the sequence of the oligonucleotides used as primers for PCR and sequencing compared to the sequences in several species of Cetartiodactyla analyzed in this study

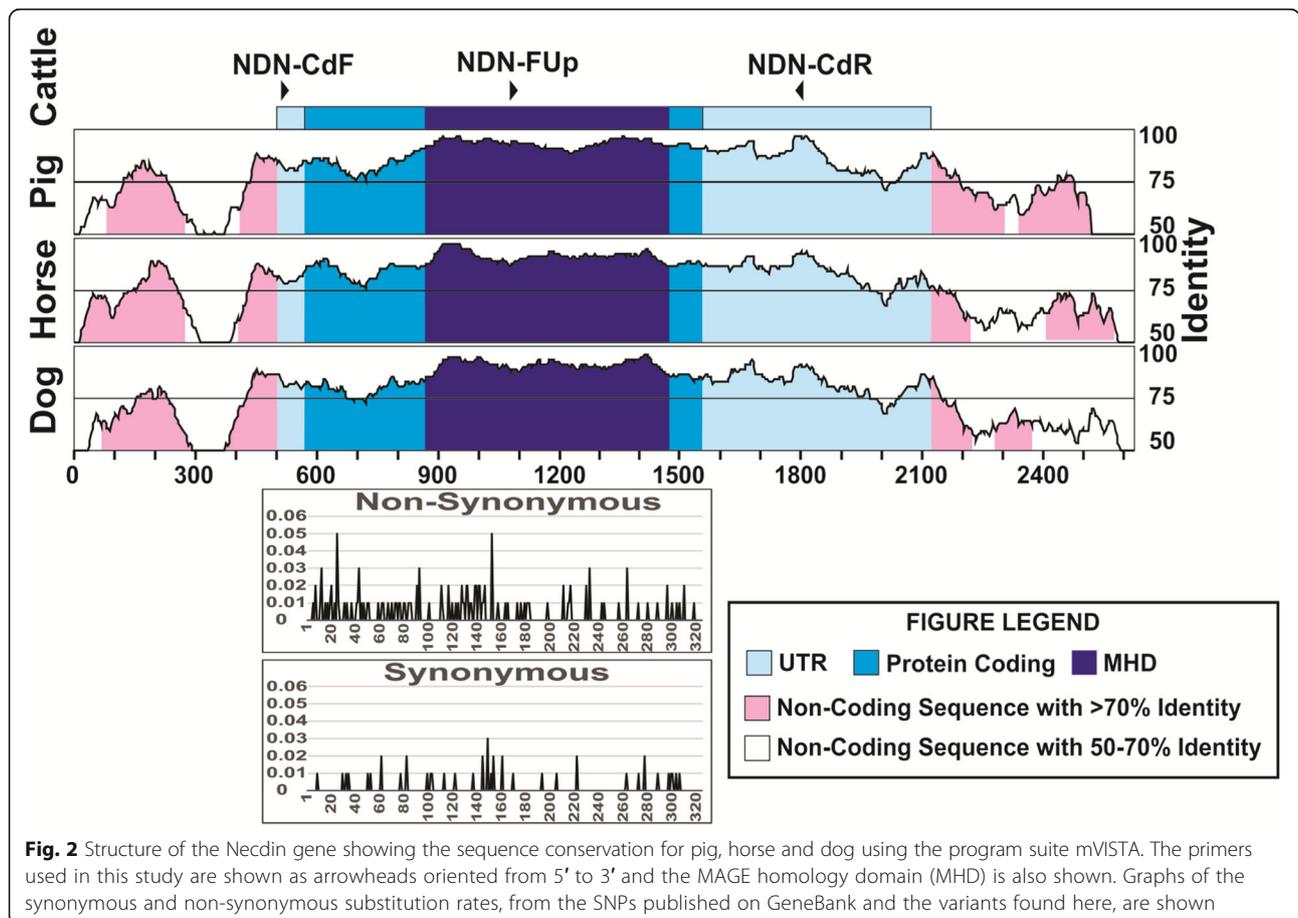
Gene	NDN-FCd	NDN-FUp	NDN-RCd
Cattle	GGAAAGCAGACTCGAAGAGC	GAGTTTTCGCTGGTCAAAGC	GCTTTCGCTTTTGTGCTACC
Yak	GGAAAGCAGACTCGAAGAGC	GAGTTTTCGCTGGTCAAAGC	GCTTTCGCTTTTGTGCTACC
Bison	GGAAAGCAGACTCGAAGAGC	GAGTTTTCGCTGGTCAAAGC	GCTTTCGCTTTTGTGCTACC
Buffalo	GGAAAGCAGACTCGAAGAGC	GAGTTTTCGCTGGTCAAAGC	GCTTTCGCTTTTGTGCTACC
Goat	GGAAAGCAGACTCGAAGAGC	GAGTTTTCGCTGGTCAAAGC	GCTTTCGCTTTTGTGCTACC
Sheep	GGAAAGCAGACTCGAAGAGC	GAGTTTTCGCTGGTCAAAGC	GCTTTCGCTTTTGTGCTACC
Oryx	GGAAAGCAGACTCGAAGAGC	GAGTTTTCGCTGGTCAAAGC	GCTTTCGCTTTTGTGCTATC
Giraffe	GGAAAGCAGACTCGAAGAGC	GAGTTTTCGCTGGTCAAAGC	GCTTTCGCTTTTGTGCTACC
Deer	GGAAAGCAGACTCGAAGAGC	GAGTTTTCGCTGGTCAAAGC	GCTTTCGCGITTTGTGCTACC
Pig	GGAGCGCAGAGTCTGAAGAGC	GAGTTCCTCGCTGGTCAAAGC	GCTTTCGCTTTTGTGCTACC
Camel	GGAGCGCACCTCTGAAGAGC	GAGTTCCTCGCTGGTCAAAGC	GCTTTCGCTTTTGTGCTACC



**Fig. 1** PCR amplification of the 1283 bp fragment containing the coding region of the Neccdin (NDN) gene in cattle (lines 1–9) yak (lines 10–11) buffalo (lines 12–14) goats (lines 15–17) and sheep (lines 18–20) run in a 1.5% agarose gel and stained with GelRed (Phenix Research Products Candler NC). M: Molecular weight marker 100pb (GENEMate Quanti-Marker UT)

and these were found in heterozygous individuals, carrying this mutation. Two of these SNPs changed from basic to polar, non-charged amino acids and one from polar, non-charged to the same type but a significant change in the structure. We found six additional individuals showing one and four individuals showing three heterozygous sites,

all of these individuals belonging to the indicine and African breeds. When analyzing all the SNPs found here and those in the cattle database (Fig. 2), we found that the N-terminal, MHD and C-terminal showed similar percentages of nucleotides with non-synonymous substitutions (12.3, 12.8 and 12.5%, respectively), and were much higher



**Fig. 2** Structure of the Neccdin gene showing the sequence conservation for pig, horse and dog using the program suite mVISTA. The primers used in this study are shown as arrowheads oriented from 5' to 3' and the MAGE homology domain (MHD) is also shown. Graphs of the synonymous and non-synonymous substitution rates, from the SNPs published on GeneBank and the variants found here, are shown

**Table 3** The nucleotide changes phylogenetically important in the NDN sequences of cattle, yak, buffalo, goat and sheep, and the amino acid changes when appropriate

Species	N	97	111	129	139	143	150	154	161	162	165	169	174	192	202	207	208	253
Cattle	49	A	A	T	A	G/A	C	C	A	G	C/T	T	A	C	T	G	A	C
Yak	3	A	A	T	A	G	C	C	A	G	C	T	A	C	T	G	A	C
Buffalo	4	G	A	T	A	G	C	C	G	G	C	C	A	C	C	G	A	C
Goat	6	G	A	C	A	G	T	A	G	G/C	C	C	G	A/C	C/T	G/A	A	C
Sheep	10	G	G	C	G	G	C	A	G	G	C	C	G	A	C	G	C	T
AA Change		N→D			S→G			P→T	E→G	E→D		S→P			S→P			P→S
Species	306	336	338	403	426	444	471	505	516	526	645	654	663	678	762	768	813	854
Cattle	G	A	C	T	G	T	C	C/T	C/T	C/T	A	G	C	C	C	A	C	A/G
Yak	G	A	C	T	G	T	C	C	C	C	A	G	C	C	G	A	C	A
Buffalo	G	A	C	T	A	T	C	C	C	A	A	C	C	C	A	T	A	A
Goat	C	A/G	C	T	G	C	C	C	C	C	G	G	C	C	C	C	C	A
Sheep	C	A	G	T/C	G	C	C/A	C	C	C	G	G	C/T	C/T	C	A	C	A
AA Change	E→D	T→S	T→S	Y→H				R→C		R→C								Y→C
Species	880	891	906	938	973	998	1016	1037	1038	1063	1082	1099	1108	1117	1122	1173	1231	1238
Cattle	A	C	C/A	C	T	C	A/C	C	A	A	-	G	G	A	A/T	G	G	T
Yak	A	C	C	C	C	C	A	C	A	A	-	G	G	A	A	G	G	T
Buffalo	A	C	C	C	C	A	A	C	C	A	-	G	A	A	A	G	G	A
Goat	A	C	C	A	C	A	A	C	C	G	C	G	G	A	A	A	A	T
Sheep	A/G	C/T	C	A	C	A	A	C/G	C	G	-	G/A	G	G	A	A	A	T
AA Change	S→G			T→N		A→E	H→P	S→C										

The numbering of the sequences was started at the transcription start site. The nucleotide in light blue represent the coding region while the darker blue represents the MAGe homology domain of the protein. N is the number of animals studied. The AA change shown in green are predicted to be neutral to the protein function, while those shown in red are predicted to be deleterious by PROVEAN software

than the percentage of synonymous substitutions (2.5, 2.6 and 4.9%, respectively).

In sheep, eight SNPs were also detected, which corresponds to 0.80 SNP per individual. Seven of the SNPs were located in the coding region and one in the 3'UTR, with three producing amino acid changes, two of which were also located in the MHD. One of these changes in the MHD was predicted to be deleterious, since it changed from a polar, non-charged to a basic amino acid. This mutation was also found in a heterozygous individual. No other heterozygous site was found in sheep. In goats, five SNPs were found, corresponding to 0.83 SNP per individual, and were localized in the coding region, but none at the MHD, and three of which produced

amino acid changes. One heterozygous site was found among the individuals analyzed.

Of the 26 nucleotide changes that differentiate the species of Bovidae studied, only 11 produced amino acid changes (non-synonymous substitutions). The analysis of all the polymorphic sites found among these species shows that N-terminal region contain 16.2% of the nucleotides that were polymorphic, while 3.7% were polymorphic in the MHD region, and 3.5% in the C-terminal region. The analysis of the biochemical characteristics in the most common form of the proteins in each species show very little difference in molecular weight, pI, net charge, instability index, aliphatic index and GRAVY (Table 4) in the Bovidae species, except for sheep, which

**Table 4** Characteristics of the NDN proteins in Bovidae deduced from their sequences, using the web tool ProtParam

Species	Molecular Weight	pI	Net Charge	Instability Index	Aliphatic Index	GRAVY <sup>a</sup>
<i>Bos taurus</i>	36,590	8.88	+4	59.4	80.8	-0.414
Yak	36,590	8.88	+4	59.4	80.8	-0.414
Bison	36,590	8.88	+4	59.4	80.8	-0.414
Buffalo	36,597	8.67	+3	60.0	80.5	-0.425
Goat	36,597	8.67	+3	60.0	80.5	-0.425
Sheep	36,615	8.88	+4	57.40	80.5	-0.439
Pig	36,382	9.05	+5	59.6	81.1	-0.381
Horse	36,406	8.34	+2	61.2	79.0	-0.415
Dog	36,469	9.03	+5	62.5	77.8	-0.459

<sup>a</sup>grand average of hydropathicity

had a higher molecular weight, instability index and GRAVY. The proteins in pig, horse and dog did show differences in several characteristics, especially in molecular weight, net charge and GRAVY.

The median-joining network analysis of the genomic sequences of the species of Bovidae, including pig, horse and dog as outgroups (Fig. 3), shows that most of the haplotypes in cattle were identical, with only few very closely related haplotypes at a low frequency. The only yak haplotype found was also very closely related to the main cattle haplotype. On the contrary, sheep haplotypes were more diverse in spite of having only 10 individuals studied. In buffalos and goats, we found two haplotypes each, with a major and a minor one.

The phylogenetic analysis of the NDN sequences of the species of Bovidae and other species in the Cetartiodactyla was carried out which included the consensus of the 1194 bp sequence obtained here for cattle, yak, buffalo, sheep and goat, plus the upstream and downstream sequences of these species from the GenBank, to compare the entire sequence of the gene. In addition, we newly assembled and annotated the necdin gene by BLAST searching in the Sequence Read Archive (SRA) database, which contains next-generation sequencing data organized by the submitted sequencing project, in the following species: giraffe (SRA: SRX1624609, SRX1624612 and SRX1624614, Taxon ID: 439328), Oryx (SRA: SRX2880697, Taxon ID: 39411), white-tailed deer (SRA: SRX2056446, Taxon ID: 9874). The consensus sequences for the entire gene in these species were obtained by the alignment of the short reads with overlaps allowing each nucleotide position to be repeated at least 5 times. No polymorphism was found in any of the sequences.

The Maximum Likelihood and Bayesian phylogenetic analysis among the species of Bovidae produced trees

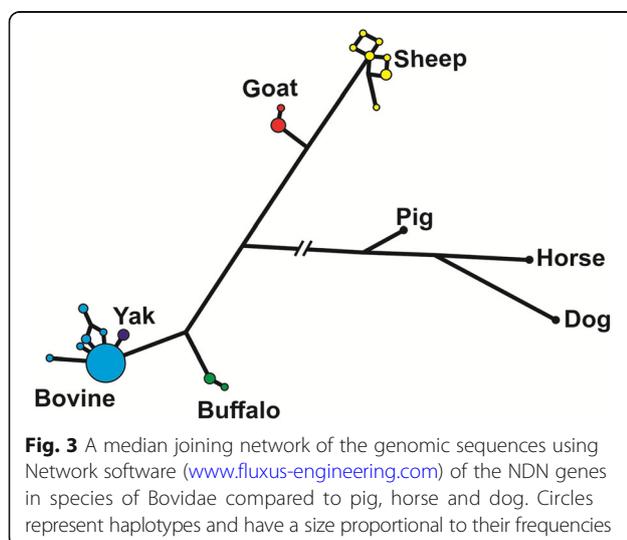
almost identical (Fig. 4), although the statistical support for the Bayesian tree was higher. In this analysis, most branches showed the highest level of support ( $\geq 94\%$ ), with the exception of deer and sheep-goat, demonstrating the utility of these gene as a good phylogenetic marker in this group.

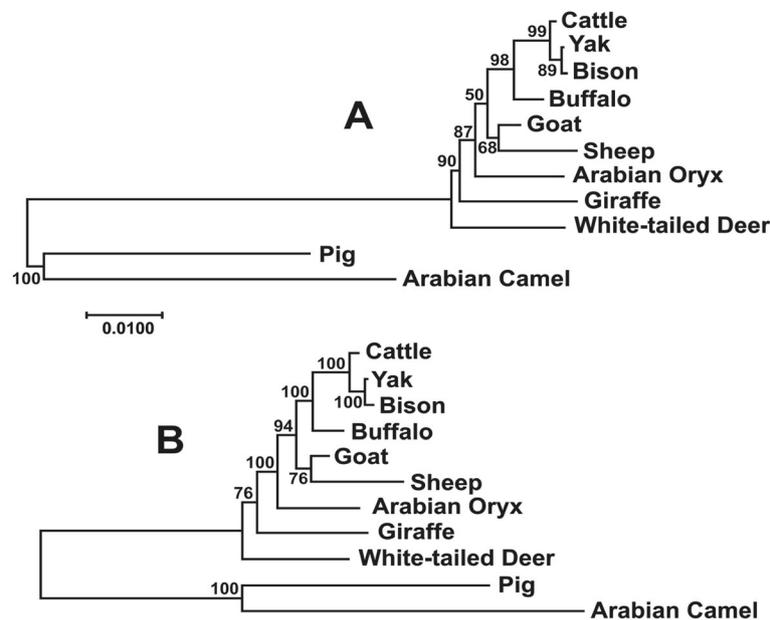
## Discussion

The use of primers localized in the most conserved regions of the 5' and 3' UTRs, should amplify all the species in the family Bovidae, since no variation or only one base change in the sequence of the primers was found in the species analyzed for this family, but should also amplify in the species of Giraffidae, Cervidae, Moschidae and Antilocaprinae. In addition, the use of the primer NDN-FCd and NDN-RCd, modified as shown, should amplify most the species from the other groups of Cetartiodactyla. This would be very useful to study the evolution of the sequence of the gene in all members of Cetartiodactyla, since it has been very difficult to establish a detailed relationship among the different clades and to trace the pattern of evolution in this group, being one of the most diversified mammalian orders.

We found higher proportion of SNPs in sheep and goats than in cattle, even though we analyzed many more animals which were representatives of several cattle breeds originated in three different continents. The lower degree of sequence variation found for cattle, could be related to a bottleneck effect that has been suggested to have occurred during the domestication process, while no trace of bottleneck have been found during the domestication in sheep and goats [34]. However, we found several cattle individuals showing heterozygous sites, which is a sign of hybridization between breeds.

The amino acid sequence of the core functional domain of the NDN protein (aa 83–292) have been reported to be highly conserved between human and mouse (91% identity) suggesting evolutionary conservation due to a key biological function, but the sequence of the N-terminal region (aa 1–82) is less conserved, with about 60% identity [12]. This agrees with the identity of 87.1% in the N-terminal region, when comparing cattle and pig sequences, while identities of 94.1 and 93.7% are shown in the MHD and C-terminal regions. This also agrees with the fact that most of the nucleotide changes differentiating the species analyzed were found towards the N-terminal region. The fact that all of the predicted deleterious SNPs were found in the MHD region highlights the selection pressure that this sequence is under and the importance of this domain for the function of the NDN gene as well as other members of the type II MAGE genes, as this region has been shown to be highly conserved [35].





**Fig. 4** Phylogenetic analysis by Maximum Likelihood (a) and Bayesian (b) methods with the sequences obtained in this study and from GenBank on other Cetartiodactyla species. The percentages of replicate trees in which the associated taxa clustered together in the bootstrap test (1000 replicates for the ML tree and 200,000 iterations for the Bayesian tree) are shown next to the branches. The tree is drawn to scale with branch lengths in the same units as those of the evolutionary distances used to infer the phylogenetic tree which are the number of base substitutions per site. For the ML method, the tree with the highest log likelihood ( $-5444.4659$ ) is shown

All the SNPs found here in the cattle breeds were novel for this species, most likely due to the higher representation of Asian and African breeds whose SNPs have not been well characterized. Interestingly, the SNPs at nucleotides 162 and 202 found in goats, producing non-synonymous changes, coincide with those reported for cattle (rs447229097 and rs434850213, respectively, [www.ncbi.nlm.nih.gov/SNP](http://www.ncbi.nlm.nih.gov/SNP)). Another SNP found at nucleotide 891 in sheep, producing a synonymous change, was also reported in cattle (rs454519300). Thus, these represent SNPs conserved across these species, where balancing selection or a similar force could have maintained them because of a role in their evolution [36].

When analyzing the SNP density in the proximal region of bovine chromosome 21, Frischknecht et al. [21] found that there is a decreased SNP density at the beginning of BTA21 (less than 1000 SNPs per Mb) relative to other locations. The lack of detailed knowledge of the genomic organization, the imprinting status and transcriptional content precluded the analysis of candidate genes, in the study by Frischknecht et al. [21] and other genome-wide association studies. However, to this region, several QTLs have been associated to reproduction [20–22, 24] and early growth [21–23], among others. Thus, detailed studies on the genetic variation on the genes in this region, including NDN, are essential to pinpoint the location and functional analysis of the molecular mechanisms affecting these QTLs.

Studies of genomic imprinting in domestic livestock has focused on imprinted genes influencing fetal growth and development, which are associated with economically important production traits in cattle, sheep and pigs, since this can have major implications for the future of animal breeding, health and management [37], thus this study is a contribution towards the association of the variation found in NDN and the possible functional implication that it can produce.

## Conclusion

We found sufficient variation in the sequence of this gene among the individuals in all the studied species, besides the selection pressure this gene should be subjected to, due to the important functions in the organisms. Because NDN function is associated with cell growth, obesity and behavior, it can have effects in economically important traits in the production of cattle, yak, sheep, goat and buffalo. In addition, the phylogenetic trees constructed with the sequences of NDN showed the same pattern in Bovidae as previous report using nuclear genes, NDN has proven to be suitable for defining the evolutionary pattern in this group, which can be very useful in the phylogenetic reconstruction or to assess the genetic differentiation of the main groups or subgroups within the family, which have been difficult to determine and be useful for conservation genetics.

**Acknowledgements**

Not applicable.

**Funding**

This work was supported by the College of Agriculture and Life Sciences, Cornell University, Ithaca, NY and Pfizer Animal Health (now Zoetis, Inc.). Additional support by National Research Initiative Competitive Grant Program (Grant No. 2006–35205-16864) from the USDA National Institute of Food and Agriculture; USDA-NIFA Research Agreements (Nos. 2009–65205-05635, 2010–34444-20729) and USDA Federal formula Hatch funds appropriated to the Cornell University Agricultural Experiment Station are gratefully acknowledged. We thank the Higher Education Commission of Pakistan for a Visiting fellowship awarded to TH. We also want to acknowledge the Berry College research start-up funds to SOP.

**Availability of data and materials**

All the sequences generated in this study from the different breeds of the species of Bovidae were published in the GenBank and their accession numbers are shown in Table 1.

**Authors' contributions**

MDD, IGI and SOP conceived the project; MDD, TH, HR, SOP and IGI designed and carried out the experiments; MDD, IGI, MEB and SOP analyzed the data; TH, HR, MEB and SOP contributed to the scientific content; MDD, IGI and SOP wrote the manuscript and TH, HR and MEB made critical revisions of the final version. All authors read and approved the final manuscript.

**Ethics approval and consent to participate**

Genomic DNA samples were obtained from the collection of the Animal Breeding, Genetics and Genomics Laboratory at the International Programs, College of Agriculture and Life Sciences, Cornell University. They were collected through different collaboration projects following the ethical guidelines of the International Council for Laboratory Animal Science (ICLAS) and the Cornell University's Institutional Animal Care and Use Committee (IACUC).

**Consent for publication**

Not applicable.

**Competing interests**

The authors declare that they have no competing interests.

**Publisher's Note**

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

**Author details**

<sup>1</sup>Department of Animal Science, Berry College, Mount Berry, GA 30149, USA. <sup>2</sup>Tecnologico de Monterrey, Escuela de Ingeniería y Ciencias, Ave. Epigmenio Gonzalez 500, 76130 Queretaro, CP, Mexico. <sup>3</sup>Department of Molecular Biology, Virtual University of Pakistan, Lahore 54000, Pakistan. <sup>4</sup>African Institute for Biosciences Research and Training, Ibadan, Nigeria. <sup>5</sup>School of Biological Sciences, Georgia Institute of Technology, Atlanta 30332, USA.

Received: 24 February 2018 Accepted: 10 December 2018

Published online: 20 December 2018

**References**

- International Union for Conservation of Nature and Natural Resources. IUCN. The IUCN red list of threatened species. Version 2014.3. 2014. Available at: <http://www.iucnredlist.org>.
- Wang Q, Yang C. The phylogeny of the Cetartiodactyla based on complete mitochondrial genomes. *Intern J Biol*. 2013;5:30–6.
- Montgelard C, Catzeflis FM, Douzery E. Phylogenetic relationships of artiodactyls and cetaceans as deduced from the comparison of cytochrome b and 12S rRNA mitochondrial sequences. *Mol Biol Evol*. 1997;14:550–9.
- Hassanin A, Douzery EJ. Molecular and morphological phylogenies of ruminantia and the alternative position of the Moschidae. *Syst Biol*. 2003;52:206–28.
- McGowen MR, Spaulding M, Gates J. Divergence date estimation and a comprehensive molecular tree of extant cetaceans. *Mol Phylogenet Evol*. 2009;53:891–906.
- Zhou X, Xu S, Yang Y, Zhou K, Yang G. Phylogenomic analyses and improved resolution of Cetartiodactyla. *Mol Phylogenet Evol*. 2011;61:255–64.
- Hassanin A, Delsuc F, Ropiquet A, Hammer C, Jansen van Vuuren B, Matthee C, Ruiz-García M, Catzeflis F, Areskoung V, Nguyen TT, Couloux A. Pattern and timing of diversification of Cetartiodactyla. *Mammalia Laurasiatheria*. As revealed by a comprehensive analysis of mitochondrial genomes. *C R Biol*. 2012;335:32–50.
- Maruyama K, Usami M, Aizawa T, Yoshikawa K. A novel brain-specific mRNA encoding nuclear protein (necdin) expressed in neurally differentiated embryonal carcinoma cells. *Biochem Biophys Res Commun*. 1991;178:291–6.
- Jay P, Rougeulle C, Massacrier A, Moncla A, Mattei MG, Malzac P, Roëckel N, Taviaux S, Lefranc JL, Cau P, Berta P, Lalande M, Muscatelli F. The human necdin gene NDN is maternally imprinted and located in the Prader-Willi syndrome chromosomal region. *Nat Genet*. 1997;17:357–61.
- Cassidy SB, Schwartz S, Miller JL, Driscoll DJ. Prader-willli syndrome. *Genet Med*. 2012;14(1):10–26.
- Muscatelli F, Abruos DN, Massacrier A, Boccaccio I, Le Moal M, Cau P, Cremer H. Disruption of the mouse Necdin gene results in hypothalamic and behavioral alterations reminiscent of the human Prader-Willi syndrome. *Hum Mol Genet*. 2000;9:3101–10.
- Chapman EJ, Knowles MA. Necdin: a multifunctional protein with potential tumor suppressor role? *Mol Carcinog*. 2009;48:975–81.
- Su AI, Wiltshire T, Batalov S, Lapp H, Ching KA, Block D, Zhang J, Soden R, Hayakawa M, Kreiman G, Cooke MP, Walker JR, Hogenesch JB. A gene atlas of the mouse and human protein-encoding transcriptomes. *Proc Natl Acad Sci U S A*. 2004;101:6062–7.
- Miller NL, Wevrick R, Mellon PL. Necdin a Prader-Willi syndrome candidate gene regulates gonadotropin-releasing hormone neurons during development. *Hum Mol Genet*. 2009;18:248–60.
- Yang H, Das P, Yu Y, Mao W, Wang Y, Baggerly K, Wang Y, Marquez RT, Bedi A, Liu J, Fishman D, Lu Z, Bast RC Jr. NDN is an imprinted tumor suppressor gene that is downregulated in ovarian cancers through genetic and epigenetic mechanisms. *Oncotarget*. 2016;7:3018–32.
- Kobayashi M, Taniura H, Yoshikawa K. Ectopic expression of necdin induces differentiation of mouse neuroblastoma cells. *J Biol Chem*. 2002;277(44):42128–35.
- Tseng YH, Butte AJ, Kokkotou E, Yechoor VK, Taniguchi CM, Kriauciunas KM, Cypess AM, Niinobe M, Yoshikawa K, Patti ME, Kahn CR. Prediction of preadipocyte differentiation by gene expression reveals role of insulin receptor substrates and necdin. *Nat Cell Biol*. 2005;7(6):601.
- Chess A. Mechanisms and consequences of widespread random monoallelic expression. *Nat Rev Genet*. 2012;13(6):421–8.
- Savova V, Chun S, Sohail M, McCole RB, Witwicki R, Gai L, Lenz TL. Genes with monoallelic expression contribute disproportionately to genetic diversity in humans. *Nat Genet*. 2016;48(3):231–7.
- Pausch H, Flisikowski K, Jung S, Emmerling R, Edel C, Götz KU, Fries R. Genome-wide association study identifies two major loci affecting calving ease and growth related traits in cattle. *Genetics*. 2011;187(1):289–97.
- Frischknecht M, Bapst B, Seefried FR, Signer-Hasler H, Garrick D, Stricker C, Fries R, Russ I, Sölkner J, Bieber A, Strillacci MG. Genome-wide association studies of fertility and calving traits in Brown Swiss cattle using imputed whole-genome sequences. *BMC Genomics*. 2017;18(1):910.
- Saatchi M, Schnabel RD, Taylor JF, Garrick DJ. Large-effect pleiotropic or closely linked QTL segregate within and across ten US cattle breeds. *BMC Genomics*. 2014 Dec;15(1):442.
- Barnwell CV, Farin PW, Ashwell CM, Farmer WT, Galphin SP Jr, Farin CE. Differences in mRNA populations of short and long bovine conceptuses on day 15 of gestation. *Mol Reprod Dev*. 2016;83(5):424–41.
- Utsunomiya YT, Carmo AS, Neves HH, Carvalho R, Matos MC, Zavarez LB, Ito PK, O'Brien AM, Sölkner J, Porto-Neto LR, Schenkel FS. Genome-wide mapping of loci explaining variance in scrotal circumference in Nelore cattle. *PLoS One*. 2014;9(2):e88561.
- Dubchak I, Ryabov DV. VISTA family of computational tools for comparative analysis of DNA sequences and whole genomes. *Methods Mol Biol*. 2006;338:69–89.
- Untergasser A, Nijveen H, Rao X, Bisseling T, Geurts R, Leunissen JAM. Primer3Plus an enhanced web interface to Primer3. *Nucleic Acids Res*. 2007;35:W71–4.
- Korber B. In: Rodrigo AG, Learn GH, editors. HIV Signature and Sequence Variation Analysis. Computational Analysis of HIV Molecular Sequences. Dordrecht, Netherlands: Kluwer Academic Publishers; 2000. p. 55–72.

28. Choi Y, Chan AP. PROVEAN web server: a tool to predict the functional effect of amino acid substitutions and indels. *Bioinformatics*. 2015;31(16):2745–7.
29. Gasteiger E, Hoogland C, Gattiker A, Duvaud S, Wilkins MR, Appel RD, Bairoch A. Protein identification and analysis tools on the ExPASy server. In: Walker JM, editor. *The proteomics protocols handbook*. New York: Humana Press; 2005. p. 571–607.
30. Tamura K, Nei M. Estimation of the number of nucleotide substitutions in the control region of mitochondrial DNA in humans and chimpanzees. *Mol Biol Evol*. 1993;10:512–26.
31. Kumar S, Stecher G, Tamura K. MEGA7: molecular evolutionary genetics analysis version 7.0 for bigger datasets. *Mol Biol Evol*. 2016;33:1870–4.
32. Felsenstein J. Confidence limits on phylogenies: an approach using the bootstrap. *Evolution*. 1985;39:783–91.
33. Ronquist F, Teslenko M, van der Mark P, Ayres DL, Darling A, Höhna S, Larget B, Liu L, Suchard MA, Huelsenbeck JP. MrBayes 3.2: efficient Bayesian phylogenetic inference and model choice across a large model space. *Syst Biol*. 2012;61:539–42.
34. Taberlet P, Coissac E, Pansu J, Pompanon F. Conservation genetics of cattle sheep and goats. *C R Biol*. 2011;334:247–54.
35. De Donato M, Peters SO, Hussain T, Rodolfo H, Thomas BN, Babar ME, Imumorin IG. Molecular evolution of type II MAGE genes from ancestral MAGED2 gene and their phylogenetic resolution of basal mammalian clades. *Mamm Genome*. 2017;28(9–10):443–54.
36. Feder ME, Mitchell-Olds T. Evolutionary and ecological functional genomics. *Nat Rev Genet*. 2003;4(8):649–55.
37. Magee DA, Spillane C, Berkowicz EW, Sikora KM, MacHugh DE. Imprinted loci in domestic livestock species as epigenomic targets for artificial selection of complex traits. *Anim Genet*. 2014;45:25–39.

**Ready to submit your research? Choose BMC and benefit from:**

- fast, convenient online submission
- thorough peer review by experienced researchers in your field
- rapid publication on acceptance
- support for research data, including large and complex data types
- gold Open Access which fosters wider collaboration and increased citations
- maximum visibility for your research: over 100M website views per year

**At BMC, research is always in progress.**

Learn more [biomedcentral.com/submissions](https://biomedcentral.com/submissions)

