Novel ovine polymorphisms and adaptive evolution in mammalian TLR2 suggest existence of multiple pathogen binding regions

S.A. Smith ¹, D. Haig, R.D. Emes *

School of Veterinary Medicine and Science, University of Nottingham, Sutton Bonington, LE12 5RD, UK

Abstract

Toll-like receptors initiate inflammatory responses following the recognition of a wide repertoire of pathogens including bacteria, fungi, protozoa and viruses. They are composed of an extracellular leucine-rich repeat domain responsible for detecting pathogen-associated molecular patterns, a membrane spanning region and an intracellular Toll/Interleukin 1 receptor domain which invokes signal transduction. Toll-like receptor 2 is the most diverse of these receptors as it recognises infectious agents from a range of pathogenic groups. Over 1400 breeds of sheep exist worldwide that inhabit a diverse range of environments, which leads to the potential contact with a wide variety of pathogens likely detected by Toll-like receptor 2. In this study, we evaluated the extent of both long term evolutionary changes, across the mammalian phylogeny of the TLR2 gene, and recent divergence of this same gene in sheep breeds. Evolutionary analyses identified positive selective pressure across the mammalian phylogeny, and differential selection pressure within the artiodactyl and primate lineage. Finally, we identified localised positively-selected sites within two regions of the extracellular domain which suggest that multiple binding regions in TLR2 may be involved in pathogen detection. These results are consistent with the hypothesis that competition between host and pathogen is driving adaptation of Toll-like receptor 2 genes. © 2014 The Authors. Published by Elsevier B.V. This is an open access article under the CC BY license (http://creativecommons.org/licenses/by/3.0/).

1. Introduction

Competition between host and pathogen is a key process in the evolution of species. This was proposed as the Red Queen Hypothesis which states that pathogen interaction with a host results in both the host and the pathogen evolving new strategies to compete with each other (Van Valen, 1973). The competition pressure exerted from both sides results in an evolutionary arms race, where both species evolve in order to equip themselves with mechanisms in an attempt to out-compete the other. This tussle results in evolutionary pressure exerted on competing groups of co-evolving genes (Woolhouse et al., 2002). It is these genes that ultimately develop adaptations and counter adaptations through mutational events which are then fixed by natural selection.

The mammalian immune system comprises two parts, the innate and the acquired immune response. Innate immunity is an ancient form of immune defence and provides an immediate response to infection (Murphy et al., 2008). Toll-like receptors (TLRs) are a major group of proteins that form part of innate immunity by recognising molecular components of infectious agents, known as pathogen associated molecular patterns (PAMPs) (Jann et al., 2008). Within pathogens, these molecular motifs are components that are integral to the survival of the organism, such as components of the cell wall and motility structures, thus resulting in them being largely conserved within pathogen genomes (Kimbrall and Beutler, 2001; Macnab, 1992).

To date, 13 TLRs have been described, although it has been shown that not all species contain this full complement of receptors. For example, TLRs 11, 12 and 13 are present in mice but not human whereas TLR10 is present in human and not mice (Hans and Hans, 2011). TLR2 is widely expressed across species and recognises the greatest number of PAMPs, detecting components from bacteria, viruses and fungi (Barbalat et al., 2009; Gay and Gangloff, 2007). TLR2 is predicted to have an expanded range of binding ligand spectrum through its unique ability to form heterodimers with TLR1 or TLR6 (Farhat et al., 2008). Single nucleotide polymorphisms (SNPs) have been identified in the coding sequence of TLR2 of different species (Bhide et al., 2009; Jann et al., 2008; Merx et al., 2007) with certain SNPs demonstrated to alter function of TLR2 of different species (Bhide et al., 2009; Jann et al., 2008; Merx et al., 2007) with certain SNPs demonstrated to alter function.

Abbreviations: ECD, extracellular domain; LRR, leucine-rich repeat; LRRNT, N-terminal leucine-rich repeat; PAMP, pathogen associated molecular patterns; PBMC, peripheral blood mononuclear cells; PCR, polymerase chain reaction; SNPs, single nucleotide polymorphisms; TIR, Toll/Interleukin 1 receptor domain; TLR, Toll-like receptors.

* Corresponding author.

E-mail addresses: sarah.smith@sheffield.ac.uk (S.A. Smith), david.haig@nottingham.ac.uk (D. Haig), richard.emes@nottingham.ac.uk (R.D. Emes).

¹ Present address: University of Sheffield, Medical School, Beech Hill Road, S10 2RX, UK.

http://dx.doi.org/10.1016/j.gene.2014.02.032
0378-1119/© 2014 The Authors. Published by Elsevier B.V. This is an open access article under the CC BY license (http://creativecommons.org/licenses/by/3.0/).
will interact with TLR2 (Entrican et al., 2010; Mun et al., 2003). To investigate potential variation of TLR2 at different evolutionary scales, novel SNPs in sheep TLR2 genes were identified and signatures of positive selection acting across the TLR2 gene were inferred. The distribution of sheep SNPs and the evolutionary imprint across the mammalian phylogeny have identified mutations with the potential to impact upon the immune response to pathogenic invasion.

2. Materials and methods

2.1. DNA samples and primer design

Genomic DNA was extracted from peripheral blood mononuclear cells (PBMC) of ten sheep breeds (n = 69). Seven breeds were of varied European origin and three of African origin (provided in the supplementary information). Published primers were used to amplify the coding sequence of TLR2 (Jann et al., 2008) (provided in the supplementary information).

2.2. DNA amplification and sequencing

PCR amplification was carried out using the proofreading KOD Hot Start DNA Polymerase (Novagen-TOYOBO, Darmstadt). PCR was performed with 160 ng of template total DNA for each 150 μl reaction with 0.2 mM of each dNTP, 1.5 mM MgSO4, 0.6 μM of each primer, 1× KOD buffer and 0.02 U/μl KOD polymerase. Each reaction was subjected to 30 cycles as per manufacturer’s instructions using an annealing temperature of 60 °C. PCR products were gel purified and DNA was extracted using NucleoSpin Extract II (Macherey-Nagel GmbH & Co.). Samples were then pooled in pairs, by breed, and sequenced using overlapping forward and reverse primers designed with Primer3 (Rozen and Skaletsky, 2000) (provided in the supplementary information).

2.3. Sequence analysis and SNP detection

Sequenced fragments were processed and assembled for each breed using the programs Pregap and Gap4 from the Staden sequence analysis package (Staden, 1996). Assembled sequences were trimmed to remove non-coding sequence. Mutations and SNPs were detected using the Contig Editor function and associated chromatograms. Bi-allelic peak signatures were identified and signatures of positive selection acting across the TLR2 gene were inferred. The distribution of sheep SNPs and the evolutionary imprint across the mammalian phylogeny have identified mutations with the potential to impact upon the immune response to pathogenic invasion.

3. Results

3.1. SNP detection in ovine TLR2

A total of 18 polymorphic sites were detected in sheep TLR2 (Table 1). One SNP, T2008C (F670L) has previously been reported (Bhide et al., 2009) and the remaining are novel. Eight of the ten sheep breeds analysed (Table 1).

3.2. Adaptive evolution in mammalian TLR2

Using a maximum likelihood approach, evidence of positive selection was detected in TLR2 since the divergence of mammals. Statistically significant evidence (posterior probability Pb > 0.95) of positive selection was found for five codons (E88, Q185, R297, L355, Q454) across the mammalian phylogeny. When using the combined branch-site test (see Methods) to compare each clade independently, significant evidence (Pb > 0.95) of positive selection was obtained for the artiodactyl clade (H99, T172), which contains the sheep lineage, but not in the other clades.
Laurasiatheria clade (which includes the artiodactyls) (Fig. 2). Positive selection was not found to be acting on Euarchontoglires but within the component primate clade two sites were identified as under selection (D38, Y276). These results are summarised in Table 2. 

All nine positively-selected sites were positioned within the extracellular domain of the ovine TLR2 (Table 2). Positively-selected sites were not detected in the transmembrane region or cytoplasmic TIR domain of TLR2 (Table 2, Fig. 3). Three sites (Y276, R297, L355) were located around the central region of the ECD involved in bacterial lipopeptide ligand detection and protein–protein heterodimerization interactions with TLR1 and TLR6 (Fig. 3). Five sites (D38, E88, H99, T172, Q185) formed another cluster towards the N terminal region of the protein between LRRNT and LRR6 with one site (Q454) identified within the C-terminal region of the extracellular domain (Fig. 3).

4. Discussion

4.1. Evidence of adaptive evolution in TLR2 mammalian phylogeny

TLR2 has previously been shown to be subject to adaptive evolution (Areal et al., 2011; Jann et al., 2008; Takaki et al., 2012; Wlasiuk and Nachman, 2010). This current study has taken a maximum likelihood approach to further explore these initial findings by comparison of evolutionary rates between specific clades of an extended mammalian phylogeny.

Specifically, five codons (E88, Q185, R297, L355, Q454) have been discovered as undergoing positive selective pressure since the divergence of the mammals (Table 2). Two specific lineages, the primates and the artiodactyls, display differential patterns of adaptive evolution compared to the rest of the mammals, at codons D38, Y276 and H99, T172, respectively. Whilst our findings support previous studies, in that positive selection is acting across the phylogeny, specific sites under selection differ.

Both Wlasiuk and Nachman (2010) and Takaki et al. (2012) concentrated their investigation on the evolution of TLR2 in the primate phylogeny. Takaki and colleagues focused on site specific divergence and adaptation in TLR2 between two closely related species of macaques. Our analyses involved lineage specific adaptation of the primate phylogeny, using 11 representative species, and reports on the site-specific adaptive-divergence of primates (D38, Y276) compared to the rest of the mammalian phylogeny. The different approaches taken between phylogenies compared to specific close range species, are likely to account for different positively selected sites within TLR2 being identified across these two studies.

The Wlasiuk and Nachman (2010) study also used 11 representative species and PAML4 to investigate selection of primate TLR2 gene. Their results initially revealed no specific sites under selection using PAML4 evolutionary analysis tool. The contrast in adaptive sites to our results is likely due to the inclusion of our extended mammalian phylogeny which provides the benefit of allowing evolutionary pressure within the primate lineage to be compared to other mammalian lineages. Wlasiuk and Nachman (2010) did however identify specific sites under selection when relaxing the inclusion criteria, posterior probability (Pb) >90% in PAML4, compared to our analysis which identified sites under selection with Pb >95%. Additionally, they employed the use of the molecular analysis tool, Datamonkey webserver, and also reported on sites under selection when at least two models reported on the same sites. The variation in approaches used can explain specific differences in results between our studies and also results generated in the Areal et al. (2011) study.

The study by Areal et al. (2011), investigated positive selection across the mammalian phylogeny of TLR2, and adopted a similar approach to Wlasiuk and Nachman (2010), in that both PAML and Datamonkey webserver were employed to independently analyse their data sets. PAML4 analysis identified two specific sites under selection across the mammalian phylogeny (296 & 453) (Areal et al., 2011). Since the species reference sequence is not detailed we have made the assumption that this corresponds to (R297 & Q454) in our analyses. The additional sites we identified as under positive selection across the mammalian phylogeny (E88, Q185, L355) are again likely to be due to increased power obtained by the inclusion of an additional 21 species compared to Areal et al. Increasing the number of quality sequences increases the power to detect accurate signatures of positive selection. Use of Datamonkey webserver revealed additional sites of interest. The approach of this package differs to PAML4. For example, the use of the REL tool makes the most assumptions of all the Datamonkey

### Table 1

SNPs detected in ovine TLR2.

| Amino acid position | 615 | 647 | 943 | 1162 | 1401 | 1688 | 1731 | 1737 | 1749 | 1949 | 2008 | 2037 | 2055 | 2089 | 2168 | 2232 | 2235 | 2238 |
|---------------------|-----|-----|-----|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|
| Codon wildtype      | CTG | GAT | CGG | GTC  | AAC  | CGC  | TCT  | CGG  | TCT  | CGG  | TCT  | CGG  | TCT  | CGG  | TCT  | CGG  | TCT  | CGG  | TCT  |
| Codon mutant        | CTG | GAT | CGG | GTC  | AAC  | CGC  | TCT  | CGG  | TCT  | CGG  | TCT  | CGG  | TCT  | CGG  | TCT  | CGG  | TCT  | CGG  | TCT  |
| Amino acid wildtype | L   | D   | R   | V    | N    | R    | D    | R    | S    | R    | F    | F    | I    | R    | R    | A    | V    | P    |
| Amino acid mutant   | L   | A   | W   | F    | N    | H    | D    | R    | R    | Q    | L    | F    | I    | C    | A    | V    | P    |
| SNP Type            | S   | NS  | NS  | NS   | S    | S    | S    | NS   | NS   | NS   | S    | S    | NS   | S    | S    | NS   | S    | S    |

NS = non-synonymous polymorphism; S = synonymous polymorphism. Amino acid one letter code names amino acid by convention. The type of nucleotide exchange for each SNP and each breed is recorded in blue 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. Nucleotides in green show breeds which are homozygote for the polymorphic type.
webserver models, and is susceptible to high rates of false positives in extreme cases (Pond and Frost, 2005). Jann et al. (2008) analyses used models in PAML4 to determine where variation exists across the mammalian phylogeny but did not determine site-specific adaptation. To allow direct comparison of results, we employed a stringent approach comparable to our previous study of TLR5 genes (Smith et al., 2012).

4.2. Comparison of adaptive evolution between mammalian TLR2 and TLR5 genes

We have previously examined evolution of mammalian TLR5 genes (Smith et al., 2012). When compared with the results of this study, a contrasting pattern of selection emerges. In particular, the artiodactyl
Fig. 2. Mammalian phylogeny detailing species analysed in evolutionary analysis for TLR2. Phylogenetic tree was created based on an accepted phylogeny (Murphy et al., 2001). Mammals from Africa or an African origin form the clade Afrotheria. Euarchontoglires form the clade of mammals that include rodents, lagomorphs and primates. Laurasiatheria forms a clade of placental mammals including carnivorans, the majority of hoofed mammals, bats, whales, shrews and hedgehogs. Artiodactyls are a clade including even-toed ungulates (information provided by the Tree of Life Project).
Table 2
Positive selection across mammalian TLR2.

| Sites analysis | LRT statistic | \( \frac{\text{dN}}{\text{dS}} \text{ (model/branch)} \) | Sites under positive selection | Domain architecture position |
|----------------|---------------|---------------------------------|-------------------------------|-----------------------------|
| All mammals (sites analysis) | 45.77 (p < 0.001) | 55.07 (p < 0.001) | 3.03 (M2) | E88* |
| | | | 1.64 (M8) | Q185* |
| | | | | R297* |
| | | | | L355** |
| | | | | Q454** |
| Branch-sites test foreground | Foreground \( \frac{\text{dN}}{\text{dS}} \) = 1 versus foreground | 2.63 (foreground) | H99* | LRR2 |
| | | | | T172 |
| Artiodactyls | 15.76 (p < 0.001) | | | LRR5 |
| | | | | LRRNT |
| Primates | 14.08 (p < 0.001) | 3.69 (foreground) | D38* | LRR9 |
| | | | | Y276 |

All mammals = positive selected detected across the mammalian phylogeny. Artiodactyls = positive selection detected in artiodactyl clade compared to the rest of the mammalian phylogeny. Primates = positive selected detected in primate lineage compared to the rest of the mammalian phylogeny. \( \frac{\text{dN}}{\text{dS}} \) = the rate ratio of non-synonymous to synonymous substitutions. The position of positively selected sites is detailed based on domain architecture predictions (see Methods). LRR = leucine rich repeat; LRRNT = leucine rich repeat N-terminus.

* Poster probabilities (Pb) of codons existing in \( \frac{\text{dN}}{\text{dS}} \) sites class.
** Pb > 0.99.

Fig. 3. Sites of positive selection and functional relevance in TLR2. (A) Schematic illustration of ECD domain architecture. Black bars indicate the ECD which is divided into three distinct regions; N terminal region, Central Region and C terminal region (Jin et al., 2007). Positive selected sites under selection are mapped to leucine-rich repeats (LRR). Regions of the protein known to be involved in function are indicated by dashed lines. 1 = Region required to respond to peptidoglycan (PGN) (Mitsuzawa et al., 2001). 2 = Region implicated in response to PGN, Lipoteichoic Acid and Zymosin (Meng et al., 2003). 3 = Region involved in synthetic ligand binding; diacylated lipopeptide ligand, Pam2CSK4 and triacylated lipopeptide ligand, Pam3CSK4, and heterodimerization (Jin et al., 2007; Kang et al., 2009). SP = signal peptide; NT = LRRNT; 1–20 = LRRs from 1–20; CT = LRRCT. (B) Sites of positive selection mapped to the crystal structure of TLR2 (Jin et al., 2007; Kang et al., 2009). (C) TLR2–TLR1/TLR6 ligand binding and dimerization residues (illustrated in grey) mapped to crystal structure of TLR2. Models generated using PDB reference number: 3a7b Chain: A. Sites in blue = sites under selection across the mammalian phylogeny. Sites in purple = artiodactyl clade and sites in orange = primate clade.
lineage for TLR5 has evidence of extensive positive selection compared with TLR2. Both receptor phylogenies contain domestic livestock species. However, TLR2 does not exhibit adaptive evolution of genus specific (porcine, ovine and bovine) lineages as was seen for TLR5. This difference may be a reflection of receptor function. Both TLR2 and TLR5 detect components of bacteria but whilst TLR5 is currently thought to act only as a homodimer and specifically detect flagellin, TLR2 shows more versatility. TLR2 is known to form dimers with TLR1 and TLR6 and recognises the greatest number of known ligands of all TLRs. This range of TLR2 increases possible combinations of interacting proteins and hence interacting ligands. Our results indicate that porcine, ovine and bovine lineages are not undergoing species specific adaptive change within the TLR2 gene but possibly utilise the strategy maintaining diversity by use of multiple TLR dimerization and ligand (and hence pathogenic species) recognition.

Most proteins have integral regions that are subject to functional constraint and non-integral regions where a relaxation of constraint is observed (Zhang et al., 2011). It is plausible that there is a wider range of essential residues within the TLR2 gene that are involved in dimerization and ligand interactions, compared to TLR5. Should this be the case, then it is expected that there would be a greater number of residues in the ECD of TLR5 that are free to undergo adaptive change, in order to adapt or enhance function, without impacting the current ligand specificity of the protein. The noticeable difference in dimerization and ligand interaction may account for the difference in selection pressure observed between the artiodactyl lineage and these two receptors.

4.3. Pattern of positively selected sites across ECD of TLR2

In the ECD, positively-selected sites were mapped to the protein structure of TLR2 (Fig 3). Sites identified may highlight possible regions of the ECD that are likely to be of biological importance. It is well documented that positively-selected sites within proteins can correlate to ligand binding regions (Ballingall et al., 2010; Furlong and Yang, 2008; Smith et al., 2012). Indeed, our recent publication on mammalian TLR5 identified a cluster of positively-selected sites within the putative ligand binding region of the ECD (Smith et al., 2012) which has been subsequently confirmed, by X-ray crystallography resolution, as a region involved in binding flagellin (Yoon et al., 2012). The first cluster of positively selected sites in TLR2 is positioned within the known region of ligand binding and dimerization with TLR1 and TLR6 (Jin et al., 2007; Kang et al., 2009). The second cluster is positioned towards the N-terminal region. Due to the extensive ligand binding repertoire of TLR2, it has been proposed that multiple binding sites may exist across the ECD (Mitsuzawa et al., 2001). TLR2 LRR-deletion studies implicated the N-terminal region, from LRRNT and up to LRR7, as integral for ligand detection for a variety of ligands including lipoteichoic acid, peptidoglycan and zymosin (Meng et al., 2003; Mitsuzawa et al., 2001). Furthermore, TLR2 LRR-deleted constructs, LRRNT-LRR7, induced an NF-κB response to the synthetic ligands Pam3CSK4 and Pam2CSK4, indicating that this region is not required to invoke a functional response to these ligands (Meng et al., 2003). To be functional as a signalling molecule TLR2 needs to be relatively conserved, but to function as a protein interacting with multiple ligands there is a need for variability within the ligand binding domain. We propose that this is achieved in TLR2 as selective pressures exerted on different parts of the ECD are driving adaptation through the competition between the host and the pathogens detected by different binding regions. As such, functionally relevant mutations in one ligand binding region may alter ligand specificity, but not necessarily impact upon pathogen recognition of the other region.

4.4. Distribution and potential impact of SNPs in ovine TLR2

Previous studies have demonstrated that not all mutations necessarily impact upon the structure or the function of a protein (Bromham and Penny, 2003; Douville et al., 2010; Kim et al., 2007). As such, it is important to identify which sites are likely to affect function.

In this study we correlated mutations to regions of known function and to sites undergoing adaptive change. Of the mutations identified in the ovine data set, five are good candidates with the potential to impact upon the biology of TLR2. Three mutations are situated in the central region involved in PAMP recognition, in close proximity to sites under positive selection. As such, these mutations have the potential to impact ligand detection. Two SNPs (D225A, R315W) result in radical amino acid replacements and one SNP (V388F) is an un-favoured amino acid replacement in extracellular regions of proteins (Bets and Russell, 2003). D225A is detected in five of the ten sheep breeds analysed whilst the remaining two SNPs were breed specific, at least in this data set. R315W was detected in the Djallonke breed and V388F was detected in the Suffolk breed and may convey breed specificity.

Two remaining SNPs of interest are positioned within the TIR signalling domain. Functional constraint was found to be acting on the TIR domain. This was expected due to its role in signalling (Werling et al., 2009). F670L, is a common variant, at least in this data set, and has previously been associated with increased susceptibility to M. avium subsp. paratuberculosis infection in Tsigai sheep and demonstrated to result in decreased cytokine (IL-4, IL-8, IL-10, IL-12 and IFN-γ) expression in dendritic cells compared with the wild type form (Bhide et al., 2009). This leads to the expectation that this SNP has the ability to alter the immune response of a wider selection of breeds and to impact a variety of PAMPs detected by ovine TLR2. Finally, R697C was located in two sheep breeds in this study, Djallonke and Red Maasai breeds. This SNP is positioned within the BB loop of the TIR domain, the region involved in dimerization and the interaction with adaptor proteins (Nyman et al., 2008). The importance of the BB loop has been demonstrated in mouse TLR4 where a SNP was found to inhibit the response to lipopolysaccharide (LPS) (Jiang et al., 2006; Poltorak et al., 1998) indicating the potential for the SNP in sheep TLR2 to affect signal transduction. Should this SNP be beneficial to these breeds in host defence then we might surmise that natural selection would act to maintain the mutation and it would spread throughout populations. However, natural selection would also act to remove deleterious mutations from the population, and their removal would require a number of generations. The random incorporation of a deleterious mutation to the host gene could inadvertently offer a competitive advantage to pathogens by altering the immune response of the host to infection.

4.5. Recent divergence within a breed diverse species

We combined our analyses of adaptive evolution across the mammalian phylogeny to examine the long term evolutionary impact across the TLR2 gene, compared to the recent, close range divergence occurring within a species by using domestic sheep as an example. At least five haplotypes have been identified in sheep, indicating multiple domestication events (Guo et al., 2005; Hiendleder et al., 2002; Meadows et al., 2007; Pedrosa et al., 2005). The multiple haplotypes within sheep provide a unique opportunity to assess the extent of recent divergence and ongoing adaptation within a species.

It is possible that certain SNPs detected in our data set (G615T, C943T, G1162T, T1401C, G1688A, G2168A) may be the result of a random mutation within a specific breed. If this is the case then the selective breeding process may still result in the increased risk of fixation or removal of a mutation from fragmented or small populations. Should, due to association with a beneficial trait, a mutation with a deleterious impact on infection protection become fixed in a population then this could have adverse effects as it spreads through the wider breed population, conversely beneficial alleles may be lost through random genetic drift. The differential distribution of SNPs may also be the result of the domestication events that gave rise to domestic sheep. Sheep breeds that arose from different domestication events may have acquired varied mutations to other founding haplotypes. Multiple domestication
events could account for the sporadic distribution of SNPs identified as breeds that may have acquired different mutations with some SNPs specific to haplotypes. There is a lack in consensus surrounding the origins of domestic sheep (Hiendleder et al., 2002). The rapid diversification of domestic sheep (Ovis aries) breeds is thought to contribute to difficulties in defining taxonomic relationships and the uncertainty in ancestral relationships further complicates the phylogeny (Guo et al., 2007). The genus Ovis includes both wild and domesticated sheep and there is debate surrounding the true ancestors of the genus Ovis (Piper and Ruvinsky, 1997). Three groups of wild sheep are often put forward as candidate ancestors of domestic sheep; mouflon, urial and argali (Hiendleder et al., 2002), thus further complicating the construction of a breed phylogeny. The multiple domestication events in sheep may account for their high genomic diversity compared to other domestic species, such as dogs and cattle (Kijas et al., 2012). Genetic variation detected in the sheep data set indicates that on-going adaptation within the TLR2 gene may be occurring locally, enabling breed/region specific responses to evolve in response to pathogens. Should this be the case then based on our evolutionary analysis of TLR2 it appears that beneficial alleles have not yet sufficiently spread throughout the species to be considered statistically significant adaptive sites. It is worth noting that with over 1400 sheep breeds (Rischkowsky and Pilling, 2007) the domestic sheep (Ovis aries) breeds is thought to contribute to diversity in domestic animals. (Hiendleder et al., 2002; Jin and Lee, 2008; Jin et al., 2007). There is clear evidence such plays a prominent role in innate immunity (Andersen-Nissen et al., 2006). Disseminating structural genomics data to the public: from a data dump to a representative annotation and space complexity. BMC Bioinformatics 5, 113.Entrican, G., Wattgeredera, S., Wheelhouse, N., Allan, A., Rocchi, M., 2010. Immunological paradigms and the pathogenesis of ovine chlamydial abortion. American Journal of Reproductive Immunology 64, 287–294. Ewing, B., Green, P., 1998. Base-calling of automated sequencer traces using phred. II. Error probabilities. Genome Research 8, 186–194. Farhat, K., Pfenningberg, S., Heine, H., Debarry, J., Lang, R., Mages, J., Buiwitt-Beckmann, U., Roeschmann, K., Jung, G., Wiesmuller, K.H., Ulmer, A.J., 2008. Heterodimerization of TLR2 with TLR1 or TLR6 expands the ligand spectrum but does not lead to differential signaling. Journal of Leukocyte Biology 83, 652–701. Furlong, R.F., Yang, Z., 2008. Diversifying and purifying selection in the peptide binding domain of RBB in mammals. Journal of Molecular Evolution 68, 384–394. Gay, N.J., Gangloff, M., 2007. Structure and function of Toll-like receptor 2-Toll-like receptor 6 heterodimer. Immunity 31, 873–884. Guo, J., Du, L.X., Ma, Y.H., Guan, W.J., Li, H.B., Zhao, Q.J., Li, X., Rao, S.Q., 2005. A novel maternal linkage revealed in sheep (Ovis aries). Animal Genetics 36, 331–336. Hans, M., Hans, V.M., 2011. Toll-like receptors and their dual role in periodontitis: a review. Journal of Oral Science 53, 263–271. Hiendleder, S., Kaupe, B., Wassmuth, R., Janke, A., 2002. Molecular analysis of wild and domestic sheep current nomenclature and provides evidence for domestication from two different subspecies. Proceedings of the Biological Sciences 269, 895–904. Jann, O.C., Werling, D., Chang, J.S., Haig, D., Glass, E.J., 2008. Molecular evolution of bovine Toll-like receptor 2 suggests substitutions of functional relevance. BMC Evolutionary Biology 8, 288. Jiang, Z., Georgel, P., Li, C., Choe, J., Crozat, K., Rutschmann, S., Du, X., Bigby, T., Mudd, S., Sovath, S., Wilson, I.A., Olson, A., Beutler, B., 2006. Details of Toll-like receptor: adapter interaction revealed by germ-line mutagenesis. Proceedings of the National Academy of Sciences of the United States of America 103, 10961–10966. Jin, M.S., Lee, J.O., 2008. Structures of the Toll-like receptor family and its ligand complex. Immunity 29, 182–191. Jin, M.S., Kim, S.E., Heo, J.Y., Lee, M.E., Kim, H.M., Paik, S.G., Lee, H., Lee, J.O., 2007. Crystal structure of the TLR1–TLR2 heterodimer induced by binding of a tri-acylated lipopeptide. Cell 130, 1071–1082. Kang, J.Y., Nan, X., Jins, M.S., You, S.J., Ryu, Y.H., Ma, S., Han, S.H., Lee, H., Paik, S.G., Lee, J.O., 2009. Recognition of lipopeptide patterns by Toll-like receptor 2-Toll-like receptor 6 heterodimer. Immunity 31, 873–883. Kijas, J.W., Lenstra, J.A., Hayes, B., Beirard, S., Neto, L.R.P., San Cristobal, M., Servin, B., McCulloch, R., Whan, V., Gietzen, K., Paiva, S., Barendse, W., Cianni, E., Raadsma, H., McEvany, J., Dalrymple, B., Consortium, I.S.G., 2012. Genome-wide analysis of the world’s sheep breeds reveals high levels of historic mixture and strong recent selection. PLoS Biology 10. Kim, H.M., Park, B.S., Kim, J.J., Kim, S.E., Lee, J., Oh, S.C., Enkhbayar, P., Matsushima, N., Lee, H., Yoo, O.J., Lee, J.O., 2007. Crystal structure of the TLR4–MD2 complex with bound endotoxin antagonist Eritoran. Cell 130, 906–917. Kimbell, D.A., Beutler, B., 2001. The evolution and genetics of innate immunity. Nature Reviews. Genetics 2, 256–267. Krog, A., Larsson, B., von Heijne, G., Sonnhammer, E.L., 2001. Predicting transmembrane protein topology with a hidden Markov model: application to complete genomes. Journal of Molecular Biology 305, 567–580. Letunic, I., Doerks, T., Bork, P., 2009. SMART 6: recent updates and new developments. Nucleic Acids Research 37, D229–D232. Macnab, R.N., 1992. Genetics and biogenesis of bacterial flagella. Annual Review of Genetics 26, 121–158. Meadows, J.R., Cemal, I., Karaca, O., Gootwine, E., Kijas, J.W., 2007. Five ovine mitochon-
Meng, G., Grabiec, A., Vallon, M., Ebe, B., Hampel, S., Bessler, W., Wagner, H., Kirschning, C.J., 2003. Cellular recognition of tri-/di-palmitoylated peptides is independent from a domain encompassing the N-terminal seven leucine-rich repeat (LRR)/LRR-like motifs of TLR2. Journal of Biological Chemistry 278, 39822–39829.

Merk, S., Neumaier, M., Wagner, H., Kirschning, C.J., Ahmad-Nejad, P., 2007. Characterization and investigation of single nucleotide polymorphisms and a novel TLR2 mutation in the human TLR2 gene. Molecular Human Genetics 16, 1225–1232.

Mitsuzawa, H., Wada, I., Sano, H., Iwaki, D., Murakami, S., Hinii, T., Matushima, N., Kuroki, Y., 2003. Extracellular Toll-like receptor 2 region containing Ser40–Ile64 but not Cys30–Ser39 is critical for the recognition of Staphylococcus aureus peptidoglycan. Journal of Biological Chemistry 278, 41350–41356.

Mun, H.S., Aosai, F., Noruse, K., Chen, M., Piao, L.X., Takeuchi, O., Akira, S., Ishikura, H., Yano, A., 2003. TLR2 as an essential molecule for protective immunity against Toxoplasma gondii infection. International Immunology 15, 1081–1087.

Murphy, W.J., Eizirik, E., Johnson, W.E., Zhang, Y.P., Ryder, O.A., O’Brien, S.J., 2001. Molecular phylogenetics and the origins of placental mammals. Nature 409, 614–618.

Murphy, K., Travers, P., Walport, M., 2008. Janeway’s Immunobiology, 7th ed. Garland Science, Taylor & Francis Group, New York.

Ngamphiw, C., Kulawonganunchai, S., Assawamakin, A., Jenwitheesuk, E., Tongsima, S., 2008. VarDetect: a nucleotide sequence variation exploratory tool. BMC Bioinformatics 9 (Suppl. 12), S9.

Niemela, R., Yang, Z., 1998. Likelihood models for detecting positively selected amino acid sites and applications to the HIV-1 envelope gene. Genetics 148, 929–936.

Nielsen, R., Yang, Z., 2000. Codon-substitution models for heterogeneous selection pressure at amino acid sites. Genetics 155, 2172–2186.

Piper, L., Rusinsky, A., 1997. The Genetics of Sheep. Cambridge University Press, Oxon.

Pond, S.L.K., Frost, S.D.W., 2005. Not so different after all: a comparison of methods for detecting amino acid sites under selection. Molecular Biology and Evolution 22, 1208–1222.

Rischkowsky, B., Pilling, D., 2007. The State of the World’s Animal Genetic Resources for Food and Agriculture. Commission on Genetic Resources for Food and Agriculture, Rome, Italy.

Rozen, S., Skaltsky, H., 2000. Prime3 on the WWW for general users and for biologist programmers. Methods in Molecular Biology 132, 365–386.

Smith, S.A., Jann, O.C., Haig, D., Russell, G.C., Werling, D., Glass, E.J., Emes, R.D., 2012. Adaptive evolution of Toll-like receptor 5 in domesticated mammals. BMC Evolutionary Biology 12, 122.

Staden, R., 1996. The Staden sequence analysis package. Molecular Biotechnology 5, 746–757.

S.A. Smith et al. / Gene 540 (2014) 217–225

Wlasiuk, G., Nachman, M.W., 2010. Adaptation and constraint at Toll-like receptors in primates. Molecular Biology and Evolution 27, 2172–2186.

Wong, W.S., Yang, Z., Goldman, N., Nielsen, R., 2004. Accuracy and power of statistical methods for detecting adaptive evolution in protein coding sequences and for identifying positively selected sites. Genetics 168, 1041–1051.

Woolhouse, M.E., Webster, J.P., Domingo, E., Charlesworth, B., Levin, B.R., 2002. Biological and biomedical implications of the co-evolution of pathogens and their hosts. Nature Genetics 32, 569–577.

Yang, Z., 2007. PAML 4: phylogenetic analysis by maximum likelihood. Molecular Biology and Evolution 24, 1596–1599.

Yang, Z., Nielsen, R., Goldman, N., Pedersen, A.M.K., 2000. Codon-substitution models for heterogeneous selection pressure at amino acid sites. Genetics 155, 431–440.

Yoon, S.I., Kurnasov, O., Natarajan, V., Hong, M., Gudkov, A.V., Osterman, A.L., Wilson, I.A., 2012. Structural basis of TLR5-flagellin recognition and signaling. Science 335, 859–864.

Zhang, C.J., Wang, J., Xie, W.B., Zhao, G., Long, M.Y., Zhang, Q.F., 2011. Dynamic programming procedure for searching optimal models to estimate substitution rates based on the maximum-likelihood method. Proceedings of the National Academy of Sciences of the United States of America 108, 7860–7865.