

RESEARCH ARTICLE

Open Access

# Adaptive evolution of Toll-like receptor 5 in domesticated mammals

Sarah A Smith<sup>1</sup>, Oliver C Jann<sup>2</sup>, David Haig<sup>1</sup>, George C Russell<sup>3</sup>, Dirk Werling<sup>4</sup>, Elizabeth J Glass<sup>2</sup> and Richard D Emes<sup>1\*</sup>

## Abstract

**Background:** Previous studies have proposed that mammalian toll like receptors (TLRs) have evolved under diversifying selection due to their role in pathogen detection. To determine if this is the case, we examined the extent of adaptive evolution in the TLR5 gene in both individual species and defined clades of the mammalia.

**Results:** In support of previous studies, we find evidence of adaptive evolution of mammalian TLR5. However, we also show that TLR5 genes of domestic livestock have a concentration of single nucleotide polymorphisms suggesting a specific signature of adaptation. Using codon models of evolution we have identified a concentration of rapidly evolving codons within the TLR5 extracellular domain a site of interaction between host and the bacterial surface protein flagellin.

**Conclusions:** The results suggest that interactions between pathogen and host may be driving adaptive change in TLR5 by competition between species. In support of this, we have identified single nucleotide polymorphisms (SNP) in sheep and cattle TLR5 genes that are co-localised and co-incident with the predicted adaptive codons suggesting that adaptation in this region of the TLR5 gene is on-going in domestic species.

**Keywords:** Toll-like receptor, SNP, Adaptive evolution, Positive selection, Sheep, Cattle

## Background

Toll-like receptors (TLRs) are type 1 transmembrane glycoproteins expressed on the cell surface and intracellular compartments of many cell types including epithelial cells and a variety of immune cells such as macrophages and dendritic cells. TLR ligands include pathogen-associated molecular pattern (PAMP) molecules, and TLRs are amongst the first receptors to respond to pathogen presence [1], hence their key role in innate immunity to infection. At least 10 TLRs have been identified in mammals and collectively these recognize a wide repertoire of microbial organisms and pathogens including bacteria, viruses, protozoa and fungi [2]. The TLR protein is comprised of three main regions: an extracellular pattern-recognition receptor domain (ECD), a transmembrane region and an intracellular TIR signalling domain [3]. The signalling domain is highly conserved across the

TLRs. In contrast, the ECD involved in pathogen detection is often variable [2].

TLR5 is known to bind bacterial flagellin [4]. Both flagellin (e.g. of *E.coli*) [4]), and the ECD of TLR5 in primates [5] and other mammals [6] show evidence of adaptive positive selection. This suggests that interspecies competition between host and pathogen is likely to be driving the co-evolution of pathogen and host. In support of this, species-specific single nucleotide variations in the TLR5 gene exist and a single nucleotide polymorphism (SNP) in the ECD of mouse, chicken and human TLR5 is associated with a species-specific response to flagellin [7,8].

The domestication of livestock by selection of desirable traits gave rise to the concept of breeds over 200 years ago [9]. This formation of breeds by selective interbreeding offers a unique opportunity to examine an accelerated process of natural selection. To investigate the evolution of the TLR5 gene in domestic livestock compared to other mammals we used phylogenetic methods to identify species-specific and branch-specific evidence of positive selection. To investigate the potential role of recent

\* Correspondence: richard.emes@nottingham.ac.uk

<sup>1</sup>School of Veterinary Medicine and Science, University of Nottingham, Sutton Bonington LE12 5RD, UK

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

**Table 1 Detection of positive selection across mammalian TLR5**

Analysis	LRT M1a versus M2a (significance)	LRT M7 versus M8 (significance)	$d_N/d_S$ (model)	Sites under positive selection <sup>a</sup>
All Mammals	18.87 (p < 0.001)	28.61 (p < 0.001)	2.85 (M2), 1.56 (M8)	G104* H592* A659*

Results of PAML site specific analysis of positive selection across mammalian TLR5 and branch-site analysis. Positively selected codons detected by both M2a and M8 models are shown \* =  $P_b > 95\%$ .

variation on evolution of the TLR5 gene we also identified known and novel SNPs in the coding region of TLR5 of sheep and cattle breeds.

## Results

### Evidence for adaptive evolution in mammalian TLR5

Positive diversifying selection acting on a gene can be inferred when the ratio of non-synonymous ( $d_N$ ) to synonymous ( $d_S$ ) substitution rates is greater than 1. This ratio  $d_N/d_S$  (also known as omega) provides a method to compare the evolutionary history of codons and lineages [10]. The parameters  $d_N$  and  $d_S$  can be estimated by a number of approaches. We applied the codon models of PAML [11] to infer estimates of parameters under a maximum likelihood framework. The results are summarized in Table 1 and Table 2. Complete results and parameter estimates for all PAML analyses are given as Additional file 1 and Additional file 2.

When comparing rates of codon evolution in TLR5 since the divergence of the mammals (sites analysis, see methods) significant evidence (posterior probability  $P_b > 0.95$ ) was found to suggest that positive selection has been acting on three codons (G104, H592, A659) Table 1. When using the branch-site test (see methods) to compare each lineage independently, significant evidence

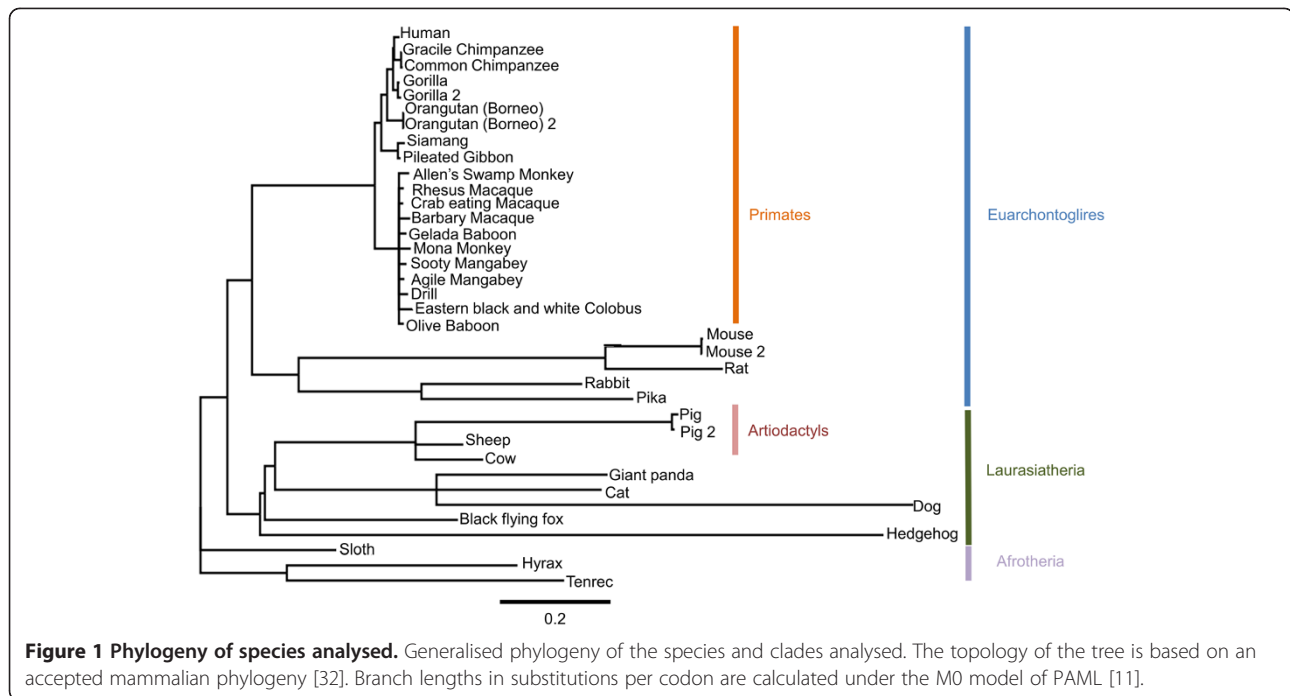
was obtained that the sloth, sheep, cattle and pig lineages were each evolving at an elevated rate compared to other branches of the phylogeny. Three positively-selected codons (I393, Q495, H720,  $P_b > 0.95$ ) were detected in the cattle lineage. One codon (K326,  $P_b > 0.99$ ) was identified in the sheep lineage and one codon (E630,  $P_b > 0.95$ ) was identified in the pig lineage (Table 2). These sites are in addition to those identified as having evolved under positive selection across the mammalian lineage.

Using the multiple branch-sites analysis (see methods), positive selection of TLR5 was detected in the artiodactyls, which contain the domestic ungulate species (cow, sheep and pig), but not in the Laurasiatheria (of which artiodactyls are a component clade) or in the separate euarchontoglires and primate clades (Figure 1). In the artiodactyls, eight codons of TLR5 were detected with significant evidence of positive selection in (L34, S268, I295, I307, I393, Q495, I621, E630, ( $P_b > 0.95$ )). Five codons (L34, S268, I295, I307, I621) were only identified as having evolved under positive selection when the members of the artiodactyla were combined in the multiple branch-sites analysis. Three codons (I393, Q495, E630) were detected previously in ungulate species-specific lineages (Table 2).

**Table 2 Detection of positive selection of mammalian TLR5 using the branch-sites test**

Foreground branch(es)	LRT Foreground $d_N/d_S = 1$ versus foreground $d_N/d_S > 1$	$d_N/d_S$ (branch)	Sites under positive selection <sup>a</sup>
Artiodactyls	35.61 (p < 0.001)	4.76 (foreground)	L34* S268** I295* I307* I393* Q495* I621* E630**
Cow	21.84 (p < 0.001)	140.91 ( foreground )	I393** Q495* H720*
Sheep	18.10 (p < 0.01)	493.28 ( foreground )	K326*
Pig	11.89 (p < 0.05)	29.07 (foreground )	E630**

Results of branch-sites analysis of positive selection within mammalian TLR5. <sup>a</sup> posterior probabilities ( $P_b$ ) of codons existing in  $d_N/d_S > 1$  site class \* =  $P_b > 0.95$  and \*\*  $P_b > 0.99$ .



When mapped onto the predicted tertiary structure of bovine TLR5 (PDB 2a0zA [12] and 1fyv [13]) the location of sites of positive selection detected by all approaches revealed a bias in their distribution. Eleven (L34, G104, S268, I295, I307, K326, I393, Q495, H592, I621, E630) of the thirteen positively-selected codons encode amino acids in the ECD (Figure 2a). When the total number of sites in the ECD compared to the rest of the protein are accounted for, this enrichment remains statistically-significant (Fisher Exact Test:  $P=0.03$ ). Additionally five of these (S268, I295, I307, K326, I393) which exhibited evidence of positive selection in the artiodactyl clade are located within the putative flagellin-binding region of the ECD close to the conserved concave surface-associated with ligand binding [7] (Figure 2b).

### SNP Detection

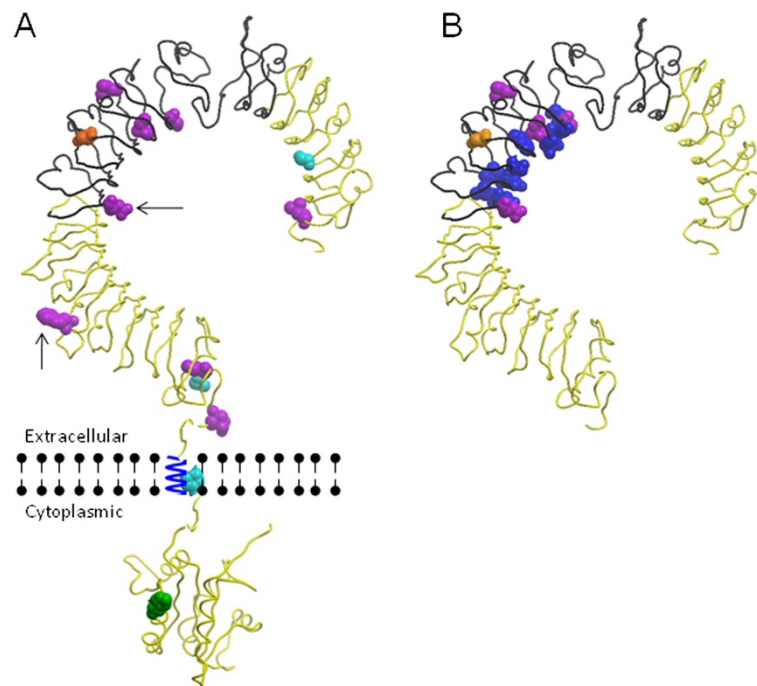
A total of 19 polymorphic sites were detected in cattle and 25 in sheep. No overlap between cattle and sheep was seen (Additional file 3 and Additional file 4). All but the Mongolian cattle breed and the Soay sheep breed showed variability within tested individuals (Figure 3).

Of the 19 polymorphisms detected in cattle, 8 are synonymous substitutions and 11 are non-synonymous substitutions. Twelve of the SNPs detected have been previously reported [14,15]. The remaining 7 SNPs were novel discoveries of this study, and three of these were non-synonymous (L34P, R59K and H262R). All except two polymorphisms are sub-species specific, with *Bos indicus* displaying the highest degree of genetic variability (details in supplementary information). Analysis of

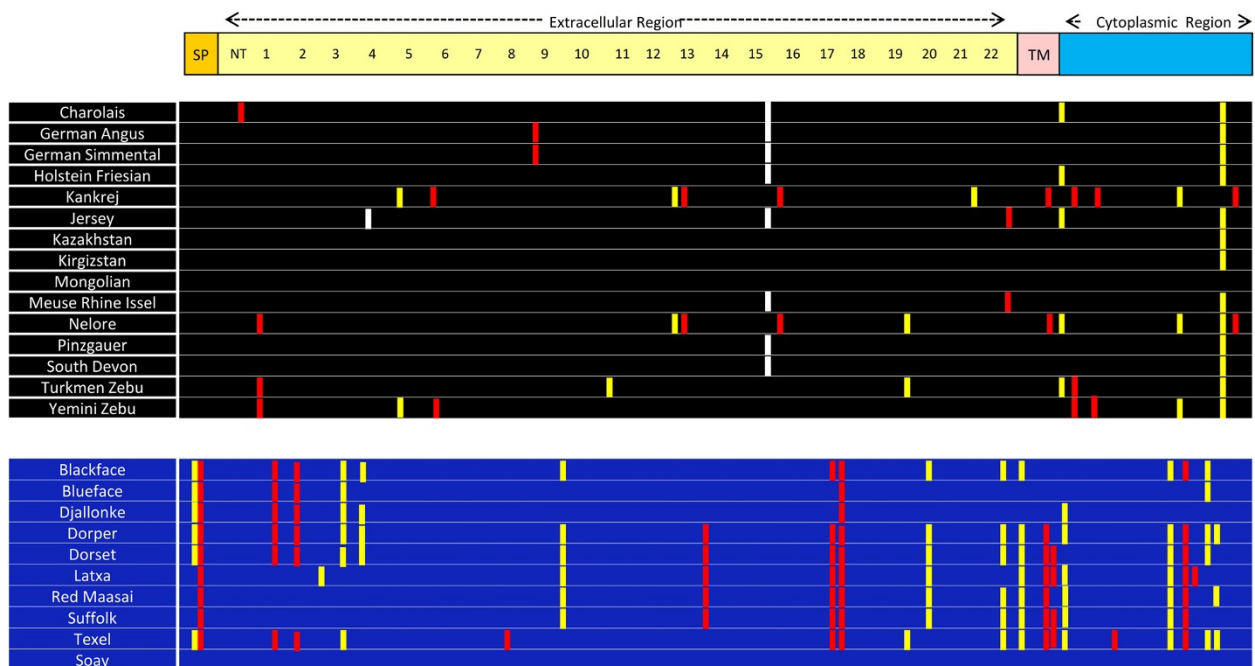
sheep breeds identified 25 novel SNPs, of which 13 were synonymous and 12 were non synonymous substitutions (Figure 3). In addition, two SNPs were identified which are predicted to cause premature stop codons in cattle TLR5 (Figure 3 and Additional file 4). R125\* was detected in the Jersey breed and has recently been reported (15). The other putative stop codon (S431\*) was detected in a selection of *Bos taurus* breeds and is a novel discovery of this study. Pseudogenes are predicted to evolve under neutral selection and as such are not subject to the same evolutionary constraints assumed for protein coding genes. Thus to avoid possible problems of including potential pseudogenes in evolutionary analyses, the stop codon variants were excluded from PAML analysis.

### Co-localisation of SNPs and Positively-Selected Sites

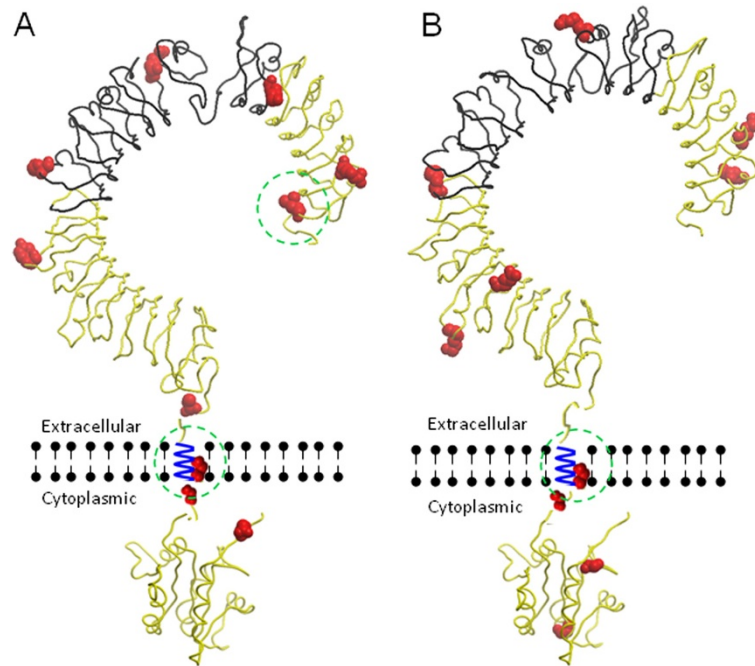
Two non-synonymous SNPs detected in the bovine species co-occurred at codons detected as evolving under positive selection. Codon L34 ( $d_N/d_S > 1$  in artiodactyls) is positioned at the N-terminal region of the extracellular domain of TLR5. Amino acid A659 ( $d_N/d_S > 1$  in all mammals) is located in the transmembrane region and is predicted to extend this domain (Figure 4 and Additional file 3 and Additional file 5). In sheep, two non-synonymous SNPs at position 2 and 3 of codon A659 were also detected (Figure 4 and Additional file 4). Due to the relatively low density of SNPs in close proximity within a single gene, these were not used to estimate population genetic measures such as linkage disequilibrium.



**Figure 2 Localisation of TLR5 sites of positive selection.** **A)** predicted tertiary protein structure of bovine TLR5 and sites of positive selection. **B)** predicted tertiary protein structure of ECD and sites of positive selection and conserved amino acids within putative flagellin binding region. Black backbone depicts putative flagellin binding region; Light blue amino acids = Positive selection acting on all mammals (G104, H592, A659); Purple amino acids = Positive selection acting on domestic ungulates (L34, S268, I295, I307, I393, Q495, I621, E630); Orange amino acid = positive selection in sheep lineage (K326). Green amino acids and black arrows = positive selection in cattle lineage (I393, Q495, H720); Dark blue amino acids = Conserved amino acids predicted to be involved in the detection of flagellin.



**Figure 3 SNPs detected in TLR5.** SNPs detected for each breed aligned to predicted secondary structure of TLR5. SP = Signal peptide; NT = LRR N-terminal; 1 – 22 in yellow = denotes each LRR; CT = LRR C-terminal; TM = transmembrane region; red bars = Non-synonymous polymorphisms; yellow bars = synonymous polymorphisms; White bars = putative stop codons; Black table = cattle breeds; Blue table = sheep breeds.



**Figure 4 Localisation of TLR5 SNPs.** Predicted tertiary protein structure of TLR5 and non-synonymous SNPs. **A)** Predicted tertiary structure of bovine TLR5 and positions of non-synonymous SNPs. 10 amino acid sites correspond to 11 detected SNPs as F679L is affected by two different SNPs in of the same codon. Dashed green circles indicate SNPs occurring at sites of positive selection (L34 and A659); **B)** Predicted tertiary structure of ovine TLR5 and positions of non-synonymous SNPs. Black backbone depicts putative flagellin binding region; Red amino acids = Non-synonymous SNPs; Dashed green circle indicate SNP occurring at sites of positive selection (A659 corresponding to M659 in sheep).

## Discussion

Whilst many studies of the phylogeny and comparative genomics of the TLR gene family exist, this is the first study to characterize the lineage-specific adaptive evolution of the TLR5 gene across clades of the mammalian phylogeny. Three codons (G104, H592, A659) exhibited clear evidence for positive selection across the mammalian phylogeny. This extends previous discoveries that detected positive selection in TLR5 in primate species [5] or in selected mammalian species [6]. Areal *et al.* showed that positive selection is seen in a number of genes of the TLR family of proteins including multiple sites of positive selection in the TLR5 genes when using a subset of mammals [6]. Whilst our findings largely support those of their study Areal *et al.* used a reduced group of animals to identify positive selection. Seventeen species were studied compared to the 37 species of mammal used in this study. The current study is proposed to have increased power to resolve true signature of positive selection. Importantly the Areal study included the chicken TLR5 gene sequence (NM\_001024586.1) in the analysis of TLR5 (see Table S5 of [6]). This would potentially introduce the problem of saturation (see Methods section). Additionally, our study included the novel analysis of groups of lineages to investigate changes in evolutionary constraint in different lineages within the mammals. When  $d_N/d_S$  between

lineages or clades were compared (single branch-sites and multiple branch-sites analysis) the artiodactyl lineage and the individual porcine, ovine and bovine lineages comprising this clade exhibited significant evidence of adaptive evolution ( $d_N/d_S > 1$ ). Within the artiodactyl clade, eight positively-selected sites were detected. Positive selective pressure on genes is symptomatic of functional adaptations acquired during the evolution of species and can promote species functional diversification [16]. This suggests that positive selection observed within the artiodactyl clade is different when compared to that seen in other mammals. We postulate that adaptive evolution observed in TLR5 of domestic livestock is a result of the breeding process. In support of this, it has been previously proposed that ruminant species are undergoing differential selective pressure in the related TLR2 genes [17]. This phenomenon may be directly caused by selective breeding resulting in a rapidly restricted population. The effective population size of all cattle breeds is known to have decreased in recent history and this may reflect initial domestication, breed formation or selection for breed specific production traits (beef or milk) [18]. However, the genetic diversity of cattle as opposed to other species such as dogs is not as low as the effective population size would suggest and high levels of divergent selection associated with immune genes amongst others are detectable



[18]. As genes such as those of the MHC show balancing selection between breeds [18], an alternative proposal is that the breeding process indirectly drove changes in host-pathogen interactions. By increasing animal density, pathogen transmission and load may also have been increased providing the selective drive for rapid adaptation of host and pathogen genes. Eleven of the 13 positively-selected sites are positioned in the ECD of TLR5 – the part of the protein involved in flagellin recognition. It is known that variability within this region influences species-specific ligand recognition [7]. Evidence of increased positive selection within the extracellular domain supports the hypothesis that competition with flagellated bacterial pathogens is driving adaptation in specific host TLR5 ECD and more precisely in the flagellin-binding region. This may be counterintuitive as many livestock will be expected to share similar microbiota and pathogens. However, TLR5-flagellin interaction has recently been mapped at the single amino acid level [7,19,20] suggesting that changes of amino-acids within either, TLR5 or flagellin can alter the species-specific TLR5 response. This in turn may influence the host range and susceptibility of infection. Such amino-acid changes could explain some of the biological differences seen in the response of different species to flagellin. Indeed, chicken TLR5 has been shown to recognise different flagellin-forms compared to human or murine TLR5 [8], and bovine TLR5 has been shown to have a reduced response to recombinant FliC of *S. typhimurium* [21]. These differences are partially based on flagellin-amino-acid differences, and alteration of these amino acids in different flagellins alters their interaction with TLR5 from different species [8,22]. It has been proposed that PAMP ligands engage with the concave surface of their cognate TLR ECDs [23]. Five codons (S268, I295, I307, K326, I393) under positive selection in the artiodactyl clade are in close proximity to this conserved cavity. The non-synonymous SNPs detected within the putative flagellin-binding region are good candidates for genetic variants most likely to impact upon the immune response mediated by TLR5. The function of some of these variants is currently being pursued. However, sites more distal to the flagellin-binding region may also be of importance in TLR ligand recognition and function, for example by altering the shape of the molecule or interfering with signal transduction. For example, in TLR3 a SNP outside of the extracellular region was found to impair receptor signalling [24]. Also a polymorphism in the transmembrane region of human TLR1 is found to regulate the innate immune response indicating that the transmembrane region plays a role in function [25].

Site-specific co-incidence of adaptive codons ( $d_N/d_S > 1$ ) and detection of SNPs suggests that adaptive evolution within these regions is on-going. A good example is codon

A659, which is identified as evolving under positive selection across the mammalian phylogeny and also three non-synonymous SNPs (2 in sheep, 1 in *Bos indicus*) that were found in this codon. SNPs detected at these positions in both sheep and cattle argues against A659 variation being stochastic but is likely to convey an advantage. A previous study of bovine TLR5 SNPs surmised that this site may be of functional significance (SNPdb ID: rs55617251) [26]. This finding was supported using the SIFT program which predicts the functional relevance of SNPs by comparison of conservation at that site [27]. We used the same approach to confirm this result but found that the SNP is now predicted to be a tolerated replacement. This conflict in results suggests caution when solely relying on algorithmic methods to predict putative sites of functional relevance. However, secondary structure analysis on bovine TLR5 revealed this SNP is predicted to alter the alpha helix structure (Additional file 5). This suggests that this amino acid site may indeed be of functional significance.

As expected, we found the internal signaling TIR domain to have evolved predominantly under purifying selection. In cattle a single codon (H720) was predicted as evolving subject to adaptive evolution. This positively-selected site is close to the BB-loop which is predicted to be involved in TLR dimerization and adaptor protein recognition [28].

Of the 25 breeds investigated in our study two breeds were found to have no detectable SNPs in TLR5. The Soay sheep breed was found to be homozygous in the 10 individuals analysed. The Soay breed is a primitive domestic breed which was introduced to Soay island of the St. Kilda archipelago in the Outer Hebrides of northern Scotland. This breed has experienced expansive population growth followed by seasonal periods of population crashes thought to be associated with parasitic helminth disease [29]. Our results indicate that this breed may have lost heterozygosity within TLR5. This raises the possibility that opportunistic bacterial infections may occur due to a poor TLR5 repertoire in this breed. However, sampling of a larger sample set for this breed should be carried out to verify this result.

Four cattle breeds (Kankrej, Nelore, Turkmen Zebu, Yemini Zebu) included in this study were *Bos indicus* subspecies and were found to contain a higher proportion of the total number of non-synonymous SNPs along with all, except two, of the total number of synonymous SNPs (Additional file 6) suggesting that the SNPs are evolutionarily recent events following the divergence of *Bos indicus* and *Bos taurus* (estimated to have diverged as early as 200,000 years ago) [30].

## Conclusions

We have shown that in agreement with other studies, positive selection is acting on mammalian TLR5.

However, our analyses have revealed that the domesticated species in the artiodactyl clade have undergone detectable diversifying selection compared with the rest of the mammals. With the ubiquity of bacterial infection why should the ungulate clade be different to the rest of the mammals? We suggest that artificial selection may have accelerated the evolutionary process resulting in positive selective pressure driving adaptation of the TLR5 gene. The nature of this selection is not known but may be due to selection for cattle resistant to bacteria during domestication or due to an increased bacterial load due to maintaining animals in closer groups and promoting bacterial infection. The concentration of positively-selected sites in close proximity to the conserved cavity of the protein supports the hypothesis that on-going competition is a driving force shaping both the bacterial flagellin and host TLR5 genes.

## Methods

### Identification of positive selection

Genes encoding mammalian TLR5 were obtained from Ensembl and GenBank (Additional file 7). Translated protein sequences were aligned using Muscle [31] and back translated to obtain a codon alignment. Sequences with an in frame stop codon were removed, and 37 sequences remained (Additional file 8). Phylogenetic analysis was based on an accepted mammalian phylogeny [32] from which an unrooted tree of aligned species was created (Additional file 9). Using this tree topology branch lengths were calculated by codeML using the M0 codon model in PAML package version 4 [11]. This tree was used as the fixed tree topology for subsequent analysis. The F3 × 4 codon frequency model calculated using the nucleotide frequencies at the three codon positions was used throughout the analysis. To detect positive selection at individual codons within a gene (sites analysis) two pairs of models were compared using codeML: M1a (neutral model) was compared with M2a (adaptive model) [33,34] and M7 (beta) was compared with M8 (beta plus omega) [35]. Statistically-significant evidence of positive selection was inferred by a likelihood ratio test (LRT) comparing 2 × the log likelihood difference of each set of nested models. These values were compared to the  $\chi^2$  distribution with 2 degrees of freedom. To ensure convergence, the analysis was conducted in triplicate with varying initial  $d_N/d_S$  values. The assumption that  $d_N/d_S$  is equal along all lineages of a phylogeny is likely to be false therefore to identify evidence of episodic selection in specific lineages the branch sites test [36] was also used. This approach allows the  $d_N/d_S$  values at each codon to vary between an *a priori* defined specific branch (foreground) where adaptive evolution is allowed ( $d_N/d_S > 1$ ) and the rest of the tree (background) where adaptive selection is not

allowed  $d_N/d_S = 1$ . Using the branch-sites test, two comparisons were conducted: (1) each branch in turn was analysed as the foreground branch and compared with the rest of the tree and (2) a multiple branch site analysis was conducted, where major taxonomic groups of animals were each in turn grouped as the foreground ( $d_N/d_S > 1$ ) lineage and the remaining lineages are grouped as background ( $d_N/d_S = 1$ ). Clades compared were primates, artiodactyls, Laurasiatheria and Euarchontoglires (Figure 1). This approach of comparison of multiple branches has been shown to improve the power of the branch-sites analysis if the underlying biological assumption to group the branches is sound [37]. This approach directly tests adaptive evolution in multiple branches and avoids the assumptions of the clades analysis model (CmC [38]) which has been recently criticized [39]. In addition the branch-sites test has been criticized as being prone to false positives when strong positive selection exists in the background branches [40]. To account for this reciprocal experiments are conducted where foreground and background branches are reversed and retested. None of these reciprocal experiments resulted in significant results, suggesting that strong positive selection of background branches does not account for the positive selection seen in foreground branches. In all branch-sites tests the Bonferroni correction [41] was employed to control the type I error rate when comparing multiple foreground lineages. In all tests if the LRT identified that a model allowing positive selection significantly improved the likelihood of the data ( $P < 0.05$  using Bonferroni corrected  $\chi^2$  critical values) codons subject to adaptive evolution were then inferred using the Bayes empirical Bayes algorithm [42]. Codons with a posterior probability of  $> 95\%$  ( $P_b > 0.95$ ) of belonging to the class  $d_N/d_S > 1$  are reported as having been subject to positive selection. All positions reported correspond to bovine TLR5 (accession: NP\_001035591.1). There has been recent discussion of the appropriateness of the branch-sites method to detect positive selection particularly by Nozawa et al. [43] who suggested that the branch-sites test is particularly susceptible to false positives when the number of nucleotide substitutions is small. Nozawa et al. proposed that a minimum of 9 substitutions are required to accurately infer positive selection along a foreground branch [43]. Whilst the validity of the approach used to reach these conclusions has also been challenged [44,45], the data analyzed here does fulfill Nozawa's suggestions. Caution should also be exercised when the number of substitutions along a lineage is too great and can lead to errors in predictions. This is due to the problem of saturation of substitutions where the true number of substitutions is masked by multiple nucleotide changes at a single site. The branches of the species of interest are not considered to be a concern in this

regard. For example, the combined branch lengths of the artiodactyl clade (cow, sheep and pig) are less than one substitution per codon. Whilst the results of any inference should be subject to future validation we are confident of the appropriateness of the tests used.

### Protein Structure Prediction

The domain architecture for TLR5 was determined by LRR finder [46], SMART6 [47] and TMHMM [48]. This largely follows published predictions [3]. Secondary Structure predictions were completed using PSI PRED [49]. Tertiary structure predictions were generated using Swiss-Model [50] using known crystallised TLR structures of highest similarity: extracellular domain template 2a0zA and TIR domain template 1fyvA. Positions of sites of positive selection and SNPs affecting particular amino acid positions were visualised using MolSoft ICM Browser [51].

### Sequence Assembly and SNP Detection

Bovine genomic DNA ( $n = 110$ ) representing 15 breeds (Additional file 10), and ovine genomic DNA ( $n = 87$ ) representing 10 breeds (Additional file 11), were extracted from blood samples using standard procedures (Qiagen, UK). Samples were not obtained for the sole purpose of this project, but were either donated (see acknowledgements) or from existing DNA banks at the Roslin Institute. The single exon encoding TLR5 was amplified using primers designed using the bovine sequence (Btau 4.0) (Additional file 12 Table S6 and Additional file 13). Amplification was performed using Taq polymerase (cattle) and KOD proof reading enzyme (sheep). Sequence reads were generated using 6 sequencing primer sets for cattle and 4 for sheep using BigDye Terminator (Applied Biosystems).

Sequences were assembled for each breed using the programs Pregel and Gap4 from the Staden sequence analysis package [52]. Variations for individuals and between breeds were detected using Contig editor. SNPs were included in analysis only when polymorphisms were detected at sites of high confidence (*Phred* score  $>30$ ). Consensus sequences for TLR5 from each breed were exported using Gap4 and polymorphic sites reported by MEGA4 [53]. Nucleotide sequences were translated to amino acid sequences to ascertain the impact of each SNP.

### Additional files

**Additional file 1: Results and parameter estimates of all single branch-sites tests.**

**Additional file 2: Results and parameter estimates of all multiple branch-site analysis.**

**Additional file 3: Positions of detected bovine SNPs.** Results of SNP detection across 15 breeds of cattle. W/T refers to wildtype; Mut refers to mutation form. NS = non-synonymous polymorphism; S = synonymous polymorphism. Amino acid one letter code names amino acid by convention. The type of nucleotide for each SNP and each breed is recorded using IUBMB single-letter code for nucleotide bases and ambiguity codes: R = A/G; Y = C/T; M = A/C; W = A/T; S = C/G; K = G/T.

**Additional file 4: Positions of detected ovine SNPs.** Results of SNP detection across 10 breeds of sheep. W/T refers to wildtype; Mut refers to mutation form. NS = non-synonymous polymorphism; S = synonymous polymorphism. Amino acid one letter code names amino acid by convention. The type of nucleotide for each SNP and each breed is recorded using IUBMB single-letter code for nucleotide bases and ambiguity codes: R = A/G; Y = C/T; M = A/C; W = A/T; S = C/G; K = G/T.

**Additional file 5: Secondary structure sequence predictions (PSI-pred) affecting SNP A659T in cattle TLR5.**

**Additional file 6: Subspecies distribution of SNPs detected in Bovine TLR5.** S: = synonymous SNP; NS: = non-synonymous SNP.

**Additional file 7: Accession numbers of all sequences compared.** Accession numbers of TLR5 coding nucleotide sequences used for PAML analysis.

**Additional file 8: Alignment of all TLR5 genes analysed in Fasta format.**

**Additional file 9: Phylogenetic tree of all TLR5 genes analysed.**

**Additional file 10: Details of bovine DNA samples.** Bovine DNA samples. Sample size and subspecies characterization for each breed is detailed.

**Additional file 11: Details of ovine DNA samples.** Ovine DNA sample set. Sample size for each breed is detailed.

**Additional file 12: Bovine primer sequences Primers used for the sequencing of the coding sequence of bovine TLR5.** Forward primer 1 and Reverse Primer 6 are positioned in the un-translated regions either side of the single exon of TLR5.

**Additional file 13: Ovine primer sequences.** Ovine TLR5 sequencing primers. Forward primer 1 and reverse primer 4 are positioned in the un-translated region either side of the single exon of TLR5.

### Competing interests

The authors declare that they have no competing interests.

### Authors' contributions

SAS designed and conducted the molecular genetic studies, performed the statistical analysis and drafted the manuscript. OCJ designed amplification primers for TLR5 and facilitated amplification and sequencing of cattle samples. OCJ, DH, GCR, DW and EJG participated in the design of the study. DH and EJG conceptualised and coordinated the study. RDE participated in conceptualisation and study design, performed the statistical analysis and helped to draft the manuscript. All authors have read and approved the final manuscript.

### Acknowledgements

Our gratitude is extended to the providers of samples used in this study: Albano Beja-Pereira CIBIO Portugal for the donation of cattle samples; Olivier Hannote, University of Nottingham, and Miika Tapio, ILRI, for sheep samples (Red Maasai, Dorper, Djallonke); and to both Esmeralda Minguijon and Ramon Juste, NIKER Tecnalia, for donating Latxa sheep samples. Our thanks also to ARK Genomics (The Roslin Institute, University of Edinburgh) for sequencing of cattle TLR5 genes. Sarah Smith was funded by a University of Nottingham studentship grant awarded to David Haig. This work was supported by the Biotechnology and Biological Sciences Research Council (BBSRC) (Institute Strategic Programme Grant funding) (EJG) a Genesis Faraday (now KTN Biosciences) facilitated project grant BB/D524040/1 jointly funded by BBSRC, The Scottish Government Rural and Environment Research and Analysis Directorate (RERAD) and Pfizer Animal Health (OCJ, EJG, DW, DH).



#### Author details

<sup>1</sup>School of Veterinary Medicine and Science, University of Nottingham, Sutton Bonington LE12 5RD, UK. <sup>2</sup>The Roslin Institute and Royal Dick School of Veterinary Studies, The University of Edinburgh, Easter Bush, Midlothian EH25 9RG, UK. <sup>3</sup>Royal Veterinary College, University of London, Hawkshead Lane, Hatfield, Herts AL9 7TA, UK. <sup>4</sup>Moredun Research Institute, Pentlands Science Park, Bush Loan, Penicuik, Midlothian EH26 0PZ, UK.

Received: 21 March 2012 Accepted: 6 July 2012

Published: 24 July 2012

#### References

1. Gay NJ, Gangloff M, Weber AN: Toll-like receptors as molecular switches. *Nat Rev Immunol* 2006, **6**:693–698.
2. Werling D, Jann OC, Offord V, Glass EJ, Coffey TJ: Variation matters: TLR structure and species-specific pathogen recognition. *Trends Immunol* 2009, **30**:124–130.
3. Matsushima N, Tanaka T, Enkhbayar P, Mikami T, Taga M, Yamada K, Kuroki Y: Comparative sequence analysis of leucine-rich repeats (LRRs) within vertebrate toll-like receptors. *BMC Genomics* 2007, **8**:124.
4. Reid SD, Selander RK, Whittam TS: Sequence diversity of flagellin (fliC) alleles in pathogenic *Escherichia coli*. *J Bacteriol* 1999, **181**:153–160.
5. Wlasiuk G, Khan S, Switzer WM, Nachman MW: A history of recurrent positive selection at the toll-like receptor 5 in primates. *Mol Biol Evol* 2009, **26**:937–949.
6. Areal H, Abrantes J, Esteves PJ: Signatures of positive selection in Toll-like receptor (TLR) genes in mammals. *BMC Evol Biol* 2011, **11**:368.
7. Andersen-Nissen E, Smith KD, Bonneau R, Strong RK, Aderem A: A conserved surface on Toll-like receptor 5 recognizes bacterial flagellin. *J Exp Med* 2007, **204**:393–403.
8. Keesstra AM, de Zoete MR, van Aubel RA, van Putten JP: Functional characterization of chicken TLR5 reveals species-specific recognition of flagellin. *Mol Immunol* 2008, **45**:1298–1307.
9. Taberlet P, Valentini A, Rezaei HR, Naderi S, Pompanon F, Negrini R, Ajmone-Marsan P: Are cattle, sheep, and goats endangered species? *Mol Ecol* 2008, **17**:275–284.
10. Yang Z: *Computational Molecular Evolution*. 1st edition. Oxford: Oxford University Press; 2006.
11. Yang Z: PAML 4: phylogenetic analysis by maximum likelihood. *Mol Biol Evol* 2007, **24**:1586–1591.
12. Bell JK, Botos I, Hall PR, Askins J, Shiloach J, Segal DM, Davies DR: The molecular structure of the Toll-like receptor 3 ligand-binding domain. *Proc Natl Acad Sci U S A* 2005, **102**:10976–10980.
13. Xu Y, Tao X, Shen B, Horng T, Medzhitov R, Manley JL, Tong L: Structural basis for signal transduction by the Toll/interleukin-1 receptor domains. *Nature* 2000, **408**:111–115.
14. Seabury CM, Cargill EJ, Womack JE: Sequence variability and protein domain architectures for bovine Toll-like receptors 1, 5, and 10. *Genomics* 2007, **90**:502–515.
15. Fisher CA, Bhattacharai EK, Osterstock JB, Dowd SE, Seabury PM, Vikram M, Whitlock RH, Schukken YH, Schnabel RD, Taylor JF, et al: Evolution of the bovine TLR gene family and member associations with *Mycobacterium avium* subspecies paratuberculosis infection. *PLoS One* 2011, **6**:e27744.
16. Vamathevan JJ, Hasan S, Emes RD, Amrine-Madsen H, Rajagopalan D, Topp SD, Kumar V, Word M, Simmons MD, Foord SM, et al: The role of positive selection in determining the molecular cause of species differences in disease. *BMC Evol Biol* 2008, **8**:273.
17. Jann OC, Werling D, Chang JS, Haig D, Glass EJ: Molecular evolution of bovine Toll-like receptor 2 suggests substitutions of functional relevance. *BMC Evol Biol* 2008, **8**:288.
18. Gibbs RA, Taylor JF, Van Tassel CP, Barendse W, Eversole KA, Gill CA, Green RD, Hamernik DL, Kappes SM, Lien S, et al: Genome-wide survey of SNP variation uncovers the genetic structure of cattle breeds. *Science* 2009, **324**:528–532.
19. Lu J, Sun PD: The structure of the TLR5-flagellin complex: a new mode of pathogen detection, conserved receptor dimerization for signaling. *Sci Signal* 2012, **5**:pe11.
20. Yoon SI, Kurnasov O, Natarajan V, Hong M, Gudkov AV, Osterman AL, Wilson IA: Structural basis of TLR5-flagellin recognition and signaling. *Science* 2012, **335**:859–864.
21. Metcalfe HJ, Best A, Kanellos T, La Ragione RM, Werling D: Flagellin expression enhances *Salmonella* accumulation in TLR5-positive macrophages. *Dev Comp Immunol* 2010, **34**:797–804.
22. Murthy KG, Deb A, Goonesekera S, Szabo C, Salzman AL: Identification of conserved domains in *Salmonella muenchen* flagellin that are essential for its ability to activate TLR5 and to induce an inflammatory response in vitro. *J Biol Chem* 2004, **279**:5667–5675.
23. Brodsky I, Medzhitov R: Two modes of ligand recognition by TLRs. *Cell* 2007, **130**:979–981.
24. Ranjith-Kumar CT, Miller W, Sun J, Xiong J, Santos J, Yarbrough I, Lamb RJ, Mills J, Duffy KE, Hoose S, et al: Effects of single nucleotide polymorphisms on Toll-like receptor 3 activity and expression in cultured cells. *J Biol Chem* 2007, **282**:17696–17705.
25. Hawn TR, Misch EA, Dunstan SJ, Thwaites GE, Lan NT, Quy HT, Chau TT, Rodrigues S, Nachman A, Janer M, et al: A common human TLR1 polymorphism regulates the innate immune response to lipopeptides. *Eur J Immunol* 2007, **37**:2280–2289.
26. Seabury CM, Seabury PM, Decker JE, Schnabel RD, Taylor JF, Womack JE: Diversity and evolution of 11 innate immune genes in *Bos taurus taurus* and *Bos taurus indicus* cattle. *Proc Natl Acad Sci U S A* 2010, **107**:151–156.
27. Ng PC, Henikoff S: SIFT: Predicting amino acid changes that affect protein function. *Nucleic Acids Res* 2003, **31**:3812–3814.
28. Nyman T, Stenmark P, Flodin S, Johansson I, Hammarstrom M, Nordlund P: The crystal structure of the human toll-like receptor 10 cytoplasmic domain reveals a putative signaling dimer. *J Biol Chem* 2008, **283**:11861–11865.
29. Bancroft DR, Pemberton JM, King P: Extensive protein and microsatellite variability in an isolated, cyclic ungulate population. *Heredity (Edinb)* 1995, **74**(Pt 3):326–336.
30. Loftus RT, MacHugh DE, Bradley DG, Sharp PM, Cunningham P: Evidence for two independent domestications of cattle. *Proc Natl Acad Sci U S A* 1994, **91**:2757–2761.
31. Edgar RC: MUSCLE: a multiple sequence alignment method with reduced time and space complexity. *BMC Bioinforma* 2004, **5**:113.
32. Murphy WJ, Eizirik E, Johnson WE, Zhang YP, Ryder OA, O'Brien SJ: Molecular phylogenetics and the origins of placental mammals. *Nature* 2001, **409**:614–618.
33. Nielsen R, Yang Z: Likelihood models for detecting positively selected amino acid sites and applications to the HIV-1 envelope gene. *Genetics* 1998, **148**:929–936.
34. Wong WS, Yang Z, Goldman N, Nielsen R: Accuracy and power of statistical methods for detecting adaptive evolution in protein coding sequences and for identifying positively selected sites. *Genetics* 2004, **168**:1041–1051.
35. Yang Z, Nielsen R, Goldman N, Pedersen AM: Codon-substitution models for heterogeneous selection pressure at amino acid sites. *Genetics* 2000, **155**:431–449.
36. Zhang J, Nielsen R, Yang Z: Evaluation of an improved branch-site likelihood method for detecting positive selection at the molecular level. *Mol Biol Evol* 2005, **22**:2472–2479.
37. Ramm SA, Oliver PL, Ponting CP, Stockley P, Emes RD: Sexual selection and the adaptive evolution of mammalian ejaculate proteins. *Mol Biol Evol* 2008, **25**:207–219.
38. Bielawski JP, Yang Z: A maximum likelihood method for detecting functional divergence at individual codon sites, with application to gene family evolution. *J Mol Evol* 2004, **59**:121–132.
39. Weadick CJ, Chang BS: An Improved Likelihood Ratio Test for Detecting Site-Specific Functional Divergence among Clades of Protein-Coding Genes. *Mol Biol Evol* 2012, **29**:1297–1300.
40. Suzuki Y: False-positive results obtained from the branch-site test of positive selection. *Genes Genet Syst* 2008, **83**:331–338.
41. Gordi T, Khamis H: Simple solution to a common statistical problem: interpreting multiple tests. *Clin Ther* 2004, **26**:780–786.
42. Yang Z, Wong WS, Nielsen R: Bayes empirical bayes inference of amino acid sites under positive selection. *Mol Biol Evol* 2005, **22**:1107–1118.
43. Nozawa M, Suzuki Y, Nei M: Reliabilities of identifying positive selection by the branch-site and the site-prediction methods. *Proc Natl Acad Sci U S A* 2009, **106**:6700–6705.
44. Yang Z, dos Reis M: Statistical properties of the branch-site test of positive selection. *Mol Biol Evol* 2011, **28**:1217–1228.

45. Yang Z, Nielsen R, Goldman N: **In defense of statistical methods for detecting positive selection.** *Proc Natl Acad Sci U S A* 2009, **106**:E95. author reply E96.
46. Offord V, Coffey TJ, Werling D: **LRRfinder: a web application for the identification of leucine-rich repeats and an integrative Toll-like receptor database.** *Dev Comp Immunol* 2010, **34**:1035–1041.
47. Letunic I, Doerks T, Bork P: **SMART 6: recent updates and new developments.** *Nucleic Acids Res* 2009, **37**:D229–D232.
48. Krogh A, Larsson B, von Heijne G, Sonnhammer EL: **Predicting transmembrane protein topology with a hidden Markov model: application to complete genomes.** *J Mol Biol* 2001, **305**:567–580.
49. McGuffin LJ, Bryson K, Jones DT: **The PSIPRED protein structure prediction server.** *Bioinformatics* 2000, **16**:404–405.
50. Arnold K, Bordoli L, Kopp J, Schwede T: **The SWISS-MODEL workspace: a web-based environment for protein structure homology modelling.** *Bioinformatics* 2006, **22**:195–201.
51. Abagyan R, Lee WH, Raush E, Budagyan L, Totrov M, Sundstrom M, Marsden BD: **Disseminating structural genomics data to the public: from a data dump to an animated story.** *Trends Biochem Sci* 2006, **31**:76–78.
52. Staden R: **The Staden sequence analysis package.** *Mol Biotechnol* 1996, **5**:233–241.
53. Tamura K, Dudley J, Nei M, Kumar S: **MEGA4: Molecular Evolutionary Genetics Analysis (MEGA) software version 4.0.** *Mol Biol Evol* 2007, **24**:1596–1599.

doi:10.1186/1471-2148-12-122

**Cite this article as:** Smith *et al.*: Adaptive evolution of Toll-like receptor 5 in domesticated mammals. *BMC Evolutionary Biology* 2012 **12**:122.

**Submit your next manuscript to BioMed Central and take full advantage of:**

- Convenient online submission
- Thorough peer review
- No space constraints or color figure charges
- Immediate publication on acceptance
- Inclusion in PubMed, CAS, Scopus and Google Scholar
- Research which is freely available for redistribution

Submit your manuscript at  
[www.biomedcentral.com/submit](http://www.biomedcentral.com/submit)

