Signatures of positive selection in *LY96* gene in vertebrates

TONGHAI DOU[1,†], MAOBIN FU[1,†], YIXIA WANG[1], YANG ZHAO[1], ZHENGSHI WANG[2], ZHENGQIAN BIAN[2] and YAN ZHOU[1,3,*]

[1]*Department of Microbiology and Microbial Engineering, School of Life Sciences, Fudan University, Shanghai 200433, People's Republic of China*

[2]*Department of General Surgery, Renji Hospital, Shanghai Jiao Tong University, School of Medicine, Shanghai 200127, People's Republic of China*

[3]*Shanghai-MOST Key Laboratory of Health and Disease Genomics, Chinese National Human Genome Center at Shanghai, Shanghai 201203, People's Republic of China*

[†]*These authors contributed equally to this work.*

**Corresponding author (Fax, 86-21-55665487; Email, zhouy@chgc.sh.cn)*

As a secreted glycoprotein that binds to the extracellular domain of Toll-like receptor 4 (*TLR4*), Lymphocyte Antigen 96 (*LY96*), also called myeloid differentiation 2 (*MD2*), is required for the activation of *TLR4* by lipopolysaccharide (LPS) and plays an important role in innate immunity, which is the first line of defence against microbial infections. Previous studies have proposed that mammalian toll-like receptors (TLRs) have evolved under diversifying selection due to their role in pathogen detection. Given the fact that *LY96* is highly functionally linked to *TLR4*, it would be interesting to test whether *LY96* is under the intense pressure of natural selection. To investigate the natural selection hypothesis, we compared the coding sequences from 13 vertebrates and evaluated the molecular evolution of *LY96* gene in these species. Result shows that natural selection at exon 4 has indeed played a role in shaping the function of *LY96* in the course of evolution. In addition to the study of Nakajima, we found the two branch nodes with Ka/Ks ratios greater than 1: the one leading to cow and pig and the other to rabbit and the primates.

[Dou T, Fu M, Wang Y, Zhao Y, Wang Z, Bian Z and Zhou Y 2013 Signatures of positive selection in *LY96* gene in vertebrates. *J. Biosci.* **38** 899–904]
DOI 10.1007/s12038-013-9377-9

1. Introduction

Lymphocyte Antigen 96 (*LY96*), also called myeloid differentiation 2 (*MD2*), was discovered by Shimazu for the first time in 1999 (Shimazu *et al.* 1999). *LY96* is a 160-amino-acid glycoprotein with five exons (Genbank accession number: NM_015364) (Gangloff and Gay 2004). *LY96* plays crucial roles in lipopolysaccharide (LPS)-induced inflammatory responses as an accessory protein of toll-like receptor 4 (*TLR4*) (Gioannini *et al.* 2004), and is associated with the development and progression of tumour such as colorectal cancer (Grondin *et al.* 2011). Human *LY96* undergoes N-linked glycosylation at Asn (26) and Asn (114), and these glycosylations are crucial for *TLR4*-mediated signal transduction of LPS (Ohnishi *et al.* 2001). *LY96* has an MD-2-related lipid-recognition (ML) domain (Inohara and Nunez 2002), which is implicated in lipid recognition, particularly in the recognition of pathogen-related products. Neither the transmembrane domain nor the intracellular domain exists in *LY96* protein.

Given that increasing evidence suggests that *TLR4* has come under natural selection pressure (Nakajima *et al.* 2008; Smirnova *et al.* 2001), there is a good chance for *LY96* to be under the intense pressure of natural selection during the

Keywords. *LY96*; Ka/Ks; positive selection; *TLR4*; vertebrates

Supplementary materials pertaining to this article are available on the *Journal of Biosciences* Website at *http://www.ias.ac.in/jbiosci/dec2013/supp/Dou.pdf*

Published online: 6 November 2013

course of evolution as well, since its function is highly linked to *TLR4*. After analysing nucleotide and protein sequences from 25 species, including eight hominoids, six Old World monkeys, eight New World monkeys, and three prosimians, Nakajima *et al.* found that non-synonymous and synonymous substitution ratio for *LY96* was 0.269 among seven primates (Nakajima *et al.* 2008). The fact Ka/Ks was lower than 1.0 indicated that *LY96* had been under the pressure of negative or purifying selection among primates. However, this result came from the level of entire coding sequence. In some cases, functional exons, especially conserved domains, are subjected to a considerably more strict selection pressure than the full-length sequences (Cruciani and Mikalsen 2007). Therefore, it is important to study the evolutionary condition of individual functional domain. Molecular evolution approaches and the software package phylogenetic analyses by maximum likelihood (PAML) were utilized to investigate the signature of selection that has acted on the *LY96* gene. Here we investigated the adaptive evolution of *LY96* at the level of individual exon and separated lineages. We found evidence of adaptive evolution of vertebrate *LY96* in the specific exon and specific lineages leading to primates and cow.

2. Material and methods

2.1 *Sequence data collection*

LY96 DNA and protein sequences were retrieved from the National Center for Biotechnology Information (NCBI) and Ensembl, which were verified subsequently by EST-Blast in NCBI (*http://blast.ncbi.nlm.nih.gov/Blast.cgi*).

Coding sequences of 13 species used in this study include *Macacamulatta* (NM_001130432.1), *Homo sapiens* (NM_015364.4), *Pan troglodytes* (NM_001130474.1), *Musmusculus* (NM_019716.2), *Rattusnorvegicus* (NM_001033690.1), *omascusleucogenys* (ENSNLET 00000002088), *Oryctolaguscuniculus* (NM_001082787.2), *Canisfamiliaris* (XM_848045.1), *Susscrofa* (ENSSSCT 00000035167), *Caviaporcellus* (ENSCPOT00000009772), *Bostaurus* (NM_001046517.1), *Gallus gallus* (XM_418301.2), and *Meleagrisgallopavo* (ENSMGAT 00000012981).

2.2 *Evolutionary analysis*

The protein coding sequences of *LY96* were aligned using MUSCLE program implemented in MEGA5 (Tamura *et al.* 2011), and then were back translated to obtain codon alignments. Every single exon homologs alignment was also performed. We manually inspected and modified alignment output to fit the demand required by software used. The phylogenetic tree of all *LY96* protein-coding sequences were reconstructed with neighbour-joining method, which was implemented in MEGA5 (figure 1).

2.3 *Detection of positive selection among LY96 exons and lineages*

Non-synonymous/synonymous substitution ratios (ω = dN/dS, also called Ka/Ks) was used for quantifying the impact of natural selection on molecular evolution (Ohta 1992).

The pair-wise computation of Ka/Ks between every exon sequence and lineage-specific model of the likelihood method were performed using the program codeML implemented in PAML software package version 4.4 (Yang 2007). To test whether different lineages are evolving at different rates by allowing for variable ω among branches in the tree, but invariable ω among different amino acid sites (Nielsen and Yang 1998), a model with one ω (H0) was compared with a 'free-ratio' model that allows each branch to have a separate ω value while keeping variation among sites constant. Then we examined every mammalian node in the phylogenetic tree generated above. 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 χ^2 distribution with 2 degrees of freedom.

3. Results and discussion

3.1 *Phylogenetic tree and exon profile analysis*

A phylogenetic tree based on amino acid sequence was generated to demonstrate the relationship of *LY96* among 13 species (figure 1). Generally speaking, genetic relationship between rodents (e.g. mouse and rat) and primates is close (Murphy *et al.* 2001), but the phylogenetic tree of *LY96* gene showed that their relationship was distant relatively.

Comparative genomics has been exploited as a common tool to analyse the conservation of individual exon based on the degree of sequence similarity. Our *LY96* exon profile analysis comprised 13 species ranging from turkey to human. Among all five exons, exon 2 is an alternative splicing exon and others are constitutive exons. Interestingly, exon 4 exhibited the highest degree of sequence conservation (termed as the identity level).

3.2 *Positive selection observed on Exon4*

The non-synonymous to synonymous rate ratio Ka/Ks (also known as dN/dS, or ω) is an indicator of the change of

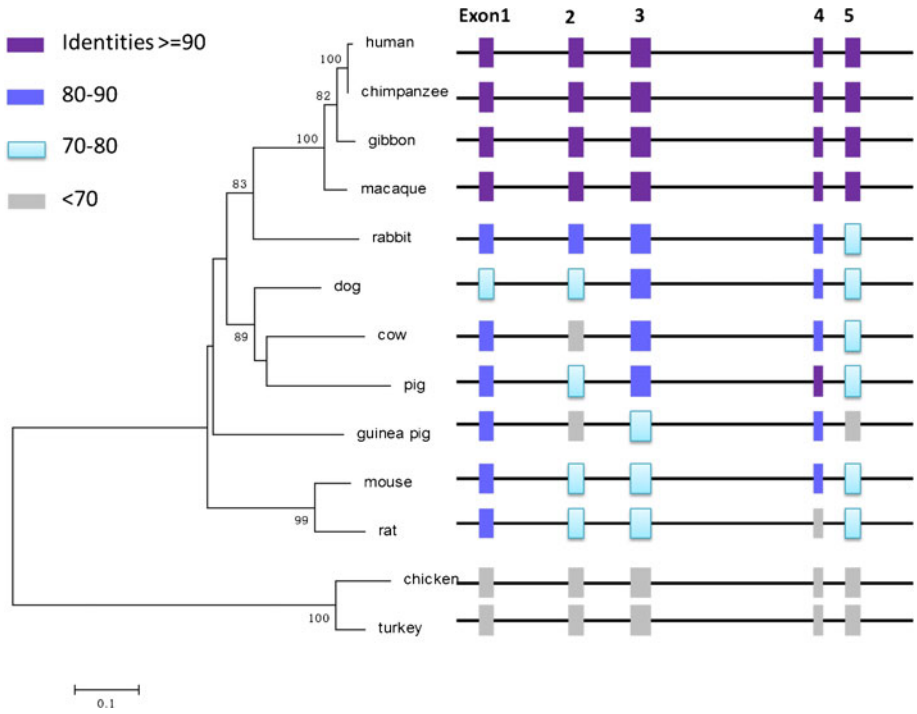

Figure 1. Phylogenetic tree and exon profile of the LY96 sequences from 13 species. The phylogenetic tree of all LY96 protein sequences were reconstructed with neighbour-joining method, which was implemented in MEGA5 (1000 bootstrap replicates, and condensed tree cut-off value 50%). This tree was used as the fixed tree topology for subsequent analysis. Straight lines stand for genomes and Squares represent exons, and different colours correspond to similarity to human exons.

selective pressures. The Ka/Ks ratios of > 1, = 1, and < 1 indicate positive selection, neutral evolution, and purifying selection on the protein involved, respectively (Ohta 1992). To investigate the evolution situation of exons, we calculated the pair-wise Ka/Ks ratio of each exon between human and mouse (figure 2a), and found that Ka/Ks ratios of both exon 2 and exon 4 were greater than 1. Especially, Ka/Ks ratio of exon4 was up to 2.38, which indicated that exon 4 had undertaken high selective pressure.

Subsequently we analysed the Ka/Ks ratios of exon 4 between human and every species respectively (figure 2b). As the results showed, except for three infinity data points (N.A.), five out of nine Ka/Ks ratios were greater than 1. It indicates that high Ka/Ks ratio of exon 4 is not a random event but an evolutionary phenomenon. Rapid evolution o fexon 4 has occurred.

To investigate the exact function of exon 4, we predicted protein domainof *LY96* gene using InterProScan software (Mulder and Apweiler 2007), and found that exon 4 was localized in MD-2-related lipid-recognition (ML) domain (IPR003172; supplementary figure 5). This domain plays roles in LPS signalling and lipid metabolism, which was a novel domain identified in MD-1, MD-2, GM2A, Npc2 and multiple proteins of unknown function in plants, animals and fungi (Inohara and Nunez 2002). Thus, rapid evolution of exon 4 may have an influence on the function change of *LY96* gene (data of other exons were shown in supplementary figures 1–4).

Evidence of positive selection was found in the region of exon 4, the middle part of MD-2-related lipid-recognition (ML) domain, which suggests that amino acid residues present in the middle part of the ML domain evolve faster than others. Exon 4 encodes amino acid residues from 110 to 128. Previous study shows that the glycosylation at Asn (114) was found to be crucial for TLR4-mediated signal transduction of LPS (Ohnishi *et al.* 2001). It has also been reported in several studies that LPS directly binds to MD-2 in a highly basic region (amino acids 119–132) (Mancek *et al.* 2002; Mullen *et al.* 2003; Gioannini *et al.* 2004). Muroi *et al.* also showed that amino acid region 108–135 of MD-2 determine the species-specific activity of Lipid IVa, the active moiety of LPS precursor, and amino acid residues 122 of mouse MD-2 was critical to determine the agonist–antagonist activity of Lipid IVa, which suggested the amino acid residue may be involved in the discrimination of lipid A structure (Muroi and Tanamoto 2006). Crystal structure of

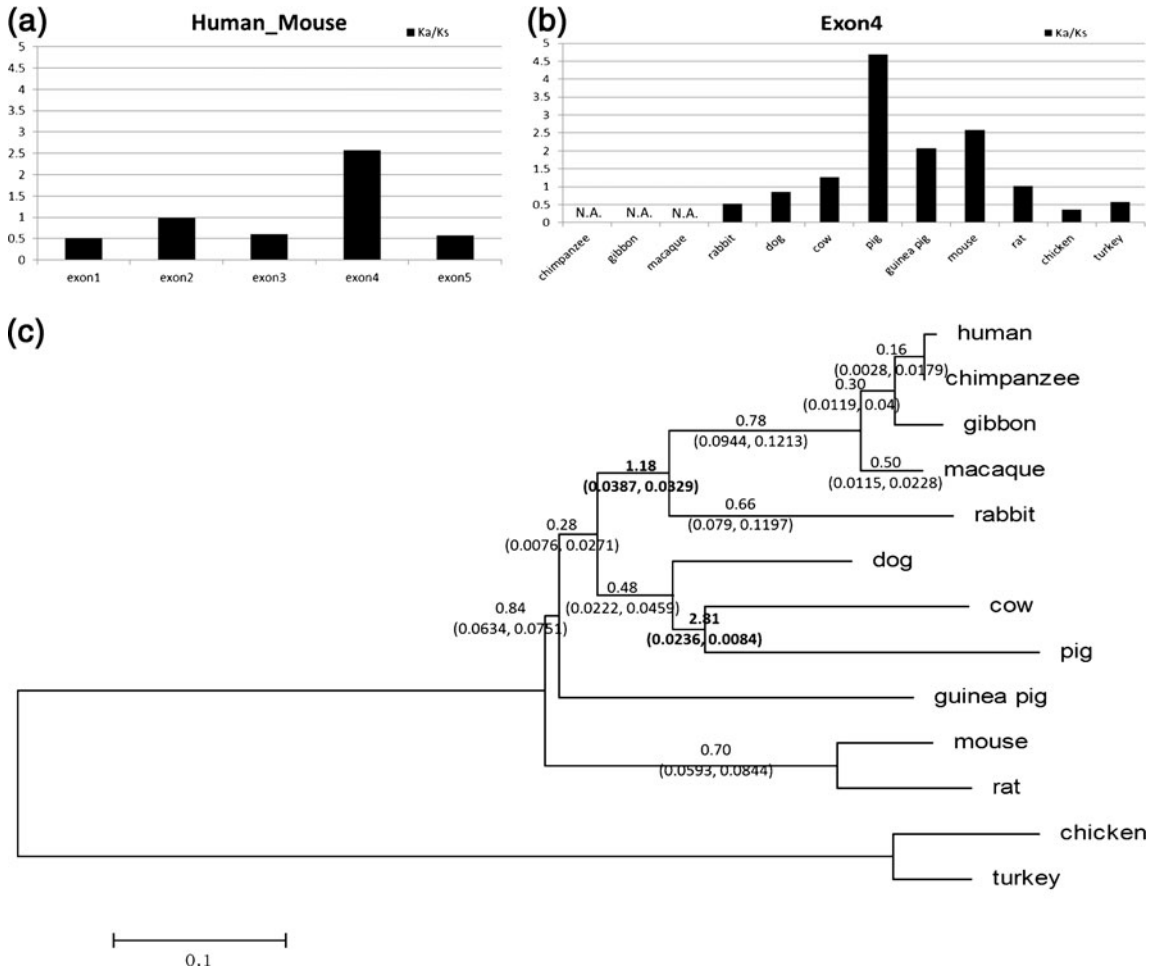

Figure 2. (**a**) The Ka/Ks ratios of different exons of LY96 between mouse and human. Y axis represents Ka/Ks ratio and X axis represents different exons. (**b**) The Ka/Ks ratios of exon4 of LY96 between human and other species. The Y axis represents Ka/Ks ratio and the X axis represents different species. In some cases, zero synonymous substitutions lead to a Ka/Ks ratio of infinity (N.A.). (**c**) Lineage-specific Ka/Ks values of LY96 in species analyzed for the entire gene. Estimated Ka/Ks values from the branch-based model are shown above branches and the estimated number of nonsynonymous and synonymous changes are shown in parentheses below branches. Branches with Ka/Ks values greater than 1 are shown in bold.

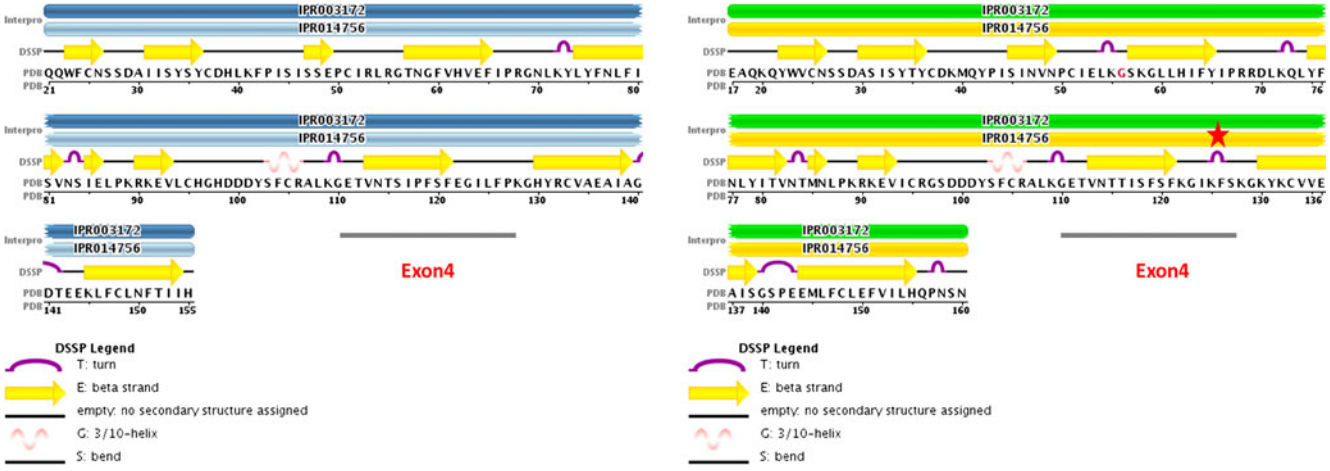

Figure 3. Secondary structure comparison of mouse (left) and human (right) LY96 proteins. The grey line represents the amino acid region coded by exon 4; the red star marks the differences between two sequences.

human and mouse *LY96* have been determined [human: 2E56:A (Ohto *et al.* 2007), mouse: 2Z64:C (Kim *et al.* 2007)]. By comparing the corresponding part of two secondary structures (Kabsch and Sander 1983), we found that there existed an extra turn in human *LY96* protein than protein of mouse, which is located exactly in exon 4 (figure 3).

Given the observation we got above, exon4 is more likely to be linked to the susceptibility to LPS in humans, variation or evidence of positive selection in which may be the result of host adaptation to pathogen evolution. The result of these variations in the function and structure of the molecule remains is to be assessed.

3.3 *Positive selection on LY96 gene lineage*

Nakajima *et al.* reported that *LY96* gene had been under the pressure of negative selection among seven primates (Nakajima *et al.* 2008). Extending to other vertebrates, we further investigated the evolution history of *LY96* among 13 species including primates. The LRT tests based on the branch model suggested that the free-ratio model fitted the data better than did the one-ratio model (H0 versus free-ratio, LRT=15.18, $p<0.001$), indicating that Ka/Ks ratios were indeed different among lineages. We found the positive selection on *LY96* gene in two branch nodes with Ka/Ks ratios greater than 1, the branch leading to cow and pig and the branch to rabbit and the primates (figure 2c). Although no adequate functional study has been constructed on *LY96* of species like cow and rabbit, the fact that Ka/Ks ratio is higher than 1 indicates that *LY96* that is evolutionarily active in certain species and has the potential to gain various functions to defense against variable pathogens.

Acknowledgements

This work was supported by the National Natural Science Foundation of China (grant nos. 31071223, and 31071158), the Ministry of Science and Technology, PRC (grant nos. 2009FY120100, 2011CB111610, and 2012AA020409), the Science and Technology Commission of Shanghai Municipality (grant nos. 11JC1400900 and 12ZR1421300).

References

Cruciani V and Mikalsen SO 2007 Evolutionary selection pressure and family relationships among connexin genes. *Biol. Chem.* **388** 253–264

Gangloff M and Gay NJ 2004 MD-2: the Toll 'gatekeeper' in endotoxin signalling. *Trends Biochem. Sci*. **29** 294–300

Gioannini TL, Teghanemt A, Zhang DS, Coussens NP, Dockstader W, Ramaswamy S and Weiss JP 2004 Isolation of an endotoxin-MD-2 complex that produces Toll-like receptor 4-dependent cell activation at picomolar concentrations. *Proc. Natl. Acad. Sci. USA* **101** 4186–4191

Grondin V, Seksik P, Dumont S, Thomas G, Trugnan G, Flejou JF, Masliah J, Wendum D and Bachelet M 2011 Regulation of colon cancer cell proliferation and migration by MD-2 activity. *Innate Immun.***17** 414–422

Inohara N and Nunez G 2002 ML - a conserved domain involved in innate immunity and lipid metabolism. *Trends Biochem. Sci*. **27** 219–221

Kabsch W and Sander C 1983 Dictionary of protein secondary structure: pattern recognition of hydrogen-bonded and geometrical features. *Biopolymers* **22** 2577–2637

Kim HM, Park BS, Kim JI, Kim SE, Lee J, Oh SC, Enkhbayar P, Matsushima N, Lee H, Yoo OJ, *et al.* 2007 Crystal structure of the TLR4-MD-2 complex with bound endotoxin antagonist Eritoran. *Cell* **130** 906–917

Mancek M, Pristovsek P and Jerala R 2002 Identification of LPS-binding peptide fragment of MD-2, a toll-receptor accessory protein. *Biochem. Biophys. Res. Commun.* **292** 880–885

Mulder N and Apweiler R 2007 InterPro and InterProScan: tools for protein sequence classification and comparison. *Methods Mol. Biol*. **396** 59–70

Mullen GE, Kennedy MN, Visintin A, Mazzoni A, Leifer CA, Davies DR and Segal DM 2003 The role of disulfide bonds in the assembly and function of MD-2. *Proc. Natl. Acad. Sci. USA* **100** 3919–3924

Muroi M and Tanamoto K 2006 Structural regions of MD-2 that determine the agonist-antagonist activity of lipid IVa. *J. Biol. Chem*. **281** 5484–5491

Murphy WJ, Eizirik E, Johnson WE, Zhang YP, Ryder OA and O'Brien SJ 2001 Molecular phylogenetics and the origins of placental mammals. *Nature* **409** 614–618

Nakajima T, Ohtani H, Satta Y, Uno Y, Akari H, Ishida T and Kimura A 2008 Natural selection in the TLR-related genes in the course of primate evolution. *Immunogenetics* **60** 727–735

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

Ohnishi T, Muroi M and Tanamoto K 2001 N-linked glycosylations at Asn(26) and Asn(114) of human MD-2 are required for toll-like receptor 4-mediated activation of NF-kappa B by lipopolysaccharide. *J. Immunol.* **167** 3354–3359

Ohta T 1992 The nearly neutral theory of molecular evolution. *Annu. Rev. Ecol. System.* **23** 263–286

Ohto U, Fukase K, Miyake K and Satow Y 2007 Crystal structures of human MD-2 and its complex with antiendotoxic lipid IVa. *Science* **316** 1632–1634

Shimazu R, Akashi S, Ogata H, Nagai Y, Fukudome K, Miyake K and Kimoto M 1999 MD-2, a molecule that confers lipopolysaccharide responsiveness on Toll-like receptor 4. *J. Exp. Med*. **189** 1777–1782

Smirnova I, Hamblin MT, McBride C, Beutler B and Di Rienzo A 2001 Excess of rare amino acid polymorphisms in the Toll-like receptor 4 in humans. *Genetics* **158** 1657–1664

Tamura K, Peterson D, Peterson N, Stecher G, Nei M and Kumar S 2011 MEGA5: Molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. *Mol. Biol. Evol.* **28** 2731–2739

Yang Z 2007 PAML 4: phylogenetic analysis by maximum likelihood. *Mol. Biol. Evol.* **24** 1586–1591

MS received 07 May 2013; accepted 26 July 2013

Corresponding editor: SARAH H ELSEA