In silico analyses of CD14 molecule reveals significant evolutionary diversity potentially associated with speciation and variable immune response in mammals

Olanrewaju B Morenikeji [1] , **Bolaji N Thomas** [Corresp. 1]

[1] Biomedical Sciences, Rochester Institute of Technology, Rochester, NY, United States of America

Corresponding Author: Bolaji N Thomas
Email address: bntsbi@rit.edu

Cluster differentiation gene (CD14) is a family of monocyte differentiating genes that works in conjunction with lipopolysaccharide binding protein (LBP), forming a complex with TLR4 or LY96 to mediate innate immune response to pathogens. In this report, we used different computational methods to elucidate the evolution of CD14 gene coding region in 14 mammalian species. Our analyses identified leucine rich repeats (LRRs) as the only significant domain across the CD14 protein of the 14 species with varying frequencies. Importantly, we found signal peptides located at mutational hotspots demonstrating this gene has ancient conservation across these species. Out of the 10 selected variants analyzed in this study, only 6 were predicted to possess significant deleterious effect. Our predicted protein interactome showed a significant varying protein-protein interaction with CD14 protein across the species. This is important for drug target and therapeutic manipulation for the treatment of many diseases. We conclude that these results contribute to our understanding of the CD14 molecular evolution, which underlays varying species response to complex disease traits.

PeerJ Preprints | https://doi.org/10.7287/peerj.preprints.27668v1 | CC BY 4.0 Open Access | rec: 19 Apr 2019, publ: 19 Apr 2019

In silico analyses of CD14 molecule reveals significant evolutionary diversity potentially associated with speciation and differential immune response in mammals

Olanrewaju B. Morenikeji, Bolaji N. Thomas

Department of Biomedical Sciences, Rochester Institute of Technology, Rochester NY

Olanrewaju B. Morenikeji: obmhst@rit.edu

Bolaji N. Thomas: bntsbi@rit.edu

***Corresponding author:**

Dr. Bolaji N. Thomas

Department of Biomedical Sciences

Rochester Institute of Technology

153 Lomb Memorial Drive

Rochester NY 14623

Office: (585) 475-6382

Fax: (585) 475-5809

Email: bntsbi@rit.edu

PeerJ Preprints | https://doi.org/10.7287/peerj.preprints.27668v1 | CC BY 4.0 Open Access | rec: 19 Apr 2019, publ: 19 Apr 2019

Abstract

Cluster differentiation gene (CD14) is a family of monocyte differentiating genes that works in conjunction with lipopolysaccharide binding protein (LBP), forming a complex with TLR4 or LY96 to mediate innate immune response to pathogens. In this report, we used different computational methods to elucidate the evolution of CD14 gene coding region in 14 mammalian species. Our analyses identified leucine rich repeats (LRRs) as the only significant domain across the CD14 protein of the 14 species with varying frequencies. Importantly, we found signal peptides located at mutational hotspots demonstrating this gene has ancient conservation across these species. Out of the 10 selected variants analyzed in this study, only 6 were predicted to possess significant deleterious effect. Our predicted protein interactome showed a significant varying protein-protein interaction with CD14 protein across the species. This is important for drug target and therapeutic manipulation for the treatment of many diseases. We conclude that these results contribute to our understanding of the CD14 molecular evolution, which underlays varying species response to complex disease traits.

Keywords: CD14, mammals, species, immune response, evolution, *in silico*

Introduction

Cluster of differentiation 14 (CD14) gene is a surface differentiation antigen preferentially expressed on mammalian monocytes, neutrophils, macrophages, and plasma cells (Baumann *et al.,* 2010; Tang *et al.,* 2017). CD14 is important in initiating a robust immune response against microbial pathogens by mediating innate immune response, in concert with several other proteins. It is a co-receptor with Toll-like receptor-4 (TLR4) to activate several intracellular signaling pathways that lead to the synthesis and release of inflammatory cytokines, antimicrobial peptides, chemokine, and other co-stimulatory molecules which in turn interact with the adaptive immune system (Hartel *et al.,* 2008). Comparative studies have shown that two or more proteins can have common evolutionary origin thereby sharing structural and functional characteristics (Kanduc 2012). CD14 molecule exists in two forms: soluble (sCD14) or membrane-bound (mCD14) (Panaro *et al.*, 2008; Xue *et al.,* 2012). There are multiple variants of CD14 that are encoded by the same protein due to alternative splicing and as such has been mapped to varying chromosomal locations in different species. For example, it is mapped to chromosome 5 in humans, 7 in cattle and 18 in mouse (Ferrero *et al.,* 1990; Le Beau *et al.,* 1993; Ibeagha-Awemu *et al.,* 2008).

Studies in human, mouse, cattle and sheep have shown that CD14 is significantly involved in innate immunity, playing major roles in susceptibility to tuberculosis, trypanosomosis, malaria and other bacterial infections (Sugawara *et al.,* 2003; Ibeagha-Awemu *et al.,* 2008; Xue *et al.,* 2012; Ojurongbe *et al.,* 2017). Other published reports have shown that there is a higher susceptibility to *Mycobacterium tuberculosis* infection in CD14 knock-out mice when compared to the wild type (Reiling *et al.,* 2002; Weiland *et al.,* 2008). Likewise, single nucleotide polymorphisms (SNPs) in CD14 gene have been associated with higher susceptibility in many disease instances (Oakley *et al.,* 2009; Liu *et al.,* 2012; Xue *et al.,* 2012; Zanoni and Granucci, 2013; Thomas *et al.,* 2015; Xue *et al.,* 2017). In fact, Song *et al*. (2014) reported how genetic heterozygosity modulate disease resistance and progression in cattle infected with bovine tuberculosis. Furthermore, comparative studies have shown that organism relatedness can be traced through their pattern of genetic divergence (Kanduc, 2012; De Donato *et al.,* 2017; Peters *et al.,* 2018).

Several sequence-based methods and tools have been developed to gather evolutionary information in related species via amino acids (aa) sequence variation and conservation of homologous proteins through multiple sequence alignment (MSA) (Hepp *et al.,* 2015, Peters *et al.,* 2018). Similarly, other computational methods are available to identify single nucleotide polymorphism (SNP) variation within and among aa sequences in multiple species, which possibly affects the stability and functionality of such proteins (Ng and Henikoff, 2006; Yue and Moult 2006; Hepp *et al.,* 2015). Many of these tools can predict the effect of SNP occurrence in protein sequences to determine whether they are disease related, deleterious or neutral. Comparative genomics therefore is a powerful tool to elucidate variants and effects among multiple species in order to detect diseases associated to variations. Variations in amino acid sequence have the ability to alter protein structure and functions like ligand binding, protein folding, impaired intracellular transport and reduced stability (Zeron-Medina *et al.,* 2013; Morisseau *et al.,* 2014; Valastyan and Lindquist 2014).

Due to the significance of CD14 gene in several disease cases in humans and other species and considering its involvement in innate immunity, we speculated that there might be evolutionary patterns of similarity and diversification that occurred during speciation, which is important for comparative immune and disease studies in different species. Here, we carried out a detailed comparative study of CD14 protein in different species to elucidate the evolutionary basis for conserved regions, active sites and mutational hotspots, which could lead to novel disease phenotypes. In addition, we examine the diversification in CD14 protein interactions within and across the species, which could be explored for therapeutic development or drug design.

Materials and Methods

Sequence retrieval and multiple sequence alignment

Complete CD14 amino acid sequences of 14 mammals were retrieved from the database of UniProtKB/Swiss-Prot (https://www.uniprot.org/uniprot/?query=CD14&sort=score). The sequences were retrieved for human (P08571), rat (Q63691), mouse (P10810), cattle (Q95122), rabbit (Q28680), monkey (B3Y6B8), gorilla (G3R4C0), sheep (W5QJA2), horse (F6VK89), pig (A7BG66), buffalo (A0A2R4SDF9), goat (ABE68725.1, from NCBI), chimpanzee (B3Y6B4) and yak (L8I9P7). We performed sequence alignment with the Multalin software (http://multalin.toulouse.inra.fr/multalin/), which does a simultaneous alignment of biological sequences with hierarchical clustering. To examine similarity between the sequences, we used SIAS (Sequence Identity And Similarity, http://imed.med.ucm.es/Tools/sias.html) with default BLOSUM62 scoring matrices. Evolutionary tree was constructed from the sampled species through Phylogeny.fr (http://www.phylogeny.fr/index.cgi) online program.

Comparative physicochemical properties of amino acid sequence in the CD14 molecule

The biochemical properties of the amino acids from the 14 mammalian species were computed with ProtParam (www.expasy.org/protparam/). The following properties were computed for each sequence: aliphatic index, instability index, protein net charge, molecular weight, and grand average of hydropathicity (GRAVY) and isoelectric point (pI).

Functional analysis, motif scanning and prediction of signal peptides

We performed functional analysis on the protein sequences in order to classify them in to super families, predict domains, repeats and find important sites that may be relevant in evolution. We scanned for the motif signatures among the amino acid sequences with the combined use of ScanProsite (https://prosite.expasy.org/) (Sigrist et al., 2010) and InterPro, an online program that analyzes protein sequences and classification (https://www.ebi.ac.uk/interpro/). The HAMAP profiles, PROSITE patterns, Pfam global models and PROSITE profiles were all included in the search. Sequence logo of the identified conserved domain in the CD14 protein among the 14 mammalian species was constructed with WebLogo (http://weblogo.berkeley.edu/logo.cgi), to show the graphical view of the region containing the conserved amino acid among the species. Furthermore, we predicted the cleavage sites and the presence of signal peptides in CD14 protein from the 14 mammalian species using SignalP 5.0 server (http://www.cbs.dtu.dk/services/SignalP/), which uses recurrent neural network architecture and deep convolution to classify signal peptides into lipoprotein signal peptides, secretory signal peptides or Tat signal peptides. In order to gain a better understanding of the localization of the protein in each species, we predicted subcellular localizations of CD14 protein using Neural Networks algorithm on DeepLoc-1.0 server (http://www.cbs.dtu.dk/services/DeepLoc/), and the construction of the subcellular pathway hierarchical tree.

Prediction analysis of amino acid substitution

The effect of the amino acid substitution was predicted using the combination of SIFT (Sorting Intolerance from Tolerance), PANTHER (Protein ANalysis THrough Evolutionary Relationship) and PROVEAN (Protein Variation Effect Analyzer). Briefly, we used human CD14 amino acid sequence to query the multiple sequence alignment (MSA) of other mammalian species in this study using SIFT which predict the tolerance or deleterious effect of substitutions for each position in the query sequence. Any position with probability less than 0.05 is classified as deleterious. We selected a total of 10 variants from the mutational hotspots as predicted by SIFT and further estimate the likelihood of the selected variants and their effects on protein function through PROVEAN and PANTHER.

Prediction of protein interactome with CD14 protein in different species

In order to establish specific interaction of CD14 with other molecules as a result of biochemical events during speciation, we used the CD14 amino acid sequence from each mammalian species in this study to predict its association with other protein groups and generate different networks using STRING, a database that predicts protein-protein interactions (https://string-db.org/). This is important in order to examine the diversity shaped by evolution in the association of CD14 gene with other molecule in different organisms. Venn diagrams were constructed for the comparison and visualization of overlapping protein-protein interaction (PPI) among different species using two web based applications

(http://bioinformatics.psb.ugent.be/software/details/Venn-Diagrams and http://bioinfogp.cnb.csic.es/tools/venny/)

Results

Comparative analysis and sequence evolutionary trace

In this study, we examined the evolutionary pattern of CD14 protein sequences in 14 mammalian species. The alignment is conserved within two groups separated into ruminants and non-ruminants. The multiple sequence alignment (MSA) identified leucine (L), aspartic acid (D), lysine (K), glutamic acid (E), valine (V), glycine (G), serine (S) and asparagine (N) as evolutionarily conserved amino acid residues, while others like proline (P), glutamine (Q), methionine (M), alanine (A), phenylalanine (F), isoleucine (I), threonine (T) and were evolutionarily varied. CD14 protein sequence is varied in both percent similarity and identity across the 14 species though they share common evolutionary origin (Figure1, 2). The percent identity of CD14 protein in monkey, gorilla, chimpanzee and human was similar while gorilla shares the closest identity with human (Table 1, 2). Among the ruminants, cattle and yak share the closet similarity compared to buffalo, sheep and goat, although the phylogenetic tree suggests that goat is distantly related. While mouse and rat cluster with the same origin, the analysis show that they share less identity (7.4%) and similarity (13.3%). Rabbit, horse and pig are distantly apart from other species, as they do not share high conservation (Table 2, Figure 2). In all, the sequence of CD14 protein in goat and horse share the least similarity (9.9% and 13.2% for goat and horse respectively) and identity (6.7% and 6.9% for goat and horse respectively) with human.

Physicochemical properties at the CD14 promoter region

ProtParam tool (www.expasy.org/protparam/) was used to compute the physical and chemical properties of CD14 amino acid sequences among the 14 species (Table 3). The aliphatic index of all the species is generally high for all species showing that the protein is thermally stable. A higher instability index was observed in the CD14 molecule of rabbit, pig and monkey (53.0, 46.8 and 45.1 respectively), indicating that the protein in less stable and hydrophobic amino acids occupy majority of the sequence, such as leucine, valine, serine and asparagine, which provides higher tolerance against diseases. The lowest instability index is observed in horse (33.5) and goat (35.1) showing that the protein is more stable in these species. CD14 protein in goat also has the lowest aliphatic index (99.7) while mouse has the highest (107.7). We observed a closer range of molecular weight among the species in this study, although gorilla, monkey, human, chimpanzee and rat had the higher molecular weight with close range (Table 3). Negative net charge, indicative that the protein is more basic than acidic, ranged from -9 to +4 as found in mouse and goat respectively. Goat, horse and gorilla has higher Ip indicating that CD14 molecule is highly basic in these species than others. The GRAVY values obtained were generally positive and higher in ruminants than non- ruminants suggesting the proteins are more hydrophobic, which enhances oligomerization and higher binding capability to different proteins.

Characterization of functional motifs and prediction of signal peptides

The CD14 amino acid sequences of the 14 mammalian species in this study were individually scanned for matches against the InterPro and PROSITE collection of protein signature databases. We found one domain (Leucine-rich repeat (LRR), PS51450) with varying frequency across the 14 species (Figure 3). Comparison of the predicted intra-domain features show one domain in human, two each in gorilla, chimpanzee, monkey, horse and pig, three each in cattle, sheep, buffalo, yak, and mouse, with the highest number (4) found in rat. Figure 4 shows the multiple sequence alignment (MSA) of the homology of LRR domain across the 14 species, showing that leucine, aspartic acid, serine and asparagine are 100% conserved in this region. The sequence logo built from the MSA of the domain is displayed in Figure 5 with the logo showing the relative frequencies of each conserved amino acid and their position in the LRR domain. The domain homology reveals that there is significant conservation of most amino acids in this region.

Furthermore, we predicted the signal peptides, position and secretory pathway of the CD14 amino acids of CD14 in the 14 species under consideration. Our analysis shows that gorilla, human, monkey and chimpanzee share the same signal peptide (VSA-TT) at the same position (19 and 20), with high likelihood (Table 4). Cattle, yak, sheep and buffalo also share the same signal peptide (VSA-DT) and position (20 and 21) although sheep has a different position (19 and 20). We observed a significant variation for the rest of the species in terms of signal peptides and their positions (Table 4). Interestingly, signal peptide for all the species (Figure 6a) except sheep (Figure 6b) share the same subcellular localization in the neural networks.

Mutational analysis of predicted variation

A total of 10 variants were selected from the predicted mutations by SIFT and the effects were tested as deleterious or not in the 14 species with PROVEAN and PANTHER. Our analysis showed that 4 of these variants (D28V, W45H, G62E, L70D) were validated mutations with deleterious effect on all species with 2 others found in few species. These variants cluster in the

C-terminus region of CD14 protein between 20 to 100 amino acids. A closer look suggests that mutational effect on the CD14 protein sequence varied from C-terminus to N-terminus with less mutational effect towards the N-terminus (Table 5). The deleterious mutations observed in our study were all at the C-terminus region thus identifying it as a mutational hotspot while Q100G, V301M, L318I, G335T, L357H and G370K mutation spots were neutral for most species. This might mean that CD14 is less conserved in this region because of evolutionary divergence of all species. However, L-H at position 357 showed a deleterious effect in cattle, yak, pig, gorilla, human, monkey, buffalo and chimpanzee, while there is also a deleterious effect of G-K at position 370 of CD14 in rat.

Protein-protein interaction cluster with CD14 gene in different species

In order to deduce protein-protein interactions (PPI) that evolved through speciation due to co-localization, additive genetic interaction, co-expression or repression and physical association with CD14 in the mammalian species under study, we used STRING to build the protein network based on collection of laboratory experimental results from the database (Figure 7) and segment the gene pool base on our phylogenetic result to build Venn diagrams for each species cluster (Figure 8a, b, c). We could not find any protein network for horse and so was excluded in the analysis. Our result shows that there is significant variation in the CD14 protein interactome across the species (Figure 7). Generally, we found that there were different proteins that clustered with CD14 in all the species. All species had 10 proteins in their cluster except cattle and goat that had eleven. Looking at the Venn diagram, rabbit had the highest CD14 PPI that is not shared with others while 3 protein set (CD14, TLR2 and TLR4) is common to members of this group (Figure 8a). Figure 8b shows the ruminant group, including goat, sheep and yak had no unique gene set, meaning the PPI is duplicated in one or two other members of the group. However, cattle has 8 unique PPI while buffalo has 4 that were not shared with others. CD14 and TLR2 are common to all in this group. Likewise, there were 8 unique PPI in human, 6 in gorilla and none found in monkey and chimpanzee (Figure 8c).

Discussion

Comparative analysis of CD14 protein in this study enhances our understanding of genome plasticity among 14 mammalian species and establishes functional, molecular and structural relationships in different clades which are important in an evolutionary trace. The significant variability in the multiple sequence alignment of CD14 molecule across the species suggests a high evolutionary divergence especially between the ruminant and non-ruminant group. This implies that CD14 amino acid (aa) sequence undergoes significant changes during speciation leading to functional and structural modification in different species. Studies have shown that variation in aa sequences could impact immunogenicity, immunotolerance and immunoreactivity (Tauber, 2004, Kanduc, 2012; Bendl *et al.*, 2014). However, we found that aa residues like leucine (L), glutamic acid (E), lysine (K), valine (V), aspartic acid (D), glycine (G), serine (S) and asparagine (N) are highly conserved, thereby retaining some degree of homology in functional, molecular and structural characteristics. In addition, this reveals the common origin between the mammalian species before divergent speciation. Based on the percentage identity and similarity, monkey, gorilla and chimpanzee are closer to human in their CD14 aa sequence, suggesting a lower degree of variation and this may infer some sort of similar CD14 expression during disease condition.

We also observed that the molecular weight, isoelectric point (Ip), instability index and net charge of CD14 protein for this group of mammals are similar, suggesting a key biochemical and immunological function is retained in these species during evolution (Saha *et al.*, 2013; Ajayi *et al.,* 2018). Of interest, the CD14 sequence in cattle and buffalo were much more than yak, despite their common origin potentially implying that domestication has not affected key biological functions in cattle, and the possibility that buffalo can also be domesticated without loss of function. Furthermore, a higher aliphatic index, net negative charge and GRAVY as shown in the physico-chemical properties of CD14 protein in mouse and rat gives an indication of high concentration of alanine, valine, isoleucine and leucine, reported to influence transcription factors, providing higher tolerance against bacterial and viral infections. This is thought to be an important evolutionary adaptation for these small animals to survive bouts of exposure to diseases in their environment, and may explain the basis for these organisms at times serving as reservoir hosts for many disease pathogens in humans. The general negative net charge of CD14 protein as observed across the species indicates an increasing reactivity and help in its receptor binding mechanism. Therefore, the higher the net charge, the more is the reactivity of the protein.

Interestingly, our motif and signal peptide scan found just one domain and one signal peptide in the entire length of CD14 aa sequence. The number of leucine-rich repeat (LRR) domains vary from species to species. The conservation of LRR domain exists in different number among species. Species with similar number of LRR profile may likely have same immunological implications. This is again a significant signature of evolution. CD14 is a co-receptor that bind with LPS, therefore a higher leucine aa profile in the molecule may accelerate its binding mechanism to receptor in a significant way because the protein plays a significant regulatory role in initiating proper innate immune response. Studies have shown that LRR domain is evolutionarily conserved in most of the innate immune related proteins in vertebrates, invertebrates and plants, providing the innate immune defense especially through the pathogen-associated molecular patterns (PAMPs) (Aylwin and Ramnik 2011). Some reports also stated that there about 2–45 leucine-rich repeats within the LRR domains, containing up to 30 residues. Classifying our mammalian species under study into ruminants versus non-ruminants, we observed that non-ruminants possess a lower number of LRR domain in their CD14 molecule (one domain in human, three in ruminants and four in rat). Notably, rat again possesses the highest number of LRR domains remarkably traceable to selection pressure across the species. Moreover, the aa sequence of this domain is highly conserved for all species under study, and are found towards the C-terminal region of CD14, justifying the fact that aa sequence variation that differentiate species are found close to the N-terminal region (Peters et al., 2018).

Our study additionally reveals varying secretory signal peptides sites in the CD14 molecule across the species. Signal peptides have been identified as hydrophobic amino acids, recognized by the signal recognition particle (SRP) in the cytosol of eukaryotic cells. Secretory signal peptide is a class of signal peptide that allows the export of a protein from the cytosol into the secretory pathway. In this, we found that human, monkey, gorilla and chimpanzee all have the same signal peptide site and position. Cattle, yak, sheep and buffalo also share the same site and position whereas goat did not, confirming why goat is significantly distant to other ruminants in our phylogenetic construction. It is unclear if this is related to disease tolerance compared to other species. However, we noted in our predicted neural network that the subcellular localization of CD14 protein goes from the extracellular through the intracellular and enters the secretory pathway for all the species, except sheep. In sheep, the subcellular localization begins from the nucleus through the mitochondrion, PTS signals and N-terminal sequences before it enters into the secretory pathway. This information may possess potential immunological consequences that will require further analysis and possibly an *in-vitro* validation.

Of most importance, a higher proportion of the predicted mutations occupying the C-terminal region of CD14 show that they are closer to the active site and may have direct structural and functional effects on CD14 protein thereby causing harmful disease phenotype or susceptibility (Malm and Nilssen, 2008). Studies have shown that the leucine-rich repeats at the C-terminal region is required for responses to smooth lipopolysaccharide, whereas the variable region (290 – 375) has been found to be necessary for response to bacterial lipopolysaccharide. Therefore, variation at this region might be traceable to varied exposure and responses to pathogens in the cause speciation.

We observed a higher proportion of deleterious mutational spots in human, monkey, gorilla and chimpanzee occupying the same loci compared to ruminants and other species. This might suggest that the vital residue conservation at this region is due to selection pressure among these species and has been maintained over time possibly because of their role in evolution, resulting in similar biological and immunological function (Feder and Mitchell-Olds, 2003; De Donato et al., 2017; Peters at al., 2018). Therefore, a perturbation of the amino acid sequence at this region could affect the protein folding, ligand binding and other functions which might be lethal or regarded as disease-causing mutation in all mammals. Understanding the molecular variation in the region could help solve the challenge of Mendelian disease phenotypes. We recommend an *in vitro* detailed study of this region in CD14 molecule to elucidate the molecular mechanism affecting functionality of this region. In all, 3 of these mutations have been characterized and verified in humans to cause disruption of active site and loss of protein activities.

Furthermore, we used the STRING database to annotate CD14 protein network with other protein molecules that may have evolved together during speciation. Significantly, we found that CD14 molecule selectively interact with other proteins from species to species. For example, in cattle, CD14 molecule interacts with 8 other proteins, which are not shared with goat, sheep and yak. In a similar vein, buffalo has 4 unique sets of protein that co-express with CD14 protein. Human and gorilla in their group has 8 and 6 genes respectively that uniquely interact with CD14 protein, which are not found in monkey and chimpanzee. These protein interactions are possibly due to the specific molecular or biochemical changes that occur in CD14 protein during selection pressure in different species. This interactome is important to note in order to understand the molecular and biochemical cascade that is shaped by evolution, which is useful for species drug target design and therapeutic treatment of many diseases. Several studies have shown that molecular association between chains of different protein molecules is geared by the electrostatic force like hydrophobic effects which define specific bimolecular interaction in different organism (Arkin et al., 2014; De Las and Fontanillo, 2010; Chen et al., 2013). The modulation of this interaction may be useful as putative therapeutic targets for disease treatment in many species. Ivanov et al (2013) have used the interaction of Tirobifan with glycoprotein IIb/IIIa as an inhibitor for cardiovascular drug discovery, likewise the interaction of Maraviroc and CCR5-gp120 for anti-HIV drug.

As shown earlier, there are variations in the number of the LRR domain among these species, possibly the lesser number of LRR domain in human is supplemented or accounted for by the functionality of other genes in the network (Thakur and Shankar, 2016). From our physicochemical properties, CD14 is classified as hydrophobic across the species due to higher proportion of LRR. The varying degree of LRR among these species is thought to affect the electrostatic force created by the hydrophobic effects of the protein. Some studies in mammals showed that diverse fungal, bacterial, viral and parasite components are sensed by the LRR domain of proteins like NOD-like receptors and Toll-like receptors. Likewise, about 34 leucine-rich repeat proteins have been associated with diseases in human. Obviously, evolutionary events have shaped the protein-protein interaction of CD14 in different species, which is thought to be significant to varying degrees of disease susceptibility and pathogen selection.

Conclusion

We have used computational methods to gather information on CD14 protein in 14 mammals. Our *in silico* comparison of CD14 amino acid sequences among these species gave molecular evidence of evolutionary events that occurred during speciation, potentially of significance in modulating innate immune response to pathogenic challenges. Obviously, this gene has been subjected to selection pressure due to sufficient sequence variation we found from one species to another. We identified mutational hotspots with damaging effects in human and other species. In particular, the signal peptides located in these mutational hotspots is of importance in immunological studies. The variants identified in this study can be further subjected to validation through *in vitro* analysis. Since CD14 molecule is an essential molecule in initiating proper immune response to pathogens and the precursor of a robust adaptive immune response, our study highlights the effect of mutations on protein structure and disease outcome, protein-protein interaction that is essential for drug design strategies and therapeutic manipulations to treat many diseases. Finally, these results contribute to our understanding the evolutionary mechanism that underlie species variation in response to complex disease traits.

Acknowledgement

This work was funded by a Laboratory and Faculty Development Award, College of Health Sciences and Technology, Rochester Institute of Technology (BNT). OBM is supported through the American Association of Immunologists Careers in Immunology Fellowship Program. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

References

Ajayi OO, Peter SO, De Donato M, Mujibi FD, Khan WA, Hussain T, Babar ME, Imumorin IG, Thomas BN. 2018. Genetic variation in N- and C-terminal regions of bovine DNAJA1 heat shock protein gene in Africa, Asian and American cattle. *Journal Genomics*. **6**:1-8. doi: 10.7150/jgen.23248

Arkin MR, Tang Y, Wells JA. 2014. Small-molecule inhibitors of protein-protein interactions: progressing towards the dream. *Cell Chemical biology* **3**:301-17. DOI:10.1016/j.chembiol.2014.09.001

Arnesen, T. 2011. Towards a functional understanding of protein N-terminal acetylation. PLoS Biology, **9**(5):e1001074. doi:10.1371/journal.pbio.100107

Bella J, Hindle KL, McEwan PA, Lovell SC. 2008. The leucine-rich repeat structure. *Cell Molecular Life Science*. **65**: 2307-2333. 10.1007/s00018-008-8019-0

Bendl J, Stourac J, Salanda O, Pavelka A, Wieben ED, Zendulka J, Brezovsky J, Damborsky J. 2014. PredictSNP: Robust and Accurate Consensus Classifier for Prediction of Disease-Related Mutations. *PLoS Computational Biology*.**10**: e1003440. doi: 10.1371/journal.pcbi.1003440 PMID: 24453961

Bendl J, Stourac J, Salanda O, Pavelka A, Wieben ED, Zendulka J, Brezovsky J, Damborsky J. 2014. PredictSNP: robust and accurate consensus classifier for prediction of disease-related mutations. *PLoS Computational Biology* **10**: e1003440.

Chen S, Krinsky BH, Long M. 2013. New genes as drivers of phenotypic evolution. *Nature Reviews Genetics*. **14**:645-60. doi: 10.1038/nrg3521.

Choi Y, Chan AP. 2015. PROVEAN web server: a tool to predict the functional effect of amino acid substitutions and indels. *Bioinformatics* **31**:2745–7.

Choi Y, Sims GE, Murphy S, Miller JR, Chan AP. 2012. Predicting the Functional Effect of Amino Acid Substitutions and Indels. PLoS ONE. **7**: e46688. doi: 10.1371/journal.pone.0046688 PMID: 23056405

De Donato M, Peters SO, Hussain T, Rodulfo H, Thomas BN, Babar ME, Imumorin IG. 2017. Molecular evolution of type II MAGE genes from ancestral MAGED2 gene and their phylogenetic resolution of basal mammalian clades. *Mammalian Genome* **28**:443–54.

Dultz E, Hildenbeute M, Martoglio B. Hochman J, Dobberstein B, Kapp K. 2008. The signal peptide of the mouse mammary tumor virus Rem protein is released from the endoplasmic reticulum membrane and accumulates in nucleoli. *Journal of Biological Chemistry*. **283**:9966–76. DOI:10.1074/jbc.M705712200

Feder ME, Mitchell-Olds T. 2003. Evolutionary and ecological functional genomics. *Nature Reviews Genetics* **4**:651-7. DOI:10.1038/nrg1128

Ferrero E, Hsieh C L, Francke U, Goyert SM. 1990. CD14 is a member of the family of leucine-rich proteins and is encoded by a gene syntenic with multiple receptor genes. *Journal of Immunology* **145**:331-336

Härtel C, Rupp J, Hoegemann A, Bohler A, Spiegler J, von Otte S, Röder K, Schultz C, Göpel, W. 2008. 159C >T CD14 genotype-functional effects on innate immune responses in term neonates. *Human Immunology* **69**: 338–343

Hepp D, Gonçalves GL, de Freitas TR. 2015. Prediction of the Damage-Associated Non-Synonymous Single Nucleotide Polymorphisms in the Human MC1R Gene. *PLoS ONE* **10**: e0121812. doi:10.1371/journal.pone.0121812

Huyen Y, Jeffrey PD, Derry WB, Rothman JH, Pavletich NP, Stavridi ES, Halazonetis TD. 2004. Structural differences in the DNA binding domains of human p53 and its C. elegans ortholog Cep-1. *Structure*.**12**:1237–1243.

Ibeagha-Awemu EM, Lee JW, Ibeagha AE, Zhao X. 2008. Bovine CD14 gene characterization and relationship between polymorphisms and surface expression on monocytes and polymorphonuclear neutrophils. *BMC Genetics* **9**:50.

Ivanov SM, Lagunin AA, Pogodin PV, Filimonov DA, Poroikov VV. 2014. Identification of drug-induced myocardial infarction-related protein targets through the prediction of drug-target interactions and analysis of biological processes. Chemical Research in Toxicology. **27**: 1263–1281.

Ivanov SM, Lagunin AA, Pogodin PV, Filimonov DA, Poroikov VV. 2015. Identification of Drug Targets Related to the Induction of Ventricular Tachyarrhythmia Through a Systems Chemical Biology Approach. *Toxicological Sciences*, **145:** 321–336. doi: 10.1093/toxsci/kfv054

Käll L, Krogh A, Sonnhammer EL. 2004. A combined transmembrane topology and signal peptide prediction method. *Journal of Molecular Biology* **338**:1027–1036.

Kamaraj B, Purohit R. 2013. In silico screening and molecular dynamics simulation of disease-associated nsSNP in TYRP1 gene and its structural consequences in OCA3. *Biomedical Research International* 697051. doi: 10.1155/2013/697051 PMID: 23862152

Kamaraj B, Purohit R. 2014. Computational Screening of Disease-Associated Mutations in OCA2 Gene. *Cell Biochemistry and Biophysics*. **68**: 97–109. doi: 10.1007/s12013-013-9697-2 PMID: 23824587

Kanduc D. 2012. Homology, similarity, and identity in peptide epitope immunodefinition. *Journal of Peptide Science* **18**: 487–494. DOI 10.1002/psc.2419

Khan JM, Ranganathan S. 2009. A multi-species comparative structural bioinformatics analysis of inherited mutations in a-D-Mannosidase reveals strong genotype-phenotype correlation. *BMC Genomics* **10**:S33 doi: 10.1186/1471-2164-10-S3-S33

Korber B. 2000. HIV signature and sequence variation analysis. In: Rodrigo Allen G, Learn Gerald H., editors. Computational Analysis of HIV Molecular Sequences. Chapter 4. Dordrecht, Netherlands: Kluwer Academic Publishers; pp. 55–72.

Kutay U, Guttinger S.2005. Leucine-rich nuclear-export signals: born to be weak. *Trends Cell Biology* **15**:121–4

Liu H-H, Hu Y, Zheng M, et al. 2012. *Cd14* SNPs regulate the innate immune response. *Molecular Immunology*, **51**:112-127

Lucchese A, Serpico R, Crincoli V, Shoenfeld Y, Kanduc D. 2009. Sequence uniqueness as a molecular signature of HIV-1-derived B-cell epitopes. *International Journal of Immunopathology and Pharmacology* **22**: 639–646.

Malm D, Nilssen Ø. 2008. Alpha-mannosidosis. Orphanet Journal of Rare Disease. **23**:21. doi: 10.1186/1750-1172-3-21.

Meng QJ, Master AM, Beesley S, Lu WQ, Gibbs J, Parks D, Collins J, Farrow S, Donn R, Ray D, Loudon A. 2008. Ligand modulation of REV-ERBα function resets the peripheral circadian clock in a phasic manner. *Journal of Cell Sciences* **121**: 3629-3635; doi: 10.1242/jcs.035048

Meng X, Noyes MB, Zhu LJ, Lawson ND, Wolfe SA. 2008. Targeted gene inactivation in zebrafish using engineered zinc-finger nucleases. *Nature Biotechnology* **26**:695-701.

Morisseau C, Wecksler AT, Deng C, Dong H, Yang J, Lee KSS, Kodani SD, Hammock BD. 2014. Effect of soluble epoxide hydrolase polymorphism on substrate and inhibitor selectivity and dimer formation. *The Journal of Lipid Research*. **55**: 1131-1138. doi: 10.1194/jlr.M049718.

Ng A, Xavier RJ. 2011. Leucine-rich repeat (LRR) proteins: Integrators of pattern recognition and signaling in immunity. *Autophagy* **7**: 9. https://doi.org/10.4161/auto.7.9.16464

Ng PC, Henikoff S. 2006. Predicting the effects of amino acid substitutions on protein function. Annu Rev Genomics Human Genetics. **7**:61-80. DOI:10.1146/annurev.genom.7.080505.115630

Nielsen H, Krogh A. 1998. Prediction of signal peptides and signal anchors by a hidden Markov model. Proc. Int. Conf. Intell. Syst. *Molecular Biology* **6**:122–130

Oakley MS, Majam V, Mahajan B, Gerald N, et al. 2009. Pathogenic roles of CD14, galectin-3 and OX40 during experimental cerebral malaria in mice. *PLoS ONE,* **4**(8):e6793

Ojurongbe O, Funwei RI, Snyder T, Aziz N, Li Y, Falade C, Thomas BN. 2017. Genetic diversity of CD14 promoter gene polymorphism (rs2569190) is associated with regulation of parasitemia but not susceptibility to *Plasmodium falciparum* infection. *Infectious Diseases: Research and Treatment,* **10**:1-6. doi: 10.1177/1178633617726781.

Panaro MA, Cianciulli A, Gagliardi N, Mitolo CI, Acquafredda A, Cavallo P, Mitolo V. 2008. CD14 major role during lipopolysaccharide-induced inflammation in chick embryo cardiomyocytes, *FEMS Immunology & Medical Microbiology,* **53**:35–45. https://doi.org/10.1111/j.1574-695X.2008.00397.x

Park KJ, Kanehisa M. 2003. Prediction of protein subcellular locations by support vector machines using compositions of amino acids and amino acid pairs. *Bioinformatics* **19**:1656–1663.

Peters SO, De Donato M, Hussain T, Rodulfo H, Babar ME and Imumorin IG. 2018. Sequence variation of necdin gene in Bovidae. *Journal of Animal Science and Technology* **60**:32. https://doi.org/10.1186/s40781-018-0191-7

Reiling N, Hölscher C, Fehrenbach A, Kröger S, Kirschning CJ, Goyert S, Ehlers S. 2002. Cutting edge: Toll-like receptor (TLR)2- and TLR4-mediated pathogen recognition in resistance to airborne infection with Mycobacterium tuberculosis. Journal of Immunology **169**:3480–3484

Rivas JL, Fontanillo C. 2010. Protein–Protein Interactions Essentials: Key Concepts to Building and Analyzing Interactome Networks. *PLoS Computational Biology*. https://doi.org/10.1371/journal.pcbi.1000807

Saha R, Saha N, Donofrio RS, Bestervelt LL. 2013. Microbial siderophores: a mini review. *Journal of Basic Microbiology*. **53**: 303–317. https://doi.org/10.1002/jobm.201100552

Sigrist CJA, Cerutti L, de Castro E, Langendijk-Genevaux PS, Bulliard V, Bairoch A, Hulo N 2010. PROSITE, a protein domain database for functional characterization and annotation. *Nucleic Acids Research* 38(Database issue):D161–D166

Song Y, Sun L, Guo A, Yang L. 2014. Toll-like receptor 6 gene polymorphisms increase the risk of bovine tuberculosis in Chinese Holstein cattle. *Acta Histochemistry*. 116:1159-62. doi: 10.1016/j.acthis.2014.06.004.

Sugawara S, Yang S, Iki K, Hatakeyama J, Tamai R, Takeuchi O, Akashi S, Espevik T, Akira S, Takada H. 2001. Monocytic cell activation by Nonendotoxic glycoprotein from Prevotella intermedia ATCC 25611 is mediated by toll-like receptor 2. *Infectious Immunology* **69**:4951-4957. DOI: 10.1128/IAI.69.8.4951-4957.2001

Tamura K, Stecher G, Peterson D, Filipski A, Kumar S. 2013. MEGA6: molecular evolutionary genetics analysis version 6.0. *Molecular Biology Evolution* **30**:2725–2729

Tang AT, Choi JP, Kotzin JJ, Yang Y, Hong CC, Hobson N, Girard R, Zeineddine HA, Lightle R, Moore T, Cao Y, Shenkar R, Chen M, Mericko P, Yang J, Li L, Tanes C, Kobuley D, Võsa U, Whitehead KJ, Li DY, Franke L, Hart B, Schwaninger M, Henao-Mejia J, Morrison L, Kim H, Awad IA, Zheng X, Kahn ML. 2017. Endothelial TLR4 and the microbiome drive cerebral cavernous malformations. *Nature*. **18**; 545(7654): 305–310. doi: 10.1038/nature22075

Tauber E, Last KS, Olive PJ, Kyriacou CP. 2004. Clock gene evolution and functional divergence. *Journal of Biological Rhythms*, **19**:445– 458.

Tauber E, Zordan M, Sandrelli F, Pegoraro M, Osterwalder N, Breda C, Daga A, Selmin A, Monger K, Benna C, Rosato E, Kyriacou CP and Costa R. 2007. Natural selection favors a newly derived timeless allele in Drosophila melanogaster. *Science* **316**: 1895–1898.

Thakur R and Shankar J. 2016. *In silico* analysis revealed high-risk single nucleotide polymorphisms in human pentraxin-3 gene and their impact on innate immune response against microbial pathogens. *Frontiers in Microbiology* **7**:192. doi: 10.3389/fmicb.2016.00192

Thomas BN, Donvito B, Cockburn I, Fandeur T, Rowe JA, Cohen JHM, Moulds JM. 2005. A complement receptor-1 polymorphism with high frequency in malaria endemic regions of Asia but not Africa. *Genes and Immunity*. **6**:31-36. doi: 10.1038/sj.gene.6364150

Valastyan JS, Lindquist S. 2014. Mechanisms of protein-folding diseases at a glance. *Disease Model and Mechanism*. **7**:9-14. doi: 10.1242/dmm.013474.

Viriyakosol S, Kirkland TN. 1996. The N-terminal half of membrane CD14 is a functional cellular lipopolysaccharide receptor. *Infection and immunity* **64**:653–656.

Xue Y, Gao WN, Chen F, Ma BB, Zhou F, Hu ZG, Long T, Zhao ZQ. 2018. CD14 gene polymorphisms associated with increased risk of bovine tuberculosis in Chinese Holstein cows. *The Veterinary Journal* **232**:1–5. doi: 10.1016/j.tvjl.2017.11.015

Xue Y, Zhao ZQ, Chen F, Zhang L, Li GD, Ma KW, Bai XF, Zuo YJ. 2012. Polymorphisms in the promoter of the CD14 gene and their associations with susceptibility to pulmonary tuberculosis. *Tissue Antigens* **80**: 437–443.

Yue P, Moult J. 2005. Identification and analysis of deleterious human SNPs. *Journal of Molecular Biology*. **10**:1263-74. DOI:10.1016/j.jmb.2005.12.025

Zanoni I, Granucci F. 2013. Role of CD14 in host protection against infections and in metabolism regulation. *Frontiers in Cellular and Infection Microbiology* **3**:1-6

Zeron-Medina J, Wang X, Repapi E, Campbell MR, Su D, Castro-Giner F, Davies B, Peterse EF, Sacilotto N, Walker GJ, Terzian T, Tomlinson IP, Box NF, Meinshausen N, De Val S, Bell DA, Bond GL. 2013. A polymorphic p53 response element in KIT ligand influences cancer risk and has undergone natural selection. *Cell*. 10:410-22. doi:10.1016/j.cell.2013.09.017

Table 1(on next page)

Table 1: Percentage identity of the CD14 protein across the mammalian species

Table 1: Percentage identity of the CD14 protein across the mammalian species

HUMAN	**100**													
RAT	23.11	100												
MOUSE	10.38	7.37	100											
CATTLE	8.31	10.21	10.92	100										
RABBIT	29.3	15.59	9.28	9.4	100									
GOAT	6.7	9.94	9.56	87.39	10.21	100								
MONKEY	95.19	23.11	10.65	8.57	29.03	6.97	100							
GORILLA	99.2	23.11	10.38	8.31	29.03	6.7	95.46	100						
SHEEP	20.75	12.39	7.92	8.89	19.4	8.35	21.29	21.02	100					
HORSE	6.88	11.29	7.71	8.81	8.26	8.81	6.61	6.88	6.88	100				
PIG	18.49	13.17	10.92	67.56	18.81	60.05	19.3	18.76	19.13	7.98	100			
BUFFALO	8.04	9.94	10.65	96.51	9.4	86.05	8.31	8.04	8.89	8.53	66.75	100		
CHIMP	98.93	23.11	10.92	8.57	29.03	6.97	95.19	99.2	21.02	6.88	19.03	8.31	100	
YAK	8.26	8.6	9.01	42.09	8.33	37.53	8.26	8.26	9.43	7.98	21.44	41.01	8.53	100
	HUMAN	**RAT**	**MOUSE**	**CATTLE**	**RABBIT**	**GOAT**	**MONKEY**	**GORILLA**	**SHEEP**	**HORSE**	**PIG**	**BUFFALO**	**CHIMP**	**YAK**

Min=6.61; Max=99.2; Mean=23.2603296703297; Standard deviation = 26.568543593553

Table 2(on next page)

Table 2: Percentage similarity in the CD14 molecule across the mammalian species

Table 2: Percentage similarity in the CD14 molecule across the mammalian species

Human	**100**													
Rat	27.95	100												
Mouse	14.48	13.38	100											
Bovine	11.52	15.05	15.3	100										
Rabbit	33.6	21.5	15.02%	13.44	100									
Goat	9.91	15.05	13.66	89.27	14.24	100								
Monkey	96.26	27.95	14.48	12.06	33.87	10.45	100							
Gorilla	99.46	27.95	14.2	11.52	33.6	9.91	96.26	100						
Sheep	26.41	16.98	13.93	10.78	25.33	10.78	26.95	26.41	100					
Horse	13.22	17.07	13.77	13.77	14.04	13.49	12.94	13.22	11.84	100%				
Pig	23.32	19.08	14.48	71.58	23.11	64.87	23.59	23.32	22.91	13.22	100			
Buffalo	11.26	14.78	15.3	97.31	13.17	88.73	11.79	11.26	10.78	12.94	71.31	100		
Chimpanzee	99.2	28.22	14.75	11.79	33.6	10.18	96	99.2	26.68	13.22	23.59	11.52	100	
Yak	12	14.24	12.29	45.3	13.97	42.09	12.26	12	13.2	14.04	26.54	44.5	12.26	100
	Human	**Rat**	**Mouse**	**Bovine**	**Rabbit**	**Goat**	**Monkey**	**Gorilla**	**Sheep**	**Horse**	**Pig**	**Buffalo**	**Chimpanzee**	**Yak**

Minimum: 9.91; Maximum: 100; Mean: 32.5818367346939; Standard deviation: 30.994144850177

Table 3(on next page)

Table 3: Physicochemical properties of the CD14 promoter region in selected mammalian species

1 **Table 3: Physico-chemical properties of the CD14 promoter region in selected mammalian species**

Species	Amino acids size	Molecular weight (Da)	Isoelectric point	Instability index	Aliphatic index	Net charge	GRAVY
Cattle	373	39666.79	5.37	41.70	102.06	-5	0.099
Mouse	366	39203.94	5.08	41.16	107.70	-9	0.051
Rabbit	372	39992.29	5.72	52.99	103.33	-5	0.041
Yak	381	40481.75	5.54	41.63	102.23	-4	0.082
Sheep	371	39368.43	5.50	40.27	101.54	-5	0.087
Goat	373	39930.28	8.47	35.07	99.71	+4	0.032
Pig	373	39724.01	5.82	46.83	103.40	-4	0.073
Horse	363	38450.27	6.19	33.47	103.06	-3	0.096
Gorilla	375	40005.15	6.10	42.27	102.80	-3	0.094
Human	375	40076.20	5.84	42.93	101.76	-5	0.083
Monkey	375	40127.19	5.69	45.10	102.80	-6	0.085
Buffalo	373	39756.09	5.84	41.49	101.80	-2	0.099
Chimpanzee	375	40135.34	5.92	43.44	104.61	-4	0.113

PeerJ Preprints | https://doi.org/10.7287/peerj.preprints.27668v1 | CC BY 4.0 Open Access | rec: 19 Apr 2019, publ: 19 Apr 2019

PeerJ Preprints | https://doi.org/10.7287/peerj.preprints.27668v1 | CC BY 4.0 Open Access | rec: 19 Apr 2019, publ: 19 Apr 2019

Table 4 (on next page)

Prediction of signal peptides and properties of the CD14 molecule in mammalian species

PeerJ Preprints | https://doi.org/10.7287/peerj.preprints.27668v1 | CC BY 4.0 Open Access | rec: 19 Apr 2019, publ: 19 Apr 2019

Table 4: Prediction of signal peptides and properties of the CD14 molecule in mammalian species

Species	Amino acids size	Cleavage position	Signal site	Probability	Likelihood	Others
Bovine	373	20 and 21	VSA-DT	0.9750	0.9992	0.0008
Mouse	366	17 and 18	ASP-AP	0.4563	0.9991	0.0009
Rabbit	372	19 and 20	AST-DT	0.6574	0.9981	0.0019
Yak	381	20 and 21	VSA-DT	0.9752	0.9993	0.0007
Sheep	371	19 and 20	VSA-DT	0.9000	0.9453	0.0547
Goat	373	20 and 21	VTA-DK	0.9642	0.9991	0.0009
Pig	373	19 and 20	VSA-AT	0.7699	0.9989	0.0011
Horse	363	14 and 15	AAT-LE	0.2069	0.675	0.325
Gorilla	375	19 and 20	VSA-TT	0.9077	0.9991	0.0009
Human	375	19 and 20	VSA-TT	0.9142	0.9991	0.0009
Monkey	375	19 and 20	VSA-TT	0.9142	0.9991	0.0009
Buffalo	373	20 and 21	VSA-DT	0.9712	0.999	0.001
Chimpanzee	375	19 and 20	VSA-TT	0.9140	0.9991	0.0009
Rat	372	17 and 18	VHA-SP	0.8795	0.9998	0.0002

Table 5(on next page)

Table 5 Prediction of amino acid mutation at the mutational hotspot of CD14 molecules in mammalian species

Table 5 Prediction of amino acid mutation at the mutational hotspot of CD14 molecules in mammalian species

Species	D28V	W45H	G62E	L70D	Q100G	V301M	L318I	G335T	L357H	G370K
Cattle	-3.289	-5.038	-2.998	-2.991	-2.095	-2.131	-1.385	-1.758	-2.634	-2.191
Mouse	-3.437	-4.803	-3.408	-1.635	-2.754	-2.009	-1.408	-1.534	-2.437	-1.828
Rabbit	-2.759	-4.293	-2.910	-4.007	-2.744	-1.969	-1.574	-0.544	-1.865	-2.451
Yak	-3.229	-5.036	-3.081	-3.188	-2.233	-2.097	-1.385	-1.575	-2.668	-2.225
Sheep	-3.559	-4.952	-4.072	-3.206	-2.312	-1.981	-1.246	-1.376	-2.335	-1.695
Goat	-3.919	-4.906	-3.964	-3.390	-2.461	-2.046	-1.476	-0.631	-1.439	-1.601
Pig	-3.712	-5.054	-3.702	-1.873	-2.329	-2.013	-1.637	-1.235	-2.902	-2.052
Horse	-3.742	-4.914	-3.513	-3.524	-2.364	-1.896	-1.412	-0.983	-2.054	-2.067
Gorilla	-3.822	-4.651	-3.216	-2.984	-2.554	-2.049	-1.446	-1.397	-3.050	-2.285
Human	-3.679	-4.680	-3.008	-3.056	-2.756	-2.043	-1.445	-1.395	-3.229	-2.305
Monkey	-3.563	-4.782	-3.238	-3.038	-2.758	-1.933	-1.444	-1.293	-3.089	-2.268
Buffalo	-3.310	-5.083	-3.497	-3.130	-2.169	-2.064	-1.390	-1.427	-3.065	-2.213
Chimpanzee	-3.472	-4.705	-3.154	-3.083	-2.591	-1.905	-1.378	-1.397	-3.088	-2.287
Rat	-3.478	-4.725	-3.373	-1.058	-2.905	-2.038	-1.351	0.464	-2.497	-2.619

Prediction (cutoff= -2.5)

PeerJ Preprints | https://doi.org/10.7287/peerj.preprints.27668v1 | CC BY 4.0 Open Access | rec: 19 Apr 2019, publ: 19 Apr 2019

Figure 1

Figure 1: Multiple sequence alignment of CD14 promoter regions between mammalian species

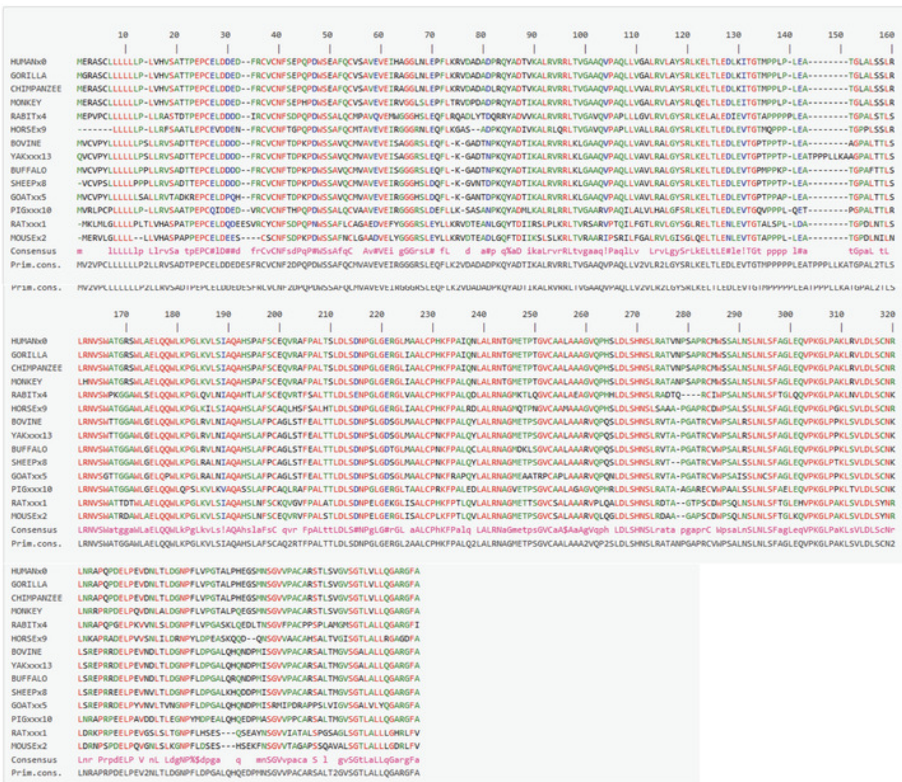


Figure 2

Figure 2: Phylogenetic tree of evolutionary relationships among taxa

The evolutionary history was inferred using the Neighbor-Joining method. The optimal tree with the sum of branch length = 1.48602764. 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. The evolutionary distances were computed using the p-distance method and are in the units of the number of amino acid differences per site. The analysis involved 14 amino acid sequences. The coding data were translated assuming a standard genetic code table.

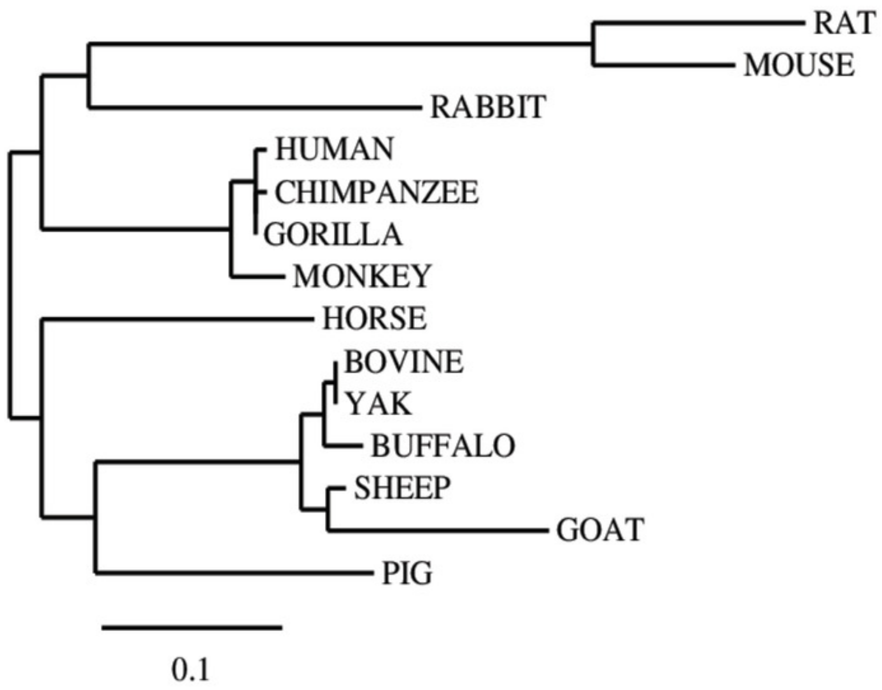


Figure 3

Figure 3: Comparison of predicted intra-domain features of CD14 protein

This comparison shows showing leucine-rich repeat (PS51450), which provide additional information about the structure and function of critical amino acids in 14 mammalian species analyzed

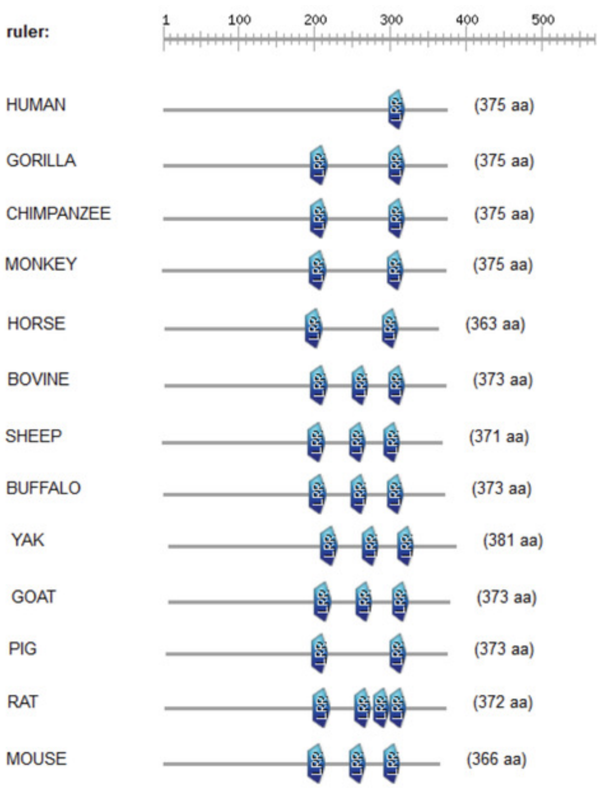


Figure 4

Figure 4: Conserved domain LRR patterns across mammalian species

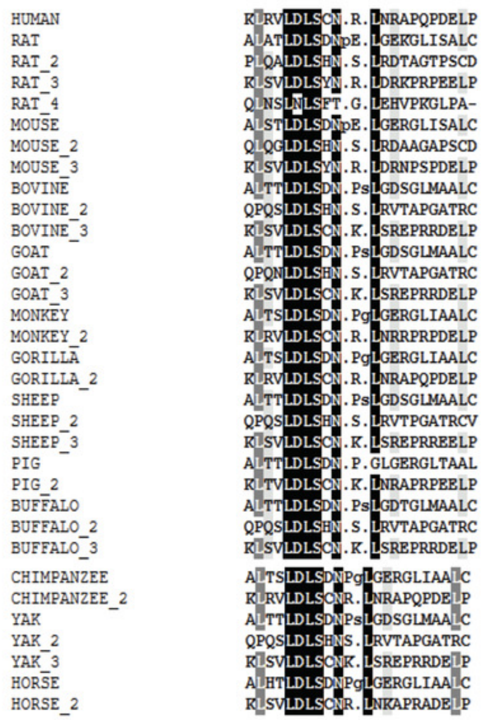
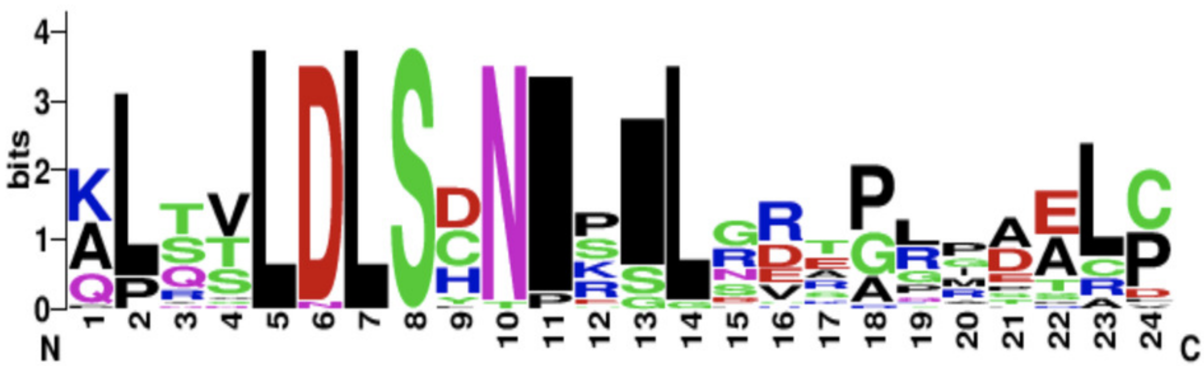


Figure 5

Figure 5: CD14 protein sequence logo displaying the most conserved domain and the positions of amino acids

Sequence logo displaying the most conserved domain and the positions of amino acids starting from the N-terminus on the left to C-terminus to the right. The relative frequency of the amino acids is shown on the y-axis.



PeerJ Preprints | https://doi.org/10.7287/peerj.preprints.27668v1 | CC BY 4.0 Open Access | rec: 19 Apr 2019, publ: 19 Apr 2019

Figure 6

Fig 6. Hierarchical tree-predicted subcellular localizations of CD14 protein using neural network algorithm

6a: Hierarchical tree for all other mammalian species analyzed

6b: Hierarchical tree for sheep only

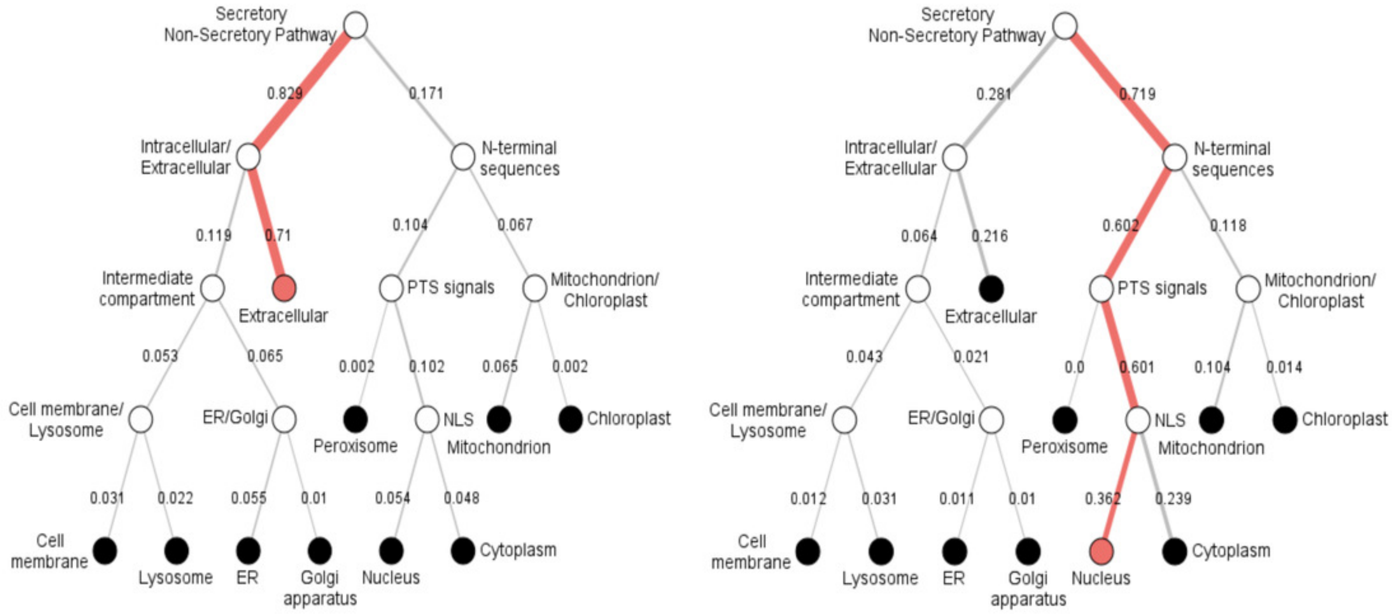


Figure 6a **Figure 6b**

Figure 7

Fig 7. The network view of predicted associations for group of proteins with CD14

The network nodes are proteins. The edges represent the predicted functional associations.The thickness of the line indicate the degree of confidence prediction for the interaction.

Red line: indicates the presence of fusion evidence

Green line: neighborhood evidence

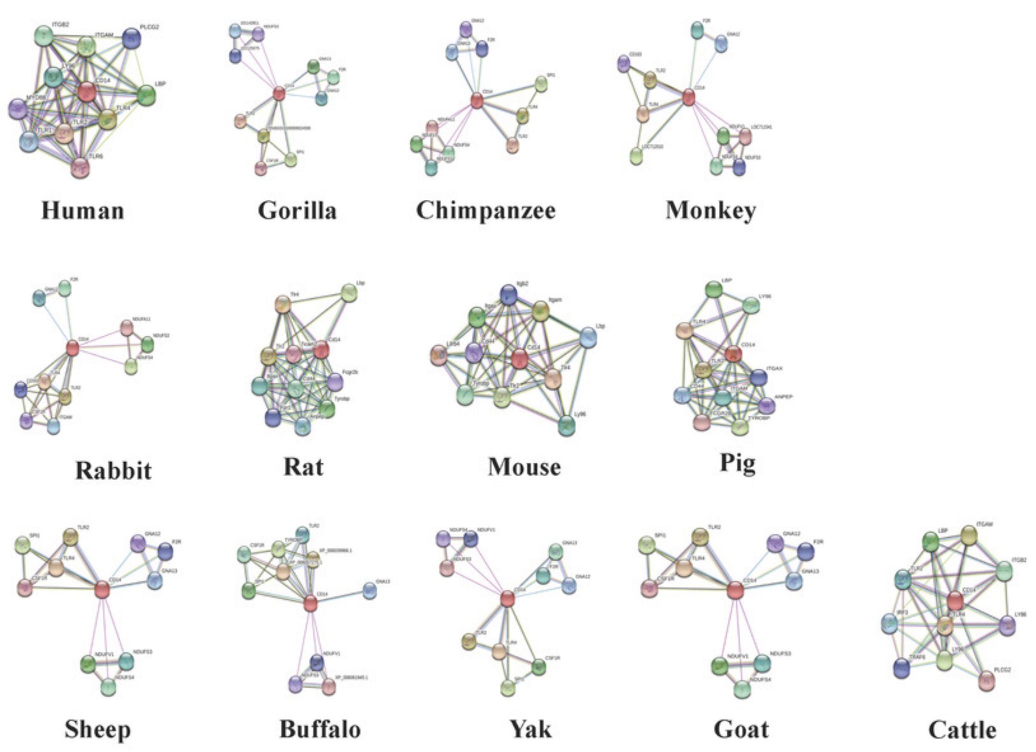
Blue line: co-occurrence evidence

Purple line: experimental evidence

Yellow line: text mining evidence

Light blue line: database evidence

Black line: co-expression evidence



PeerJ Preprints | https://doi.org/10.7287/peerj.preprints.27668v1 | CC BY 4.0 Open Access | rec: 19 Apr 2019, publ: 19 Apr 2019

Figure 8

Figure 8: Venn diagram showing the proportion of intersection and unique genes depicting evolutionary diversity of CD14 molecule

8a: Comparison and visualization of protein interaction with CD14 molecule in pig, rabbit, mouse and rat

8b: Comparison and visualization of protein interaction with CD14 molecule in cattle, yak, sheep, goat and buffalo

8c: Comparison and visualization of protein interaction with CD14 molecule in human, gorilla, chimpanzee and monkey

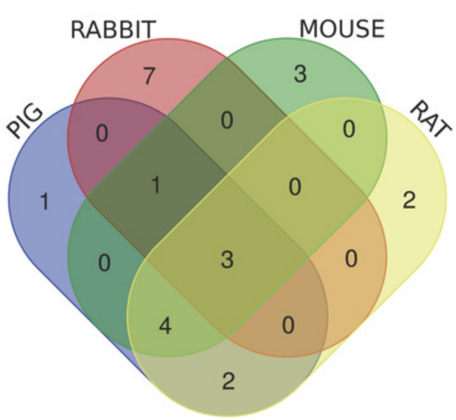


Figure 8a

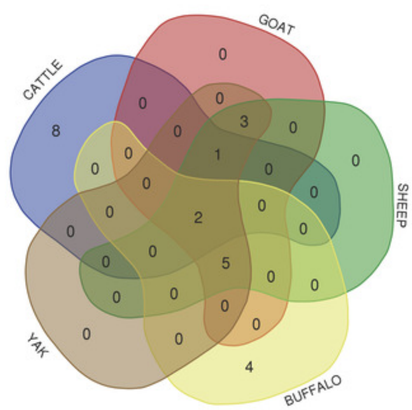


Figure 8b

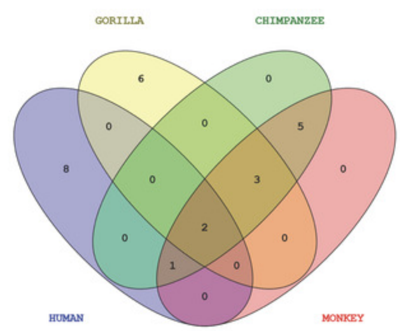


Figure 8c

PeerJ Preprints | https://doi.org/10.7287/peerj.preprints.27668v1 | CC BY 4.0 Open Access | rec: 19 Apr 2019, publ: 19 Apr 2019