EVOLUTIONARY ANALYSIS OF SUIDAE TLR SIGNALING PATHWAY

BY

KWAME ASANTE DARFOUR-ODURO

DISSERTATION

Submitted in partial fulfillment of the requirements
for the degree of Doctor of Philosophy in Animal Sciences
in the Graduate college of the
University of Illinois at Urbana-Champaign, 2015

Urbana, Illinois

Doctoral committee:

Professor Lawrence B. Schook, Chair, Director of Research
Professor Bryan White
Associate Professor Alfred L. Roca
Assistant Professor Jian Ma

Abstract

The innate immune system represents the first line of host defense against pathogens. Toll-like receptors (TLRs) are the best characterized pathogen recognition receptors (PRRs) of the innate immune system. TLRs detect distinct pathogen associated molecules and engage downstream intracellular signaling cascades eventually leading to an innate immune response. Thus, the TLR signaling pathway is under selective pressure from pathogens and is essential for host survival. The family Suidae (domestic pigs and their wild relatives) has diverged over extended evolutionary periods in diverse environments, suggesting that adaptation in response to endemic infectious agents may have occurred. It is therefore expected that the TLR signaling pathway might have played a crucial role in the survival of members of the family Suidae. However, the extent to which pathogen mediated selection pressures have influenced the evolution of family Suidae TLR signaling pathway genes is not well understood. Investigating the role of pathogens in the genetic variation within the TLR signaling pathway genes will indicate the critical role of these genes in host defense against prior and present infections.

In this dissertation I first determined whether members of the family Suidae TLRs have evolved adaptively at the interspecies level. To this end, coding sequences of bacterial sensing TLR (*TLR1, TLR2* and *TLR6*) and viral sensing TLR (*TLR3, TLR7* and *TLR8*) were retrieved from resequencing libraries of one animal representing each of the following members of the family Suidae: *Sus scrofa* from Europe, *Sus scrofa* from Asia, *Potamochoerus larvatus*, *Potamochoerus porcus* and *Phacochoerus africanus* of sub Saharan African origin and *Sus verrucosus*, *Sus celebensis*, *Sus scebifrons*, *Sus barbatus* and *Babyrousa babyrussa* from island Southeast Asia. Evolutionary analyses of the aligned TLR sequences were done using Maximum likelihood (ML) approaches in a phylogenetic framework. These analyses indicated that persistent positive selection pressures have acted on amino acid residues across the family Suidae in both bacterial and viral sensing TLRs. There were more amino acid sites under positive selection in bacterial sensing TLRs than their viral sensing counterparts. Lineage specific positive selective events, where positive selective were inferred for particular family Suidae species were

suggestive of species specific pathogens that are acting as selection pressures. Some of the amino acid residues under positive selective pressure were involved in radical amino acid substitutions and were also within or in close proximity to functionally important domains of the TLR proteins. Thus, such residues may have disease resistance implications for the family Suidae.

Adaptive evolution was also investigated at the intraspecies level. The European *Sus scrofa* (wild boars and domestic pigs) diverged from their Asian counterparts over 1million years ago. Distinct pathogen mediated selective pressures might have acted on these geographically separated groups leading to signatures of adaptation in host genes. To investigate whether positive selection pressures have acted on TLRs of *Sus scrofa* of European and Asian origins, coding sequences of bacterial sensing TLR (*TLR1, TLR2 and TLR6*) and viral sensing TLR (*TLR3, TLR7 and TLR8*) were retrieved from resequencing libraries of 15 wild boars and 25 domestic pigs of European origin and 5 wild boars and 22 domestic pigs of Asian origin. Within and between populations analyses of positive selection indicated that the *TLR2* gene was under positive selective pressure in European *Sus scrofa* but not in the Asian *Sus scrofa*. Specifically, the derived allele (nucleotide: A; Amino acid: Threonine) of *TLR2* SNP A376G (Alanine 126 Threonine) was under positive selection. The frequency of the derived allele was 83.33% within European wild boars, 98.00% within domestic pig breeds of European origin, 40.00% within Asian wild boars and 11.36% within Asian domestic pigs. The age of the derived allele was 163,000 years which roughly coincided with a time of *Sus scrofa* population expansion. This population expansion might have created an environment for pathogen transmission providing the selective force for adaptation at host genes. Three dimensional crystal structure of the *TLR2* protein indicated the derived allele under positive selection was located within the N-terminal domain of the extracellular domain where single amino acid substitutions are likely to affect protein function.

TLRs act in concert with other genes in the signaling pathway to elicit innate immune response to invading pathogens. A key question in molecular evolution is whether parameters of signaling pathways have an influence on how genes evolve. The evolution of the entire TLR signaling pathway in the context of parameters of the pathway was therefore investigated. In particular, I investigated whether gene

position within TLR signaling pathway of the family Suidae have an effect on evolutionary rate of genes. The goal here was to determine whether genes upstream and downstream have distinct selection pressures acting on them and the factors responsible for such a polarity. To do this, genes within the TLR signaling pathway as indicated in the KEGG database were retrieved from resequencing libraries of *Sus scrofa* from Europe, *Sus scrofa* from Asia, *Potamochoerus larvatus*, *Potamochoerus porcus* and *Phacochoerus africanus* of sub Saharan African origin and *Sus verrucosus*, *Sus celebensis*, *Sus scebifrons*, *Sus barbatus* and *Babyrousa babyrussa* from island Southeast Asia. After screening sequences for suitability for evolutionary analyses, 33 genes remained. Gene position, the number of protein-protein interactions (connectivity), protein length, length of 3' untranslated region and codon bias measured as Effective Number of Codons (ENC) were estimated for each gene sequence alignment. Gene position was significantly negatively correlated with evolutionary parameters ω (dn/ds) and dn suggesting that downstream genes in the pathway were more selectively constrained than upstream genes. Protein length and connectivity were also significantly correlated with evolutionary parameters ω (dn/ds) and dn such that downstream genes had shorter protein lengths and were connected to more proteins. As gene position was also significantly correlated with protein length and connectivity, the polarity in evolutionary rate along the TLR signaling pathway is due to differences in protein length and connectivity between upstream and downstream genes within TLR signaling pathway of the family Suidae.

To my wife Sandra Asantewaa Darfour-Oduro,
for her love and encouragement

Acknowledgements

I would like to thank my advisor Lawrence (Larry) B. Schook for giving me the opportunity to work in his lab. Larry has given me enough freedom to pursue my research interest and has also shaped the way I think and write as a scientist. I am extremely grateful to Larry for sponsoring my trips abroad to meet and interact with top scientists in the research area of pig evolution and genomics.

I also want to thank my committee member Alfred L. Roca for his guidance and helpful feedback on my manuscripts. I thank my committee member Jian Ma for helpful discussions at the initial stages of my research work and for introducing me to the linux operating system. I am further grateful to Jian Ma for allowing me to do bioinformatics work on his server and for asking his student Ryan Cunningham to assist me with python and perl programming during the initial stages of my research work. Thank you to Sandra Rodriguez-Zas for her helpful suggestions during my preliminary examination. I owe additional thanks to Bryan A. White for accepting to be on my thesis committee at a very short notice. Additionally for making me have access to whole genome sequence data used for my studies and the use of their server, I would like to thank Martien Groenen and Hendrik-Jan Megens of the Wageningen University in the Netherlands. I am grateful to Martien for allowing me to stay in his lab for one week to acquaint myself with some bioinformatics tools necessary for my research work and to Hendrik-Jan for making available some very useful scripts to manipulate sequence data. I also thank Laurie Rund of the Schook lab for her useful comments during lab meetings and her support in ensuring a good working environment. I am also grateful to all my lab mates for creating a good environment for research. Finally, I would like to thank my family for their encouragement and my wife Sandra Darfour-Oduro for her love, kindness and constant encouragement.

Table of contents

Chapter 1: General Introduction

Immunity

The health of vertebrates is constantly threatened by infectious pathogenic agents such as bacteria, viruses, parasites and fungi. Vertebrates have evolved the immune system to eliminate pathogens from the body. Upon encountering an infectious agent, the first line of defenses of the immune system is physical and chemical barriers. Physical barriers include the skin that forms an impenetrable barrier of epithelium to potentially infectious pathogens and mucous membranes linings of the mouth and nose. Chemical barriers such acidity on skin, lysozyme, gastric juice and saliva destroy infectious agents at body surfaces, openings and inner linings. When the physical and chemical barriers are breached by pathogens, the complement system, a collection of blood and cell surface proteins, named for its ability to complement the antibacterial properties of antibodies, can recognize and destroy the invading pathogens. Phagocytes such as macrophages and neutrophils can also affect the pathogens that breach the physical and chemical barriers by ingesting and destroying them. Macrophages and neutrophils can be attracted to the area of infection by chemical substances released by the pathogens. The complement system also marks pathogens with molecular flags for their destruction by phagocytes. These defenses collectively constitute the innate immune system.

The innate immune system provides an immediate defense against pathogenic infections. Upon breaching the line of defense provided by the complement system and the phagocytes of the innate immune system, antigen specific lymphocytes target the infectious agents and persist to provide long term immunological memory that enables efficient response to specific pathogens in the future. This lymphocyte mediated response constitutes the adaptive immune system. Various antigens are detected by T- and B- cell receptors formed by somatic gene arrangements that create a multitude of receptor specificities [1]. Major Histocompatibility Complex (MHC) class I and class II molecules are cell surface glycoproteins that mediate the presentation of antigens to T- and B- cell receptors, and also play an important role in priming/triggering adaptive immune responses when the bound peptide is recognized as

foreign [2]. The adaptive immune system is more complex than the innate immune system and is slower in response usually taking days or weeks. In order to defend against an invading pathogen, the innate immune system must reliably discriminate between invading pathogens (non-self-antigens) and self-molecules. This is achieved by pattern recognition receptors located on the surfaces of immune cells.

Pattern recognition receptors

PRRs are innate immune response-expressed proteins that recognize evolutionary conserved microbial structures known as pathogen associated molecular patterns (PAMPs) and initiate the innate and adaptive immune response. The PAMPs include proteoglycans, lipopolysaccharides, glycoproteins and nucleic acid motifs that are shared by different microorganisms. These PAMPS are critical to the survival of the microorganisms. PRRs are germline encoded and have been grouped into five families based on homology. The families include Toll-like receptors (TLRs), C-type lectin receptors (CLRs), cytosolic NOD-like receptors (NLRs), RIG-I-like receptors (RLRs) and AIM2-like receptors (ALRs).

PRRs are characterized by the location of their cognate ligands into those responsible for extracellular derived-ligands and soluble cytosolic ligands respectively [3]. TLRs and CLRs are found at cell surfaces or endocytic compartments and detect microbial ligands in the extracellular space and within endosomes whereas NLRs, RLRs and ALRs are located within the cytoplasm and detect intracellular pathogens. Engagement of PAMPs by PRRs triggers intracellular signaling cascades which results in the expression of proinflammatory molecules. Furthermore, PRRs also detect induced-or aberrant self and missing self [4], indicating that PRRs mediated immune response can occur in the absence of infectious agents.

An important property of PRRs is that a number of different PRRs are engaged by a given pathogen [5]. Thus, the different PRRs have ligands in common. Furthermore, different PRRs can recognize distinct PAMPs. The ability to recognize distinct PAMPs and PAMPs shared by most pathogens broadens the range of pathogens these receptors can detect enabling the host to deal with numerous infections. PRRs have other functions beyond their role in triggering immune response. They

may regulate cell proliferation, survival, apoptosis and tissue repair [6,7]. The TLR family is the best characterized PRR in terms of structure, function, ligands and the evolutionary pressure imposed by pathogens that they detect.

Structure and ligands of TLRs

In terms of structure, all TLRs are characterized by an extracellular domain which recognizes PAMPs, a transmembrane domain and Toll/interleukin receptor (TIR) domains which initiates signaling (Figure 1.1). The TLR extracellular domain consists of 19 to 25 copies of a motif known as leucine-rich repeat (LRR). LRRs are typically made up of 20-30 amino acid residues [8], folded in beta-strands arranged in sheets and helices that are linked by loops [9]. The copies of consecutive LRR motifs together have a horseshoe shape with an interior (inner concave surfaces) parallel beta sheet and an exterior (convex surface) array of helices. Each LRR motif consists of a variable region and a conserved "LxxLxLxxNxL" region. Other hydrophobic amino acids can substitute the conserved leucine residues [10]. The LRR region has cysteine clusters flanking each side at their termini, denoted as LRRNT and LRRCT [10]. These LRRNT and LRRCT modules stabilize the protein by protecting its hydrophobic core from exposure to solvent [9].

Most mammalian species have between 10 to 15 TLRs. In humans, 10 TLRs have been identified. TLRs can be divided into two groups based on their cellular location (Figure 1.2). TLRs (*TLR1, TLR2, TLR4, TLR5*, *TLR6* and *TLR10*) are expressed on cell surface and recognize predominantly bacterial ligands and several fungal and parasite ligands (Table 1) whereas *TLR3, TLR7, TLR8* and *TLR9* are expressed within endosome and recognize single and double-stranded RNA from viruses and CpG DNA [3]. In addition to recognizing microbial PAMPs, TLRs also recognize endogenous ligands. Endogenous TLR ligands are molecules derived from host tissues or cells [11]. The TLRs can further be classified based on the conformation of the beta sheet. The beta sheet of *TLR3*, *TLR5*, *TLR7*, *TLR8* and *TLR9* has uniform twist angles and radii throughout the entire protein [9,12]. Conversely *TLR1*, *TLR2, TLR4*, *TLR6*

and *TLR10* have two structural transitions in their beta sheet and therefore their extracellular domains have three subdomains: N-terminal domain, central domain and C-terminal domain [9,13]. In vertebrates, the N-terminal domain determines PAMPs recognition specificity [14]. The border between the central domain and C-terminal (LRR motifs 9-12) of In *TLR1* and *TLR2* extracellular domains harbors ligand binding pockets on the convex side [15,16].

Some TLRs form heterodimers that bring their TIR domains together in order to recognize their ligands and to initiate signaling. *TLR1* and *TLR2* form a heterodimer to recognize triacylated lipopeptides whereas *TLR6* and *TLR2* heterodimerization recognize diacylated lipopeptides [17,18]. Similarly, *TLR4* forms a heterodimer with MD-2 to interact with lipopolysaccharide of Gram-negative bacteria [19,20]. *TLR2* in association with *TLR1* or *TLR6* recognize numerous PAMPs including mycoplasma lipopeptides, fungal zymosan, peptidoglycan, lipopeptides and lipoproteins of gram positive bacteria [21]. Heterodimerization therefore broadens the spectrum of ligands TLRs recognize. Jin et al. [12] have undertaken crystallographic studies of *TLR1-TLR2* and have proposed how TLR heterodimerization occurs. They indicated that ligand-induced dimerization brings the C termini of *TLR1* and *TLR2* close to each other, promoting the heterodimerization of their intracellular TIR domains and providing a scaffold for the recruitment of other proteins and initiation of signaling.

Origin and evolution of TLRs

It has been hypothesized that vertebrate TLRs arose as a result of an ancient gene duplication event that has subsequently given rise to two large families of TLRs. Indeed, phylogenetic analysis of vertebrate TLRs revealed two strongly supported clusters of TLRs with one clade containing the *TLR1* family (*TLR1*, *TLR2*, *TLR6* and *TLR10*) and another clade that included the remainder of mammalian TLRs [22]. These two groups arose prior to the divergence of protostomes (primitive invertebrates) and deuterostomes (erchinoderms and chordates) [22]. Subsequent evolution of these two gene families have been influenced by a complex history of gene duplication, gene conversion, positive selection and co-

evolution [23,24]. Within avian and mammalian genomes, genes of the *TLR1* gene family appear in tandem, suggesting successive rounds of tandem gene duplication from an ancestral gene [23]. Related TLR genes are also found in plants [25], indicating the ancient nature of host defense mechanism.

Vertebrate Toll-like receptors are so named due to their homology with the Toll receptor originally identified in *Drosophila melanogaster*. In *Drosophila melanogaster,* the Toll receptor is involved in dorsal-ventral patterning and antifungal and antibacterial immune response [26–28]. The first TLR to be identified as an orthologue of Drosophila Toll was *TLR4* [29,30] and the first hint of the pathogen recognition functions of TLRs was through the discovery that lipopolysaccharides responses were abolished in mice with spontaneous *TLR4* mutations and *TLR4*-knock out mice [31–33]. The similarity of the toll and TLRs has engendered much research interest. Two evolutionary models have been proposed as likely explanations for this similarity; a convergent evolution, where Toll and TLR signaling pathways evolved independently to adapt insects and vertebrates environments to similar environments [34] and a divergent evolution, where Toll and TLR signaling pathways have a common ancestor. Of the two models, the convergent evolution has been supported by several lines of evidence [35]. Convergent evolution has also been hypothesized as the force behind the similarity in the innate immune systems of plants and animals [36].

TLR signaling pathway

Signaling pathways are systems of proteins that act in an orchestrated fashion to modulate cell response to external and internal stimuli. These pathways have receptors that detect stimulus and trigger a cascade of events where each protein in the pathway changes the conformation (usually through phosphorylation and dephosphorylation) of the next protein down the pathway. The final effect of the signaling pathway is to trigger a response such as the activation of gene transcription.

TLR receptors detect PAMPs and trigger signaling pathways that results in increased expression of multiple inflammatory genes to eliminate the invading pathogens. In terms of the TLR signaling, two

pathways are distinguishable based on two master adaptors (Fig 1.3). The *MyD88* adaptor pathway (*MyD88* dependent pathway) is used by all TLRs except *TLR3* principally to induce the release of inflammatory cytokines while the *TRIF* adaptor dependent pathway (*MyD88* independent pathway) is used by *TLR3* and *TLR4* to induce the release of type I interferon production. *TLR4* therefore has the most complex signaling arrangement in comparison with all other TLRs.

MyD88 has a death domain and a TIR domain. Upon ligand binding, dimerization of TLRs occurs leading to interactions of the TLR TIR domains and a subsequent recruitment of *MyD88* through its TIR domain. Through interactions of the Death domains of both *MyD88* and *IRAK4*, *IRAK4* becomes activated. *IRAK4* phosphorylates *IRAK1*. Phosphorylated *IRAK1* in turn activates *TRAF6* which then undergoes ubiquitination resulting in the activation of NFκB. The activated NFκB leads to the expression of inflammatory cytokine genes [3]. Like *MyD88*, the *TRIF* adaptor in the *TRIF* dependent pathway has a TIR domain that is recruited by the *TLR3* and *TLR4*. *TRIF* then activates the downstream kinases, *TBK1* and *IKKε* leading to the phosphorylation and activation of *IRF3* [37]. Activated *IRF3* then controls the transcription of type I IFNS and IFN inducible genes [37]. Over activation of the TLR signaling can result in the onset of autoimmune and inflammatory disorders [38,39]. Thus, tight regulation of the TLR signaling is required. Among the numerous mechanisms to prevent aberrant induction of cytokines are degradation and sequestration of signaling molecules, transcription inhibition and inhibitory signals from certain receptors that antagonize PRR signaling [40].

Crosstalk between MyD88 dependent and independent pathways

Given that many TLRs recognize similar pathogenic products, it is conceivable that there is crosstalk between their signaling pathways: the *MyD88* dependent and independent pathways. Indeed, it has been observed that some *MyD88* dependent pathway genes (*Il1β*, *Cxcl1*, *Tnf*) were still induced in *MyD88* knock-out conditions [41], indicating a likely role of the *MyD88* independent pathway. The induction of the *Il1β*, *Cxcl1*, *Tnf* genes in *MyD88* knock-out conditions are as a result of crosstalk

between *TRIF* , a MyD88 independent pathway gene to *TRAF6*, a *MyD88* dependent pathway gene and between *TRIF* to *TAB/TAK* complex also belonging to the *MyD88* dependent pathway [42]. The interaction of *TRIF* and *TRAF6* to directly activate NFκB has also been demonstrated in other studies [43,44]. It has also been established that simultaneous stimulation of the *MyD88* dependent and *MyD88* independent pathways results in a greater immune response. For example, combinatorial stimulation of the *MyD88* dependent pathway with the synthetic ligand R848 and the *MyD88* independent pathway with Poly I:C induced higher cytokines [45]. Reduced transcript levels of *TLR1* and *TLR2* in the ileum have been reported for both *MyD88* and *TRIF* deficient mice, suggestive of a crosstalk between *TRIF* and MyD88 signaling pathways in the small intestine [46]. TLR signaling therefore involve genes interacting with each other in both sequential process within particular pathways and a network of crosstalk between pathways.

Single nucleotide polymorphisms within TLR signaling pathway genes

Single nucleotide polymorphisms (SNPs) refer to single allele mutations in the genomic sequence of an organism. Single base substitutions (nonsynonymous nucleotide polymorphisms) in genes can influence its expression or the function of the protein encoded by the gene [47–49]. Genetic variability in TLR molecules results in differences in susceptibility to infectious and inflammatory diseases across and within species [14]. Both synonymous and nonsynonymous polymorphisms have been identified in the promoter and coding regions of several TLRs and their association with infectious diseases have been documented [50]. In humans, the Asp299Gly amino acid substitution in *TLR4* is associated with a decrease in airway response to inhaled bacterial lipopolysaccharide [51] and with increased mortality in septic shock [52]. Polymorphisms in *TLR2* and *TLR4* pathways have been shown to regulate inflammatory response to bacterial ligands [53,54]. In wild rodents, an association between polymorphisms within *TLR2* and Borrelia infection has been reported [55]. A SNP of swine *TLR2* C406G is related to the prevalence of pneumonia [56]. Attempts to replicate findings on TLR polymorphisms and

their association with diseases have yielded variable results due to small sample sizes and the use of different populations. Apart from the TLRs, mutations within other genes in the TLR signaling pathway have been reported to result in susceptibility to diseases. A rare mutations affecting the MyD88 dependent pathway causes pyogenic bacteria diseases in childhood [57,58] and another rare mutation within the TRIF dependent pathway causes herpes simplex virus encephalitis [59,60]. Thus, genes within the TLR signaling pathway are under evolutionary pressure mediated by infectious disease pathogens.

Evolutionary pressure on genes

Natural selection pressures acting through climate, diets and pathogens have shaped genetic variation present in organisms. The response of organisms to such selective pressures results in three types of natural selection; positive selection which increases the frequency of favored alleles, purifying selection which eliminates detrimental alleles and balancing selection which results in diversity at a locus [61]. Advantageous genetic diversity at a locus can be maintained by overdominance (higher fitness of heterozygotes), frequency-dependent selection which refers to the situation where an allele's effect on fitness varies with its frequency, fluctuating selection where selection changes in time or space and pleiotropy where selection on a variant that affects multiple traits [62]. Balancing selection prevent fixation and maintain high genetic diversity levels within the species [63].

Positive selection has engendered more interest due to its adaptive value. Organisms carrying advantageous mutations are better adapted to their environment and tend to survive and reproduce. Positive selection has been used to identify genes putatively involved in species innovations and population adaptations [64–67], genes linked to disease [68–70], and sites within genes involved in antiviral or antibiotic resistance [71]. Positive selection can be detected for a whole protein coding gene, regions or codons within the gene using interspecies divergence data (variation representing substitutions between species) and intraspecific (within population) polymorphism data.

Codon based models of molecular evolution can be used to detect codons under positive selection in a sequence alignment from interspecies divergence data. An excess of nonsynonymous substitutions (dn) over synonymous substitutions (ds) at a codon sites indicates positive selection, suggesting replacement substitutions increase fitness [72]. Approaches based on population genetics have been used to infer positive selection at the intraspecies level. Geographically restricted positive selection has been investigated by comparing the degree of population differentiation at immune genes and putatively neutrally evolving loci. The theory here is that selection pressure that is geographically heterogeneous will result in increased differentiation at immune genes. High population differentiation, decreased polymorphisms between populations and expected linkage disequilibrium relative to neutral expectations constitutes some of the population genomic signatures of positive selection. Local adaptation at the molecular level occurs when alternative alleles are favored in different geographic locations. This process leads to low levels of nucleotide and allelic variation within population under directional selection [73] because frequency of alleles conferring adaptation increases.

Evolutionary pressure of infectious disease agents on TLRs

As pathogens evolve to subvert the host immune system, host immune genes also evolve in response. The arms race between hosts and microbial pathogens (host-parasite co-evolution) influences variation in the response to infectious disease agents at individuals, population, species levels and within higher order taxanomic units [74]. TLRs might be under positive selection due to co-evolutionary arms race with their microbial pathogens as they lie directly at the host-environment interface and target microbial molecules [75]. They may also be under purifying selection as they detect conserved molecular motifs. Previous studies using interspecies divergence data have inferred positive selection for codons within almost all TLRs investigated.

Studies across primates and across a wider range of species (Artiodactyla, rodents, primates, carnivores and Lagomorphs) have inferred positive selection for codons in viral (*TLR3*, *TLR7*, *TLR8* and

TLR9) and bacterial sensing TLRs (*TLR1*, *TLR2*, *TLR4* and *TLR6*) [75,76]. Positive selections have been inferred in *TLR1*LA, *TLR2*A, *TLR2*B, *TLR3, TLR4, TLR5* and *TLR15* across seven distantly related birds' species [77]. Recently, signatures of positive selection have been inferred in *TLR4* and *TLR7* across several rodent species [78] and several birds species in *TLR1*LA, *TLR2*A, *TLR2*B, *TLR3, TLR4, TLR5,TLR7, TLR15* and *TLR21* [79]. The following insights emerge from these previous studies. First, investigations on selection pressure acting on TLRs are an active area of research. Second, these studies suggest that positive selection has played a role in the evolution of most species TLRs. Third, the extracellular domain has more codons under positive selection than the transmembrane domain of TLRs. This is expected given that extracellular domains binds with PAMPs and the two may co-evolve while the transmembrane domain is involved in signaling and is therefore conserved. Furthermore, apart from the study of Areal et al. [76] which inferred a similar proportion of positively selected codons in both viral sensing and bacterial sensing TLRs, the other studies reported of more codons under positive selection in bacterial sensing than viral sensing TLRs, indicating that viral pathogens impose a greater selective constraint than their bacterial counterparts. For studies that used slightly different species within the same genera (Alcaide et al., 2011 and Grueber et al., 2014), some codon sites inferred to be under positive selection were different for both studies. These studies also used distantly related species. Studies using closely related species are required in order to gain new insights into how pathogens shape TLR evolution.

Studies at the population level have corroborated findings at the interspecies level that TLRs have evolved adaptively. For example, there is a strong evidence of recent positive selection in Europeans *TLR10-TLR1-TLR6* gene cluster with *TLR1* SNP 1805G (602S) as the real target of selection (Barreiro et al., 2009). Interestingly, this SNP results in impairment of NFκB activation leading to the avoidance of excessive TLR mediated inflammatory response [70], indicating how evolutionary analysis augment experimental studies in determining TLR polymorphisms of clinical relevance. The adaptor protein *MyD88* displays signatures of recent positive selection worldwide (human population in sub Saharan Africa, Europe and East Asia) whereas other adaptor proteins such as *TRIF* and *TRAM* showed evidence

of positive selection restricted to certain populations [80]. However in comparison with TLRs, the adaptor proteins are under stronger selective constraints [80], reflecting the different roles of the two groups of proteins; the TLRs being involved in pathogen recognition and the adaptor proteins participating in signaling in response to infections.

Further evidence of adaptive evolution of TLRs has been found in chicken populations. Similar to the study in humans [80], TLRs were inferred to have evolved adaptively in domestic chicken populations whereas cytokines, mediating molecules that initiate proinflammatory signals in the immune system in response to pathogens had signals suggestive of balancing selection [81]. The domestication process in chicken populations led to higher population densities creating a conducive environment for pathogen spread and this could have constituted an adaptive force on chicken TLRs [81]. Intraspecies analysis of three subspecies of chimpanzees indicated that positive selection has targeted a limited number of TLRs within specific subspecies [82]. The authors inferred that *TLR2* and *TLR4* showed signals of recent positive selection in *Pan troglodytes elliotii* of central Africa but not in *Pan troglodytes verus* of Western African, likely due to geographical differences in pathogen distribution. Comparison of pigs (wild boars and domestic pigs) from different geographic locations (Europe vs Asia) indicated that the snp *TLR4* C7485A was under positive selection [83]. However this study used a limited panel of 54 snps and only two TLR genes (*TLR2* and *TLR4* among 17 other immune related genes) and might have excluded other snps that are likely to be under positive selection. Analysis involving more TLR genes and an expanded panel of snps will be necessary to understand pathogen mediated TLR evolution in wild and domestic pigs. It is obvious that TLRs have been shown to evolve adaptively at both the interspecies and population levels as a result of pathogen-mediated selective pressures. However, given that the pathogens differ greatly from one host species to the other, new insights on TLR evolution can be gained from other species that have not yet been studied to a greater extent.

Protein regions targeted by infectious agents mediated selection pressure

There appears to be a bias in the region of a protein targeted by selection pressures. For instance, disease causing mutations are found in physical interface of protein-protein interactions [84]. Furthermore, positive selection pressure tend to target amino acids exposed to the surface of protein 3D structures whereas purifying selection pressure selects against amino acids buried in the 3D structure of protein because they may disrupt protein function [85–88]. Particular protein domains are also targeted by positive selective pressure. The N and C terminal domains of 3D protein structures appear to contain more positively selected sites [89]. Proteins have conformationally structured regions containing α-helices and β-sheets and intrinsically disordered regions (IDR) that are conformationally flexible. IDRs are polypeptide segments that are not likely to form well-defined 3D structures but are still functional [90,91]. IDRs generally tolerate increased genetic variation and have more positively selected residues than structured regions [92–94]. This has been largely attributed to lack of constraints on maintaining packing interactions that is responsible for purifying selection in structured sequences [95].

Post translational modifications (PTMs) are biochemical alterations of amino acids that extend the functional repertoires of protein [96]. PTMs affect the stability, interaction potential and localization of proteins within the cell [97]. Given these essential roles of PTMs, pathogens have developed strategies to interfere with host PTMs for their survival. Host PTMs are thus a target for pathogen mediated selection pressures. PTM regions are enriched in disordered sequences [96] and are thus likely to evolve adaptively. Indeed, strong signal for positive selection was evident in the N-linked glycosylation site of the envelope protein of the St. Louis encephalitis virus [98]. Knowledge of the protein regions targeted by selection pressure is essential in inferring the functional significance of amino acid residues in the absence of clinical or experimental studies.

Evolution of genes in the context of gene pathways and networks

Genes encode proteins that that do not act alone but operate as components of complex pathways and networks. For example, in the TLR signaling pathway, the receptors detect PAMPs that eventually trigger an immune response through interactions of proteins downstream of the signaling pathway. A problem in molecular evolution is to identify factors that influence the rate of protein evolution in the context of complex pathways and networks. The availability of whole genome sequences of numerous species has made such evolutionary studies possible. Taking into account the network within which a gene operates will provide insights into the evolutionary forces acting on it [99].

Researchers have investigated how position of genes within molecular pathways influences selective forces acting on these genes. There are conflicting results as to which genes (early functioning or downstream genes) within molecular pathways are under stronger selective constraints (relatively smaller dn/ds ratio). A hypothesis predicts that genes found early in pathways are subject to stronger selective constraint than downstream genes as mutations in early functioning genes are likely to have greater pleiotropic effects and affect all downstream phenotypes [100]. In keeping with this hypothesis, upstream genes evolved slower than downstream genes in the plant anthocyanin biosynthetic pathway [101]. This pattern is also observed in the plant carotenoid biosynthetic pathway enzymes where the most downstream enzyme is under the least constraint [102]. In contrast, studies on some signaling pathways indicate that downstream genes tend to be under stronger purifying selection than upstream genes. For instance, in the insulin/TOR signaling pathway in Drosophila and vertebrates, downstream genes evolve more slowly than their upstream counterparts [103]. A possible explanation for the relatively stronger purifying selection in downstream genes within signaling pathway is that downstream genes are located within a more stable cytoplasmic milieu whereas upstream genes encode receptors that lie in the host-environment interface and are therefore subject to adaptive evolutionary changes [104,105]. A detailed analysis of various signal transduction pathways in several organisms indicate differences in the way position of genes within networks impact evolutionary rate [104]: Situations where no relationship exist between gene position and the strength of purifying selection, upstream genes evolving faster and

upstream genes being more conserved have all been observed. Thus, it appears the relationship between positions of genes in a network varies depending partly on the function of the pathway and the species within which the pathway is being studied. A comprehensive understanding of the relationship will require analysis of different pathways and networks within different organisms.

Protein-protein interaction network and Evolutionary rate

The overall functions of molecular pathways and networks are dependent on the interactions among proteins constituting the pathways. Both protein-protein interactions (PPI) in terms of physical interaction between proteins and functional associations have been shown to influence protein evolutionary rate. It has been demonstrated that protein with more interactors evolves more slowly [106,107] because a greater proportion of the protein is involved in protein function. In contrast, [108] proteins in the center of networks have slower evolutionary rate, regardless of the number of PPI partners[108]. A conclusive study carried out in yeast indicated no correlation between connectivity and protein evolutionary rate. A negative correlation observed for some highthroughput datasets may have been due to artifacts of the data sets [109]. The lack of consensus on how the number of PPI affect protein evolution is further supported by a study [110] that confirmed that correlation between evolutionary rate and number of PPIs varies considerably across different protein interaction datasets. Furthermore, similarity in evolution rates for interacting proteins has been observed and this is attributed to their coevolution [106]; substitutions in one protein resulting in selective pressure for reciprocal changes in interacting proteins.

Biological features of PPI partners have also been implicated as having an influence on the evolutionary rates of proteins within a network. For example, proteins interacting with proteins of different function, and thus involved with multi different biological processes evolve slowly as compared with proteins that interact with proteins of the same function [111]. Makino and Gojobori [111] again showed that proteins in dense parts of PPI (forming a cluster) tend to evolve faster than those in sparse parts of PPI. They speculated that proteins in sparse parts of the PPI may be indispensable and therefore

under high selective constraints owing to likely scarceness of substitutable PPI partners. Considering the influence of the type of interactions among proteins on protein evolution, proteins involved in obligate (proteins never found out of complex with each other) interactions are more evolutionary constrained than those involved in transient interactions [112]. The development of highthroughput technology continues to make available large numbers of PPI within model organisms, providing an opportunity to further explore the effect of protein-protein interactions on protein evolution within several pathways and networks.

Network evolution and length of 3'UTR

The role of the length of the 3' untranslated regions (3'UTR) in affecting the evolution of proteins is mediated through microRNAs**.** MicroRNAs are ancient, short noncoding RNA molecules that regulate the transcriptome through post-transcriptional mechanisms. The 3'UTR region of a gene typically has binding sites for microRNAs influencing stability, localization and translation of messenger RNA [113]. Expression of genes encoding functionally important proteins is subject to robust regulation at the transcriptional and post-translational levels and therefore, genes under more intense regulation by miRNAs are expected to evolve at slower rates at the protein level [114]. A relationship between the length of the 3'UTR and protein evolutionary rate is expected as longer 3'UTRs will have more binding sites for miRNAs. Consistent with this expectation, a significant negative correlation between 3'UTR length and evolutionary rate for both humans and mice proteins have been inferred [114]. A study conducted for the *Drosophila* Toll-Imd signaling pathway [115] also reported a negative correlation between 3'UTR length and evolutionary rate.

Network evolution and codon bias

Codon bias refers to the phenomenon where synonymous codons are used with different frequencies in a variety of organisms [116]. Measures of codon bias usage include Effective Number of

Codons (ENC) and Codon Adaptation Index (CAI). ENC measures the magnitude of codon bias for an individual gene, with values ranging from 20 for a gene with extreme bias using one codon per amino acid, to 61 for a gene with no bias using synonymous codons equally [117]. The CAI estimates the extent of bias towards codons that are preferred in highly expressed genes with range of values 0 to 1, with 1 representing a stronger codon usage bias and a higher expression level [117]. The existence of codon bias can be explained by selection or mutation [116]. The selectionist explanation hypothesizes that codon bias is maintained by selection because it contributes to the efficiency and/or accuracy of protein expression. In contrast, the mutational explanation proposes that the existence of codon bias is due to some codons being more mutable done others (nonrandomness in the mutational patterns). An acceptable model for the existence of codon bias called mutation-selection-drift balance model of codon bias [118] proposes that selection favors preferred codons whereas mutation and genetic drift allow minor codons to exist. Codon bias has been shown to have negative correlation with protein evolution in unicellular organisms as diverse as *S. cerevisiae* [119,120] and multicellular organisms such as *Drosophila* species [103]. Gene expression, which is also known to influence evolutionary rate of proteins has been shown to be correlated with codon bias, and therefore there is the need to control for gene expression levels as a confounding factor.

A full understanding of the factors influencing protein evolution in molecular pathways or networks will require investigations across a broader range of organisms as different forces seem to drive evolution in different species. For example, transcriptional abundance (Codon adaptation index, gene expression level and protein abundance) is the most important factor influencing evolutionary rate in yeast [121] while protein-protein associations was an important contributor to protein evolution in bacteria [122]. Codon adaptation index (CAI) has been shown to be the most important factor influencing evolutionary rate in *E. coli* and *B. subtilis* [123]. Results of the influence of factors on evolutionary rate come with an important caveat. Factors influencing protein evolutionary rate are often interrelated. Thus, it is necessary to employ partial correlation and multivariate regression analysis in order to identify direct and indirect effects of these factors on protein evolution.

Suidae evolutionary history

The Suidae (pigs and their wild relatives) are one of the most successful families within the order of Artiodactyla [124]. The family Suidae consists of over 12 closely related species that have survived to date [125]. They are classified into six genera; *Babyrousa babyrusa* from South East Asia, *Porcula* from India, *Potamochoerus* (bush pig and red river hog), *Phacocherus* (common and desert warthogs) and *Hylochoerus* from sub-Saharan Africa and the *Sus* (domestic and wild pigs) from Eurasia [126]. Six out of the eight species of the genus *Sus* are endemic to Southeast Asia (SEA). Thus members of the Suidae have spread widely into different habitats.

Species within Suidae have evolved over a relative short time of 1-10 million years [125] and had their greatest adaptive radiation in the late Oligocene/early Miocene [127]. The Suidae radiation was characterized by relatively rapid speciation into new territories [128]. *Babyrousa babyrussa* stands well apart phylogenetically from the other members of the family Suidae and its relationship to other genera of the family Suidae is contentious. The *Babyrousa babyrussa* has been considered as an independent subfamily separate from the subfamily Suinae comprising the *Sus*, *Potamochoerus* and *Hylochoerus* [129] and also as part of the Suinae [130]. The genus *Sus* originated between 5.0 and 1.2 Mya [127,131] and differentiated into several lineages during the Late Pliocene and Early Pleistocene due to repeated connection and isolation of islands during sea level fluctuations [132]. Within the *Sus*, a deep split has been observed between *S.verrucosus* and other Island SEA *Sus* indicating that S. verrucosus represents a distinct lineage [133]. Analyses based on whole genome data have revealed that the speciation process of *Sus* from ISEA involved intra and inter specific gene flow and diversification [133].

Sus scrofa is the wild ancestor of the domestic pig. *Sus scrofa* diverged from *Sus* some 4 mya [133]. *Sus scrofa* originated from South-East Asia and dispersed into India and East Asia and moved westwards until they reached Europe [134]. This was followed by genetic isolation of the Eastern and Western gene pools occurring at approximately 1.2 mya [133]. The initial divergence between Asian and

European *Sus scrofa* may be the result of cooler climate during the Calabrian period that isolated populations in in small refugia across Eurasia [133]. The domestication process of *Sus scrofa* (wild boars) begun 10,000 years ago from Asia and Europe and introgression between wild pigs and domestic pigs might have taken place since the initial domestication event [135].

The origin of suids from Africa is contentious. Six subfamilies of the family Suidae colonized Africa from Eurasia at least six times during the Plio-Pleistocene [136]. On the basis of cranial and dental similarities, it has been proposed that some sub-Saharan African suids are more closely related to species from Eurasia [137], indicating a possible Eurasian origin of the African suids. However, this is inconsistent with the finding that the ancestors of the sub-Saharan suid genera (*Potamochoerus, Phacochoerus and Hylochoerus*) came from Africa and not Eurasia [138]. Attempts have been made to address this inconsistency using mitochondrial and nuclear DNA sequences of extant genera of Suidae from Eurasia [138]. The study concluded that the ancestors of the extant African suids evolved separately from the ancestors of modern day *Sus* and *Porcula* in Eurasia before colonizing Africa.

Phylogenetic analysis of species of the family Suidae indicates that the species form clusters consistent with their geographic distribution. For example, the phylogenetic tree of Suidae constructed using mitochondrial DNA cytochrome b sequences showed that African suids clustered in a single clade while Eurasian species formed another clade [139]. A similar observation has been made using mitochondrial and nuclear DNA where all sub-Saharan suids cluster in a monophyletic clade separate from Eurasian *Sus* species (Figure 1.4) [138]. Using near complete genome sequences, a well resolved tree for species within the genus *Sus* has been obtained (Figure 1.5) [133], where *Sus scrofa* form a cluster separate from the *Sus* restricted to Island and Mainland Southeast Asia.

Infectious disease challenges of Suidae

Suidae are a threat to human and domestic pig health as they serve as a reservoir for a number of infectious diseases. For example, Aujeszky's disease also known as pseudorabies is one of the economically important infectious diseases of swine for which suids are the natural host [140]. Due to

cross species transmission of viruses and bacteria, wild animals that have close phylogenetic relationships with domestic animals are at the greatest risk of infectious diseases mediated decline [141]. Indeed within the Artiodactyla, species in the families Bovidae and Suidae have been identified as more threatened by parasites than species in other families [141]. Wild boars and domestic pigs share many pathogens such as classical swine fever virus [142], African swine fever virus [143], Aujeszky's disease virus [144] and porcine reproductive and respiratory syndrome virus [145]. Among the host of naturally occurring diseases that affect both wild boars and domesticated pigs, diseases which have viruses and some bacteria as their causal agents have fatal consequences on the species either as juveniles or adults and include African swine fever, pseudorabies, Classical swine fever and foot and mouth diseases. Thus, infectious diseases have implications for the conservation and survival of populations of members of the Suidae.

Populations from different continents have historically been exposed to different pathogen mediated selective pressures [146]. For example, early cattle populations from Asia were exposed to rinderpest [147] and cattle populations from Sub-Saharan Africa have been exposed to trypanosomiasis [148]. Within the Suidae, different viral subpopulations in African and Eurasian Suidae host species have been reported. Two different lineages of ϒ1 endogenous retroviruses (ERVs), remnants of an exogenous viral form, corresponded to host phylogeny, one of Eurasian and another of African species have been observed [149]. Furthermore, *Phacochoerus africanus* (African warthog), *Potamochoerus larvatus* (bush pig) and *Hylochoerus meinertzhageni* (giant forest hog) are susceptible to African swine fever virus and other bacterial, viral and parasitic diseases of domestic swine [143,150]. Classical swine fever has become endemic in wild boar populations of eastern and Western Europe with spreading of the virus to domestic pigs [151].

Variation in tolerance for the same disease is well documented in members of the Suidae. For example, experimental infection of *Phacochoerus africanus* and *Potamochoerus larvatus* with classical swine fever virus resulted in subtle histological lesions in *Phacochoerus africanus* but overt clinical signs in *Potamochoerus larvatus* [152]. This variation in tolerance had also been demonstrated where virulent isolates of the African swine fever causes a rapidly fatal hemorrhagic fever in domestic pigs but species

(*Phacochoerus africanus*) endogenous to Africa tolerate the infection [153]. Differences at nonsynonymous amino acid sites in candidate genes have been implicated as influencing the level of tolerance among these species. Three amino acid differences in the candidate gene *RELA* (p65; v-rel reticuloendotheliosis viral oncogene homolog A) of warthog and domestic pigs causes reduced NF-κB activity in vitro for warthog RELA but not for domestic pig RELA.

In the absence of sufficient data on pathogens that have mediated selective pressure on species in the past, evolutionary analysis has been relied on to identify genes that may have been crucial in host survival. Evolutionary studies have proven to augment clinical studies in finding disease resistance genes [35] and also serve as a first step in characterizing genes for subsequent experimental work. The family Suidae has evolved overtime and into diverse environments where they likely encountered numerous challenges from infectious disease agents. However, knowledge of how pathogen mediated selection pressures have shaped the evolution of their immune related genes is scanty. Such knowledge is necessary in determining which genes and amino acid sites within these gene products have been important in adaptation of the Suidae members to past and present infections. Given that the innate immune system is the first line of defense against pathogens and TLRs have receptors for the recognition for almost all classes of pathogens ranging from bacteria, viruses, fungi and parasites [154], the family Suidae TLR signaling pathway was chosen for evolutionary analysis in this thesis work. Another criterion for choosing the TLR signaling pathway is that as compared to other PRRs where signaling is insufficient to mount an effective immune response in some cases, TLR mediated signaling is adequate and essential [155]. Evolutionary studies within the Suidae genomes have become possible due to recent developments in next generation sequencing technology, the availability of whole genome-sequence data for Suidae members at both the species and population (wild boars and domestic pigs) level and the availability of a high quality annotated sequence of the porcine (*Sus scrofa*) genome.

Thesis outline

Chapter 1 of this thesis, described early on is a general introduction covering review on immunity, TLRs in the context of their structure, origin, signaling and the evolutionary pressure imposed by pathogens on the TLRs. Also reviewed is the evolution of genes in the context of pathway and network parameters. Chapter 1 ends with the evolution of the Suidae and diseases that affect the health of members of the Suidae. In chapter 2, bacterial and viral sensing TLR sequences of 10 members of the family Suidae were obtained and the following questions were addressed using codon based models in a phylogenetic framework: 1) whether there is evidence of persistent positive selection at TLRs across members of the family Suidae and 2) to determine whether restricted lineages within the Suidae demonstrate TLR positive selection. Chapter 3 describes TLR adaptation within *Sus scrofa* (wild boar and domestic pig populations). Bacterial and viral sensing TLRs were obtained for wild boars and domestic pigs of European and Asian origins. Population genetics approaches were then used the determine 1) whether adaptive selection pressures on pigs from both European and Asian environments was evident and 2) whether different selective pressures for each geographic environment were identified. In chapter 4, the evolutionary rate of genes within the family Suidae TLR signaling pathway is investigated in the context of network parameters. Thirty three gene orthologs within the TLR signaling pathway of 10 members of the family Suidae were obtained. Codon based and multivariate analyses were then used to seek answers to the following questions 1) Is there a relationship between the strength of purifying selection and gene position in the TLR signaling pathway 2) Are there any network parameters that might be contributing to the polarity in the strength of purifying selection .

Figures and Table

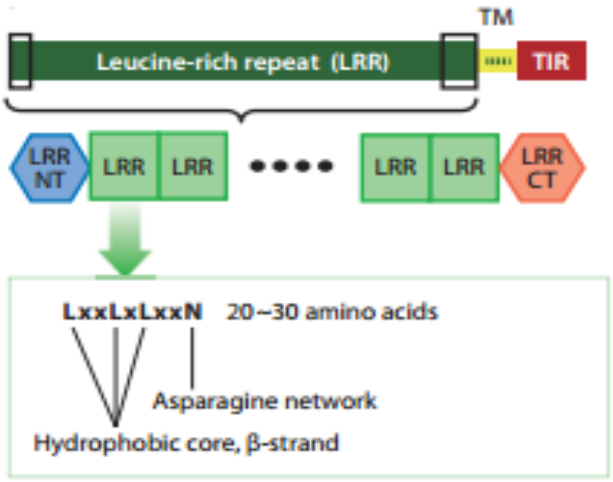

Figure 1.1. TLR domain organization. The leucine-rich repeat region is the extracellular domain required for detecting PAMPs. TM represents the transmembrane domain. TIR represents the Toll/interleukin receptor domain. The N and C termini of the LRR region is covered by LRRNT and LRRCT respectively (adapted from [8]).

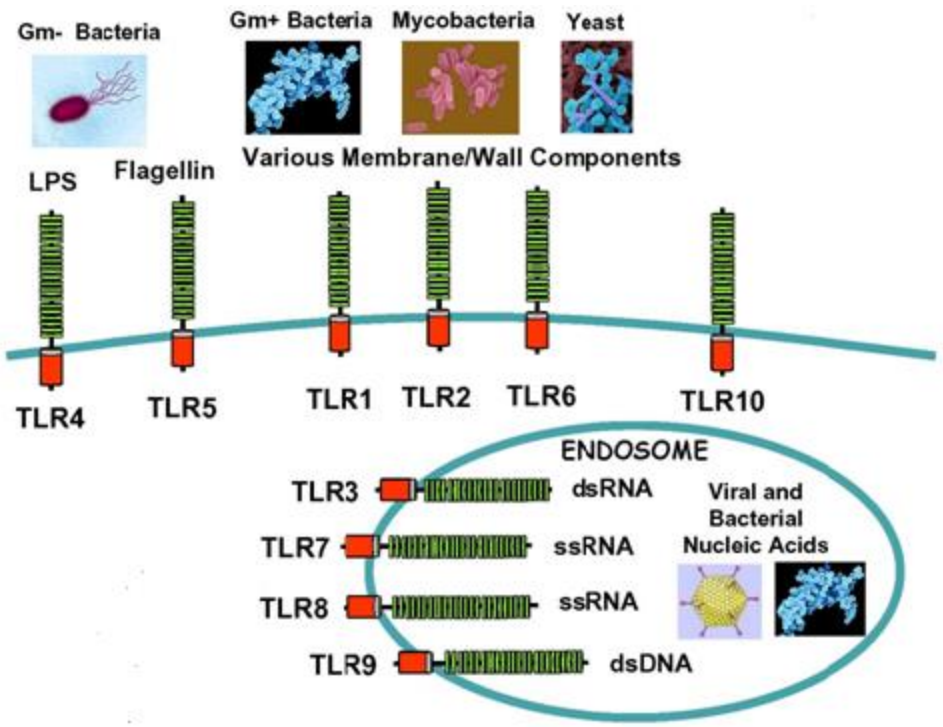

Figure 1.2. Cellular location of TLRs. TLR1, TLR2, TLR4, TLR5, TLR6 and TLR10 are located on the cell surface whereas TLR3, TLR7, TLR8 and TLR9 are located within the endosome. Also shown are the various TLR ligands (adapted from https://mcb.illinois.edu/faculty/profile/trapping).

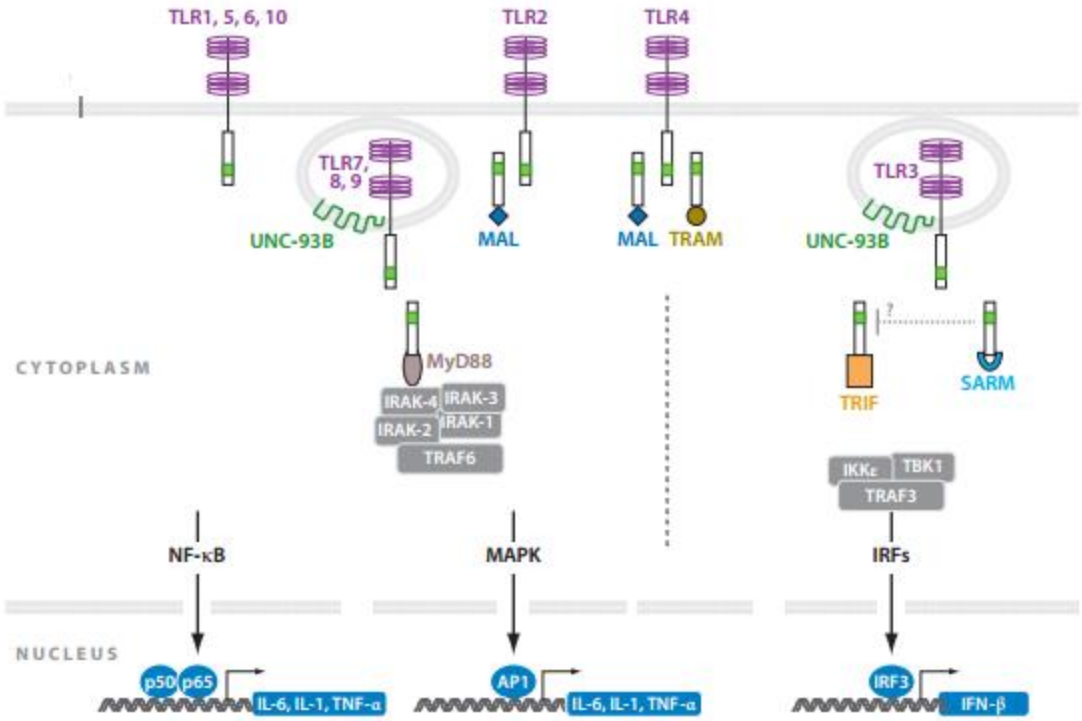

Figure 1.3. The TLR signaling pathway. *TLR4* and *TLR3* utilize the TRIF/MyD88 independent pathway. *TLR4*, *TLR1*, *TLR2*, *TLR5*, *TLR6*, *TLR7*, *TLR8*, *TLR9* and *TLR10* utilize the MyD88 dependent pathway of TLR signaling (adapted from [35]).

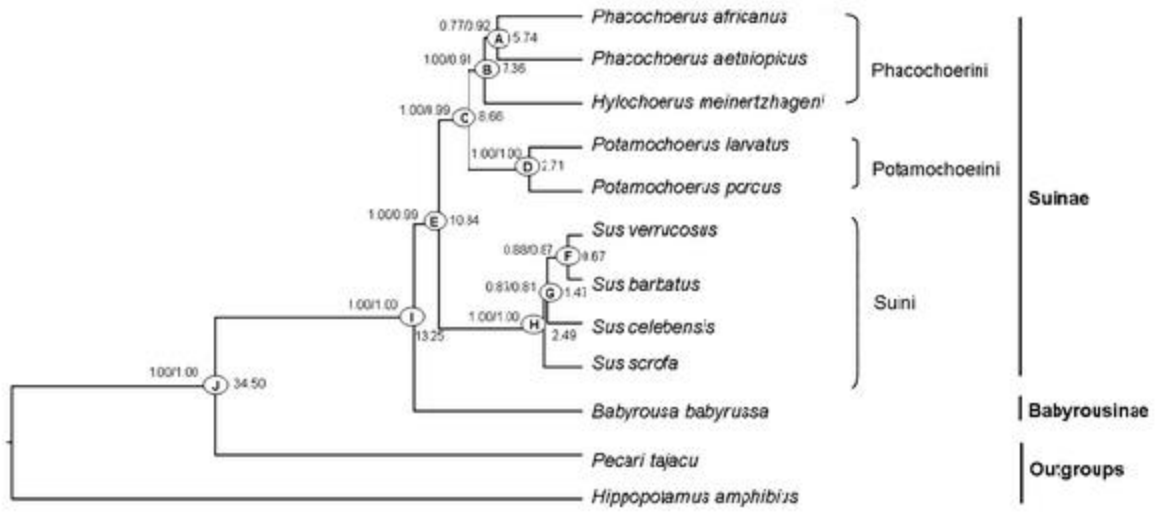

Figure 1.4. Relationship among species of the family Suidae based on a combined dataset of mitochondria and nuclear DNA sequences. Numbers after nodes represent divergence times and numbers on branches are the posterior probabilities. All species from sub-Saharan Africa (belong to the Phacochoerini and Potamochoerini) form a monophyletic clade and are separated from their Eurasian counterpart (the Suini) (adapted from [138]).

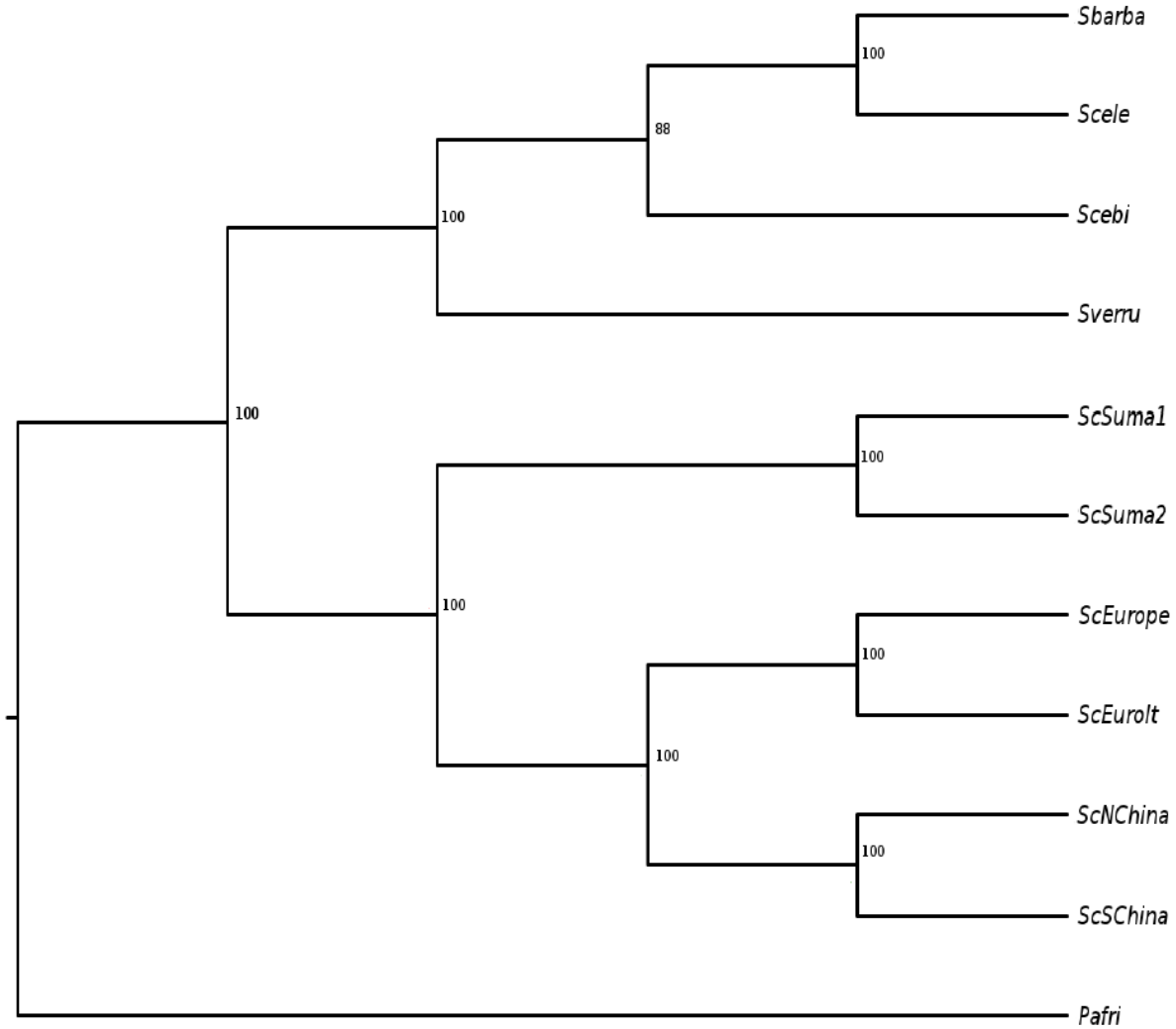

Figure 1.5. Relationship among species within the genus *Sus* based on near complete genomes. Numbers indicate support values. *S. scrofa* Sumatra; ScEuroIt = *S. scrofa* Italy; ScEurope = *S. scrofa* Europe; Sbarba = *S.barbatus*; Scebi = *S. celebensis*; Sverru= *S. verrucosus*; ScNChina = *S. scrofa* North China; ScSChina = *S. scrofa* South China. The relationship is consistent with the geographic distribution of the species; *Sus scrofa* is distributed across Eurasia and form a cluster separate from all other species of the genus *Sus* that are restricted to Island and mainland Southeast Asia (adapted from [133]).

Table 1.1

TLR	**Ligands**
TLR1	triacyl lipoproteins
TLR2	lipoprotein/lipopeptides, gram positive peptidoglycan, lipoteichoic acids, Zymosan (*Saccharomyces*), atypical lipopolysaccharides (*Leptospira interrogans*, *Porphyromonas gingivalis*), structural viral proteins (Herpes simplex virus, Cytomegalovirus)
TLR3	Double stranded RNA, polycytidic acid (poly I:C)
TLR4	Lipopolysaccharide (Gram-negative bacteria)
TLR5	flagellin
TLR6	diacyl lipoproteins
TLR7	Single stranded RNA
TLR8	Single stranded RNA
TLR9	Unmethylated CpG motifs (bacteria and viruses)
TLR10	Unknown

Table 1.1: Principal ligands of the Toll-like receptors

References

1. Hirano M, Das S, Guo P, Cooper MD. The evolution of adaptive immunity in vertebrates. Adv Immunol. 2011;109: 125–57. Available: http://www.ncbi.nlm.nih.gov/pubmed/21569914

2. Klein J, Sato A. The HLA system. First of two parts. N Engl J Med. 2000;343: 702–9. Available: http://www.ncbi.nlm.nih.gov/pubmed/10974135

3. Akira S, Uematsu S, Takeuchi O. Pathogen recognition and innate immunity. Cell. 2006;124: 783–801. Available: http://www.ncbi.nlm.nih.gov/pubmed/16497588

4. Janeway CA, Medzhitov R. Innate immune recognition. Annu Rev Immunol. 2002;20: 197–216. Available: http://www.ncbi.nlm.nih.gov/pubmed/11861602

5. Mogensen TH. Pathogen recognition and inflammatory signaling in innate immune defenses. Clin Microbiol Rev. 2009;22: 240–73, Table of Contents. Available: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=2668232&tool=pmcentrez&rendertype=abstract

6. Fukata M, Chen A, Klepper A, Krishnareddy S, Vamadevan AS, Thomas LS, et al. Cox-2 is regulated by Toll-like receptor-4 (TLR4) signaling: Role in proliferation and apoptosis in the intestine. Gastroenterology. 2006;131: 862–77. Available: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=2169292&tool=pmcentrez&rendertype=abstract

7. Rakoff-Nahoum S, Medzhitov R. Role of toll-like receptors in tissue repair and tumorigenesis. Biochem Biokhimiia. 2008;73: 555–61. Available: http://www.ncbi.nlm.nih.gov/pubmed/18605980

8. Kang JY, Lee J-O. Structural biology of the Toll-like receptor family. Annu Rev Biochem. 2011;80: 917–41. Available: http://www.ncbi.nlm.nih.gov/pubmed/21548780

9. Jin MS, Lee J-O. Structures of the toll-like receptor family and its ligand complexes. Immunity. 2008;29: 182–91. Available: http://www.ncbi.nlm.nih.gov/pubmed/18701082

10. Matsushima N, Tanaka T, Enkhbayar P, Mikami T, Taga M, Yamada K, et al. Comparative sequence analysis of leucine-rich repeats (LRRs) within vertebrate toll-like receptors. BMC Genomics. 2007;8: 124. Available: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=1899181&tool=pmcentrez&rendertype=abstract

11. Yu L, Wang L, Chen S. Endogenous toll-like receptor ligands and their biological significance. J Cell Mol Med. 2010;14: 2592–603. Available: http://www.ncbi.nlm.nih.gov/pubmed/20629986

12. Jin MS, Kim SE, Heo JY, Lee ME, Kim HM, Paik S-G, et al. Crystal structure of the TLR1-TLR2 heterodimer induced by binding of a tri-acylated lipopeptide. Cell. 2007;130: 1071–82. Available: http://www.ncbi.nlm.nih.gov/pubmed/17889651

13. Kang JY, Nan X, Jin MS, Youn S-J, Ryu YH, Mah S, et al. Recognition of lipopeptide patterns by Toll-like receptor 2-Toll-like receptor 6 heterodimer. Immunity. 2009;31: 873–84. Available: http://www.ncbi.nlm.nih.gov/pubmed/19931471

14. Werling D, Jann OC, Offord V, Glass EJ, Coffey TJ. Variation matters: TLR structure and species-specific pathogen recognition. Trends Immunol. 2009;30: 124–30. Available: http://www.ncbi.nlm.nih.gov/pubmed/19211304

15. Omueti KO, Beyer JM, Johnson CM, Lyle EA, Tapping RI. Domain exchange between human toll-like receptors 1 and 6 reveals a region required for lipopeptide discrimination. J Biol Chem. 2005;280: 36616–25. Available: http://www.ncbi.nlm.nih.gov/pubmed/16129684

16. Botos I, Segal DM, Davies DR. The structural biology of Toll-like receptors. Structure. 2011;19: 447–59. Available: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3075535&tool=pmcentrez&rendertype=abstract

17. Takeuchi O, Kawai T, Mühlradt PF, Morr M, Radolf JD, Zychlinsky A, et al. Discrimination of bacterial lipoproteins by Toll-like receptor 6. Int Immunol. 2001;13: 933–40. Available: http://www.ncbi.nlm.nih.gov/pubmed/11431423

18. Takeuchi O, Sato S, Horiuchi T, Hoshino K, Takeda K, Dong Z, et al. Cutting edge: role of Toll-like receptor 1 in mediating immune response to microbial lipoproteins. J Immunol. 2002;169: 10–4. Available: http://www.ncbi.nlm.nih.gov/pubmed/12077222

19. Ohto U, Fukase K, Miyake K, Satow Y. Crystal structures of human MD-2 and its complex with antiendotoxic lipid IVa. Science. 2007;316: 1632–4. Available: http://www.ncbi.nlm.nih.gov/pubmed/17569869

20. Park BS, Song DH, Kim HM, Choi B-S, Lee H, Lee J-O. The structural basis of lipopolysaccharide recognition by the TLR4-MD-2 complex. Nature. Macmillan Publishers Limited. All rights reserved; 2009;458: 1191–5. Available: http://dx.doi.org/10.1038/nature07830

21. Takeuchi O, Hoshino K, Kawai T, Sanjo H, Takada H, Ogawa T, et al. Differential roles of TLR2 and TLR4 in recognition of gram-negative and gram-positive bacterial cell wall components. Immunity. 1999;11: 443–51. Available: http://www.ncbi.nlm.nih.gov/pubmed/10549626

22. Hughes AL, Piontkivska H. Functional diversification of the toll-like receptor gene family. Immunogenetics. 2008;60: 249–56. Available: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=2431982&tool=pmcentrez&rendertype=abstract

23. Huang Y, Temperley ND, Ren L, Smith J, Li N, Burt DW. Molecular evolution of the vertebrate TLR1 gene family--a complex history of gene duplication, gene conversion, positive selection and co-evolution. BMC Evol Biol. 2011;11: 149. Available: http://www.biomedcentral.com/1471-2148/11/149

24. Roach JC, Glusman G, Rowen L, Kaur A, Purcell MK, Smith KD, et al. The evolution of vertebrate Toll-like receptors. Proc Natl Acad Sci U S A. 2005;102: 9577–82. Available: http://www.pnas.org/content/102/27/9577

25. Takeda K, Kaisho T, Akira S. Toll-like receptors. Annu Rev Immunol. 2003;21: 335–76. Available: http://www.ncbi.nlm.nih.gov/pubmed/12524386

26. Anderson K V, Schneider DS, Morisato D, Jin Y, Ferguson EL. Extracellular morphogens in Drosophila embryonic dorsal-ventral patterning. Cold Spring Harb Symp Quant Biol. 1992;57: 409–17. Available: http://www.ncbi.nlm.nih.gov/pubmed/1339676

27. Lemaitre B, Nicolas E, Michaut L, Reichhart JM, Hoffmann JA. The dorsoventral regulatory gene cassette spätzle/Toll/cactus controls the potent antifungal response in Drosophila adults. Cell. 1996;86: 973–83. Available: http://www.ncbi.nlm.nih.gov/pubmed/8808632

28. Williams MJ, Rodriguez A, Kimbrell DA, Eldon ED. The 18-wheeler mutation reveals complex antibacterial gene regulation in Drosophila host defense. EMBO J. 1997;16: 6120–30. Available: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=1326296&tool=pmcentrez&rendertype=abstract

29. Gay NJ, Keith FJ. Drosophila Toll and IL-1 receptor. Nature. 1991;351: 355–6. Available: http://www.ncbi.nlm.nih.gov/pubmed/1851964

30. Medzhitov R, Preston-Hurlburt P, Janeway CA. A human homologue of the Drosophila Toll protein signals activation of adaptive immunity. 1997;388: 394–397.

31. Poltorak A, He X, Smirnova I, Liu MY, Van Huffel C, Du X, et al. Defective LPS signaling in C3H/HeJ and C57BL/10ScCr mice: mutations in Tlr4 gene. Science. 1998;282: 2085–8. Available: http://www.ncbi.nlm.nih.gov/pubmed/9851930

32. Qureshi ST, Larivière L, Leveque G, Clermont S, Moore KJ, Gros P, et al. Endotoxin-tolerant mice have mutations in Toll-like receptor 4 (Tlr4). J Exp Med. 1999;189: 615–25. Available: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=2192941&tool=pmcentrez&rendertype=abstract

33. Hoshino K, Takeuchi O, Kawai T, Sanjo H, Ogawa T, Takeda Y, et al. Cutting edge: Toll-like receptor 4 (TLR4)-deficient mice are hyporesponsive to lipopolysaccharide: evidence for TLR4 as the Lps gene product. J Immunol. 1999;162: 3749–52. Available: http://www.ncbi.nlm.nih.gov/pubmed/10201887

34. Leulier F, Lemaitre B. Toll-like receptors--taking an evolutionary approach. Nat Rev Genet. 2008;9: 165–78. Available: http://www.ncbi.nlm.nih.gov/pubmed/18227810

35. Casanova J-L, Abel L, Quintana-Murci L. Human TLRs and IL-1Rs in host defense: natural insights from evolutionary, epidemiological, and clinical genetics. Annu Rev Immunol. 2011;29: 447–91. Available: http://www.ncbi.nlm.nih.gov/pubmed/21219179

36. Yue J-X, Meyers BC, Chen J-Q, Tian D, Yang S. Tracing the origin and evolutionary history of plant nucleotide-binding site-leucine-rich repeat (NBS-LRR) genes. New Phytol. 2012;193: 1049–63. Available: http://www.ncbi.nlm.nih.gov/pubmed/22212278

37. Fitzgerald KA, McWhirter SM, Faia KL, Rowe DC, Latz E, Golenbock DT, et al. IKKepsilon and TBK1 are essential components of the IRF3 signaling pathway. Nat Immunol. 2003;4: 491–6. Available: http://www.ncbi.nlm.nih.gov/pubmed/12692549

38. Drexler SK, Foxwell BM. The role of toll-like receptors in chronic inflammation. Int J Biochem Cell Biol. 2010;42: 506–18. Available: http://www.ncbi.nlm.nih.gov/pubmed/19837184

39. Midwood KS, Piccinini AM, Sacre S. Targeting Toll-like receptors in autoimmunity. Curr Drug Targets. 2009;10: 1139–55. Available: http://www.ncbi.nlm.nih.gov/pubmed/19715534

40. Kawai T, Akira S. TLR signaling. Cell Death Differ. 2006;13: 816–25. Available: http://www.ncbi.nlm.nih.gov/pubmed/16410796

41. Hirotani T, Yamamoto M, Kumagai Y, Uematsu S, Kawase I, Takeuchi O, et al. Regulation of lipopolysaccharide-inducible genes by MyD88 and Toll/IL-1 domain containing adaptor inducing IFN-beta. Biochem Biophys Res Commun. 2005;328: 383–92. Available: http://www.ncbi.nlm.nih.gov/pubmed/15694359

42. Selvarajoo K, Helmy M, Tomita M, Tsuchiya M. Inferring the mechanistic basis for the dynamic response of the MyD88-dependent and –independent pathways. International Conference on Molecular Systems Biology. 2008. Available: http://www.researchgate.net/publication/230823766_Inferring_the_mechanistic_basis_for_the_dynamic_response_of_the_MyD88-dependent_and_independent_pathways

43. Han K-J, Su X, Xu L-G, Bin L-H, Zhang J, Shu H-B. Mechanisms of the TRIF-induced interferon-stimulated response element and NF-kappaB activation and apoptosis pathways. J Biol Chem. 2004;279: 15652–61. Available: http://www.ncbi.nlm.nih.gov/pubmed/14739303

44. Xu L-G, Wang Y-Y, Han K-J, Li L-Y, Zhai Z, Shu H-B. VISA is an adapter protein required for virus-triggered IFN-beta signaling. Mol Cell. 2005;19: 727–40. Available: http://www.ncbi.nlm.nih.gov/pubmed/16153868

45. Suet Ting Tan R, Lin B, Liu Q, Tucker-Kellogg L, Ho B, Leung BPL, et al. The synergy in cytokine production through MyD88-TRIF pathways is co-ordinated with ERK phosphorylation in macrophages. Immunol Cell Biol. 2013;91: 377–87. Available: http://www.ncbi.nlm.nih.gov/pubmed/23567895

46. Hörmann N, Brandão I, Jäckel S, Ens N, Lillich M, Walter U, et al. Gut microbial colonization orchestrates TLR2 expression, signaling and epithelial proliferation in the small intestinal mucosa. PLoS One. 2014;9: e113080. Available: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=4232598&tool=pmcentrez&rendertype=abstract

47. Parter M, Kashtan N, Alon U. Facilitated variation: how evolution learns from past environments to generalize to new environments. PLoS Comput Biol. 2008;4: e1000206. Available: http://journals.plos.org/ploscompbiol/article?id=10.1371/journal.pcbi.1000206

48. Collins FS, Guyer MS, Charkravarti A. Variations on a theme: cataloging human DNA sequence variation. Science. 1997;278: 1580–1. Available: http://www.ncbi.nlm.nih.gov/pubmed/9411782

49. Risch N, Merikangas K. The future of genetic studies of complex human diseases. Science. 1996;273: 1516–7. Available: http://www.ncbi.nlm.nih.gov/pubmed/8801636

50. Medvedev AE. Toll-like receptor polymorphisms, inflammatory and infectious diseases, allergies, and cancer. J Interferon Cytokine Res. 2013;33: 467–84. Available: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3760066&tool=pmcentrez&rendertype=abstract

51. Arbour NC, Lorenz E, Schutte BC, Zabner J, Kline JN, Jones M, et al. TLR4 mutations are associated with endotoxin hyporesponsiveness in humans. Nat Genet. 2000;25: 187–91. Available: http://www.ncbi.nlm.nih.gov/pubmed/10835634

52. Schröder NWJ, Schumann RR. Single nucleotide polymorphisms of Toll-like receptors and susceptibility to infectious disease. Lancet Infect Dis. 2005;5: 156–64. Available: http://www.ncbi.nlm.nih.gov/pubmed/15766650

53. Shey MS, Randhawa AK, Bowmaker M, Smith E, Scriba TJ, de Kock M, et al. Single nucleotide polymorphisms in toll-like receptor 6 are associated with altered lipopeptide- and mycobacteria-induced interleukin-6 secretion. Genes Immun. 2010;11: 561–72. Available: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3518443&tool=pmcentrez&rendertype=abstract

54. Etokebe GE, Skjeldal F, Nilsen N, Rodionov D, Knezevic J, Bulat-Kardum L, et al. Toll-like receptor 2 (P631H) mutant impairs membrane internalization and is a dominant negative allele. Scand J Immunol. 2010;71: 369–81. Available: http://www.ncbi.nlm.nih.gov/pubmed/20500688

55. Tschirren B, Andersson M, Scherman K, Westerdahl H, Mittl PRE, Råberg L. Polymorphisms at the innate immune receptor TLR2 are associated with Borrelia infection in a wild rodent population. Proc Biol Sci. 2013;280: 20130364. Available: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3619520&tool=pmcentrez&rendertype=abstract

56. Muneta Y, Minagawa Y, Kusumoto M, Shinkai H, Uenishi H, Splichal I. Development of allele-specific primer PCR for a swine TLR2 SNP and comparison of the frequency among several pig breeds of Japan and the Czech Republic. J Vet Med Sci. 2012;74: 553–9. Available: http://www.ncbi.nlm.nih.gov/pubmed/22156387

57. Picard C, Puel A, Bonnet M, Ku C-L, Bustamante J, Yang K, et al. Pyogenic bacterial infections in humans with IRAK-4 deficiency. Science. 2003;299: 2076–9. Available: http://www.ncbi.nlm.nih.gov/pubmed/12637671

58. Von Bernuth H, Picard C, Jin Z, Pankla R, Xiao H, Ku C-L, et al. Pyogenic bacterial infections in humans with MyD88 deficiency. Science. 2008;321: 691–6. Available: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=2688396&tool=pmcentrez&rendertype=abstract

59. Casrouge A, Zhang S-Y, Eidenschenk C, Jouanguy E, Puel A, Yang K, et al. Herpes simplex virus encephalitis in human UNC-93B deficiency. Science. 2006;314: 308–12. Available: http://www.ncbi.nlm.nih.gov/pubmed/16973841

60. Zhang S-Y, Jouanguy E, Ugolini S, Smahi A, Elain G, Romero P, et al. TLR3 deficiency in patients with herpes simplex encephalitis. Science. 2007;317: 1522–7. Available: http://www.ncbi.nlm.nih.gov/pubmed/17872438

61. Karlsson EK, Kwiatkowski DP, Sabeti PC. Natural selection and infectious disease in human populations. Nat Rev Genet. Nature Publishing Group, a division of Macmillan Publishers Limited. All Rights Reserved.; 2014;15: 379–93. Available: http://dx.doi.org/10.1038/nrg3734

62. Andrés AM, Dennis MY, Kretzschmar WW, Cannons JL, Lee-Lin S-Q, Hurle B, et al. Balancing selection maintains a form of ERAP2 that undergoes nonsense-mediated decay and affects antigen presentation. PLoS Genet. 2010;6: e1001157. Available: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=2954825&tool=pmcentrez&rendertype=abstract

63. Ferrer-Admetlla A, Bosch E, Sikora M, Marquès-Bonet T, Ramírez-Soriano A, Muntasell A, et al. Balancing selection is the main force shaping the evolution of innate immunity genes. J Immunol. 2008;181: 1315–22. Available: http://www.ncbi.nlm.nih.gov/pubmed/18606686

64. Voight BF, Kudaravalli S, Wen X, Pritchard JK. A map of recent positive selection in the human genome. PLoS Biol. 2006;4: e72. Available: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=1382018&tool=pmcentrez&rendertype=abstract

65. Bakewell MA, Shi P, Zhang J. More genes underwent positive selection in chimpanzee evolution than in human evolution. Proc Natl Acad Sci U S A. 2007;104: 7489–94. Available: http://www.pnas.org/content/104/18/7489.abstract

66. Sabeti PC, Varilly P, Fry B, Lohmueller J, Hostetter E, Cotsapas C, et al. Genome-wide detection and characterization of positive selection in human populations. Nature. Nature Publishing Group; 2007;449: 913–8. Available: http://dx.doi.org/10.1038/nature06250

67. Kosiol C, Vinar T, da Fonseca RR, Hubisz MJ, Bustamante CD, Nielsen R, et al. Patterns of positive selection in six Mammalian genomes. PLoS Genet. 2008;4: e1000144. Available: http://journals.plos.org/plosgenetics/article?id=10.1371/journal.pgen.1000144

68. Rockman M V, Hahn MW, Soranzo N, Loisel DA, Goldstein DB, Wray GA. Positive selection on MMP3 regulation has shaped heart disease risk. Curr Biol. 2004;14: 1531–9. Available: http://www.ncbi.nlm.nih.gov/pubmed/15341739

69. Vamathevan JJ, Hasan S, Emes RD, Amrine-Madsen H, Rajagopalan D, Topp SD, et al. The role of positive selection in determining the molecular cause of species differences in disease. BMC Evol Biol. 2008;8: 273. Available: http://www.biomedcentral.com/1471-2148/8/273

70. Barreiro LB, Ben-Ali M, Quach H, Laval G, Patin E, Pickrell JK, et al. Evolutionary dynamics of human Toll-like receptors and their different contributions to host defense. PLoS Genet. 2009;5: e1000562. doi:10.1371/journal.pgen.1000562

71. Stanhope MJ, Lefébure T, Walsh SL, Becker JA, Lang P, Pavinski Bitar PD, et al. Positive selection in penicillin-binding proteins 1a, 2b, and 2x from Streptococcus pneumoniae and its

correlation with amoxicillin resistance development. Infect Genet Evol. 2008;8: 331–9. Available: http://www.ncbi.nlm.nih.gov/pubmed/18394970

72. Poon AFY, Frost SDW, Pond SLK. Detecting signatures of selection from DNA sequences using Datamonkey. Methods Mol Biol. 2009;537: 163–83. Available: http://www.ncbi.nlm.nih.gov/pubmed/19378144

73. Moeller DA, Tiffin P. Geographic variation in adaptation at the molecular level: a case study of plant immunity genes. Evolution. 2008;62: 3069–81. Available: http://www.ncbi.nlm.nih.gov/pubmed/18786191

74. Dawson HD, Loveland JE, Pascal G, Gilbert JGR, Uenishi H, Mann KM, et al. Structural and functional annotation of the porcine immunome. BMC Genomics. 2013;14: 332. Available: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3658956&tool=pmcentrez&rendertype=abstract

75. Wlasiuk G, Nachman MW. Adaptation and constraint at Toll-like receptors in primates. Mol Biol Evol. 2010;27: 2172–86. Available: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3107592&tool=pmcentrez&rendertype=abstract

76. Areal H, Abrantes J, Esteves PJ. Signatures of positive selection in Toll-like receptor (TLR) genes in mammals. BMC Evol Biol. 2011;11: 368. Available: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3276489&tool=pmcentrez&rendertype=abstract

77. Alcaide M, Edwards S V. Molecular evolution of the toll-like receptor multigene family in birds. Mol Biol Evol. 2011;28: 1703–15. Available: http://www.ncbi.nlm.nih.gov/pubmed/21239391

78. Fornůsková A, Vinkler M, Pagès M, Galan M, Jousselin E, Cerqueira F, et al. Contrasted evolutionary histories of two Toll-like receptors (Tlr4 and Tlr7) in wild rodents (MURINAE). BMC Evol Biol. 2013;13: 194. Available: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3848458&tool=pmcentrez&rendertype=abstract

79. Grueber CE, Wallis GP, Jamieson IG. Episodic positive selection in the evolution of avian toll-like receptor innate immunity genes. Ellegren H, editor. PLoS One. Public Library of Science; 2014;9: e89632. Available: http://dx.plos.org/10.1371/journal.pone.0089632

80. Fornarino S, Laval G, Barreiro LB, Manry J, Vasseur E, Quintana-Murci L. Evolution of the TIR domain-containing adaptors in humans: swinging between constraint and adaptation. Mol Biol Evol. 2011;28: 3087–97. Available: http://www.ncbi.nlm.nih.gov/pubmed/21659570

81. Downing T, Lloyd AT, O'Farrelly C, Bradley DG. The differential evolutionary dynamics of avian cytokine and TLR gene classes. J Immunol. 2010;184: 6993–7000. Available: http://www.ncbi.nlm.nih.gov/pubmed/20483729

82. Quach H, Wilson D, Laval G, Patin E, Manry J, Guibert J, et al. Different selective pressures shape the evolution of Toll-like receptors in human and African great ape populations. Hum Mol Genet. 2013;22: 4829–40. Available:

http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3820138&tool=pmcentrez&rendertype=abstract

83. Chen S, Gomes R, Costa V, Santos P, Charneca R, Zhang Y, et al. How immunogenetically different are domestic pigs from wild boars: a perspective from single-nucleotide polymorphisms of 19 immunity-related candidate genes. Immunogenetics. 2013;65: 737–48. Available: http://www.ncbi.nlm.nih.gov/pubmed/23846851

84. Wang X, Wei X, Thijssen B, Das J, Lipkin SM, Yu H. Three-dimensional reconstruction of protein networks provides insight into human genetic disease. Nat Biotechnol. 2012;30: 159–64. Available: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3708476&tool=pmcentrez&rendertype=abstract

85. Bustamante CD, Townsend JP, Hartl DL. Solvent accessibility and purifying selection within proteins of Escherichia coli and Salmonella enterica. Mol Biol Evol. 2000;17: 301–8. Available: http://www.ncbi.nlm.nih.gov/pubmed/10677853

86. Liu J, Zhang Y, Lei X, Zhang Z. Natural selection of protein structural and functional properties: a single nucleotide polymorphism perspective. Genome Biol. 2008;9: R69. Available: http://genomebiology.com/2008/9/4/R69

87. Petersen L, Bollback JP, Dimmic M, Hubisz M, Nielsen R. Genes under positive selection in Escherichia coli. Genome Res. 2007;17: 1336–43. Available: http://genome.cshlp.org/content/17/9/1336.abstract

88. Pentony MM, Winters P, Penfold-Brown D, Drew K, Narechania A, DeSalle R, et al. The plant proteome folding project: structure and positive selection in plant protein families. Genome Biol Evol. 2012;4: 360–71. Available: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3318447&tool=pmcentrez&rendertype=abstract

89. Ridout KE, Dixon CJ, Filatov DA. Positive selection differs between protein secondary structure elements in Drosophila. Genome Biol Evol. 2010;2: 166–79. Available: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=2997536&tool=pmcentrez&rendertype=abstract

90. Tompa P. Intrinsically disordered proteins: a 10-year recap. Trends Biochem Sci. 2012;37: 509–16. Available: http://www.ncbi.nlm.nih.gov/pubmed/22989858

91. Forman-Kay JD, Mittag T. From sequence and forces to structure, function, and evolution of intrinsically disordered proteins. Structure. 2013;21: 1492–9. Available: http://www.ncbi.nlm.nih.gov/pubmed/24010708

92. Nilsson J, Grahn M, Wright APH. Proteome-wide evidence for enhanced positive Darwinian selection within intrinsically disordered regions in proteins. Genome Biol. 2011;12: R65. Available: http://genomebiology.com/2011/12/7/R65

93. Schlessinger A, Schaefer C, Vicedo E, Schmidberger M, Punta M, Rost B. Protein disorder--a breakthrough invention of evolution? Curr Opin Struct Biol. 2011;21: 412–8. Available: http://www.ncbi.nlm.nih.gov/pubmed/21514145

94. Dunning LT, Dennis AB, Thomson G, Sinclair BJ, Newcomb RD, Buckley TR. Positive selection in glycolysis among Australasian stick insects. BMC Evol Biol. 2013;13: 215. Available: http://www.biomedcentral.com/1471-2148/13/215

95. Goldman N, Thorne JL, Jones DT. Assessing the impact of secondary structure and solvent accessibility on protein evolution. Genetics. 1998;149: 445–58. Available: http://www.genetics.org/content/149/1/445.abstract

96. Reimand J, Wagih O, Bader GD. Evolutionary Constraint and Disease Associations of Post-Translational Modification Sites in Human Genomes. PLoS Genet. 2015;11: e1004919. Available: http://journals.plos.org/plosgenetics/article?id=10.1371/journal.pgen.1004919

97. Seet BT, Dikic I, Zhou M-M, Pawson T. Reading protein modifications with interaction domains. Nat Rev Mol Cell Biol. 2006;7: 473–83. Available: http://www.ncbi.nlm.nih.gov/pubmed/16829979

98. Baillie GJ, Kolokotronis S-O, Waltari E, Maffei JG, Kramer LD, Perkins SL. Phylogenetic and evolutionary analyses of St. Louis encephalitis virus genomes. Mol Phylogenet Evol. 2008;47: 717–28. Available: http://www.ncbi.nlm.nih.gov/pubmed/18374605

99. Alvarez-Ponce D. The relationship between the hierarchical position of proteins in the human signal transduction network and their rate of evolution. BMC Evol Biol. 2012;12: 192. Available: http://www.biomedcentral.com/1471-2148/12/192

100. Cork JM, Purugganan MD. The evolution of molecular genetic pathways and networks. Bioessays. 2004;26: 479–84. Available: http://www.ncbi.nlm.nih.gov/pubmed/15112228

101. Rausher MD, Lu Y, Meyer K. Variation in constraint versus positive selection as an explanation for evolutionary rate variation among anthocyanin genes. J Mol Evol. 2008;67: 137–44. Available: http://www.ncbi.nlm.nih.gov/pubmed/18654810

102. Livingstone K, Anderson S. Patterns of variation in the evolution of carotenoid biosynthetic pathway enzymes of higher plants. J Hered. 100: 754–61. Available: http://www.ncbi.nlm.nih.gov/pubmed/19520763

103. Alvarez-Ponce D, Aguadé M, Rozas J. Network-level molecular evolutionary analysis of the insulin/TOR signal transduction pathway across 12 Drosophila genomes. Genome Res. 2009;19: 234–42. Available: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=2652205&tool=pmcentrez&rendertype=abstract

104. Jovelin R, Phillips PC. Expression level drives the pattern of selective constraints along the insulin/Tor signal transduction pathway in Caenorhabditis. Genome Biol Evol. 2011;3: 715–22. Available: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3157841&tool=pmcentrez&rendertype=abstract

105. Song X, Jin P, Qin S, Chen L, Ma F. The evolution and origin of animal Toll-like receptor signaling pathway revealed by network-level molecular evolutionary analyses. Wicker-Thomas C, editor. PLoS One. Public Library of Science; 2012;7: e51657. Available: http://dx.plos.org/10.1371/journal.pone.0051657

106. Fraser HB, Hirsh AE, Steinmetz LM, Scharfe C, Feldman MW. Evolutionary rate in the protein interaction network. Science. 2002;296: 750–2. Available: http://www.ncbi.nlm.nih.gov/pubmed/11976460

107. Fraser HB, Wall DP, Hirsh AE. A simple dependence between protein evolution rate and the number of protein-protein interactions. BMC Evol Biol. 2003;3: 11. Available: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=166126&tool=pmcentrez&rendertype=abstract

108. Hahn MW, Kern AD. Comparative genomics of centrality and essentiality in three eukaryotic protein-interaction networks. Mol Biol Evol. 2005;22: 803–6. Available: http://www.ncbi.nlm.nih.gov/pubmed/15616139

109. Batada NN, Hurst LD, Tyers M. Evolutionary and physiological importance of hub proteins. Russell R, editor. PLoS Comput Biol. Public Library of Science; 2006;2: e88. Available: http://dx.plos.org/10.1371/journal.pcbi.0020088

110. Pang K, Cheng C, Xuan Z, Sheng H, Ma X. Understanding protein evolutionary rate by integrating gene co-expression with protein interactions. BMC Syst Biol. 2010;4: 179. Available: http://www.biomedcentral.com/1752-0509/4/179

111. Makino T, Gojobori T. The evolutionary rate of a protein is influenced by features of the interacting partners. Mol Biol Evol. 2006;23: 784–9. Available: http://www.ncbi.nlm.nih.gov/pubmed/16407461

112. Teichmann SA. The constraints protein-protein interactions place on sequence divergence. J Mol Biol. 2002;324: 399–407. Available: http://www.ncbi.nlm.nih.gov/pubmed/12445777

113. Hughes TA. Regulation of gene expression by alternative untranslated regions. Trends Genet. 2006;22: 119–22. Available: http://www.ncbi.nlm.nih.gov/pubmed/16430990

114. Cheng C, Bhardwaj N, Gerstein M. The relationship between the evolution of microRNA targets and the length of their UTRs. BMC Genomics. 2009;10: 431. Available: http://www.biomedcentral.com/1471-2164/10/431

115. Han M, Qin S, Song X, Li Y, Jin P, Chen L, et al. Evolutionary rate patterns of genes involved in the Drosophila Toll and Imd signaling pathway. BMC Evol Biol. 2013;13: 245. Available: http://www.biomedcentral.com/1471-2148/13/245

116. Hershberg R, Petrov DA. Selection on codon bias. Annu Rev Genet. 2008;42: 287–99. Available: http://www.ncbi.nlm.nih.gov/pubmed/18983258

117. Zhao S, Zhang Q, Chen Z, Zhao Y, Zhong J. The factors shaping synonymous codon usage in the genome of Burkholderia mallei. J Genet Genomics. 2007;34: 362–72. Available: http://www.ncbi.nlm.nih.gov/pubmed/17498635

118. Bulmer M. The selection-mutation-drift theory of synonymous codon usage. Genetics. 1991;129: 897–907. Available: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=1204756&tool=pmcentrez&rendertype=abstract

119. Coghlan A, Wolfe KH. Relationship of codon bias to mRNA concentration and protein length in Saccharomyces cerevisiae. Yeast. 2000;16: 1131–45. Available: http://www.ncbi.nlm.nih.gov/pubmed/10953085

120. Sharp PM. Determinants of DNA sequence divergence between Escherichia coli and Salmonella typhimurium: codon usage, map position, and concerted evolution. J Mol Evol. 1991;33: 23–33. Available: http://www.ncbi.nlm.nih.gov/pubmed/1909371

121. Drummond DA, Raval A, Wilke CO. A single determinant dominates the rate of yeast protein evolution. Mol Biol Evol. 2006;23: 327–37. Available: http://mbe.oxfordjournals.org/content/23/2/327

122. Wei W, Zhang T, Lin D, Yang Z-J, Guo F-B. Transcriptional abundance is not the single force driving the evolution of bacterial proteins. BMC Evol Biol. 2013;13: 162. Available: http://www.biomedcentral.com/1471-2148/13/162

123. Rocha EPC, Danchin A. An analysis of determinants of amino acids substitution rates in bacterial proteins. Mol Biol Evol. 2004;21: 108–16. Available: http://www.ncbi.nlm.nih.gov/pubmed/14595100

124. Hardjasasmita HS. Taxonomy and phylogeny of the Suidae (Mammalia) in Indonesia /. Leiden : Rijksmuseum van Geologie en Mineralogie,; 1987.

125. Groenen MAM, Megens H-J, Semiadi G, Schook LB. Insights from whole genome sequence alignment of the domestic (Sus scrofa) and Javan Warty (Sus verrucosus) pigs. Plant and Animal Genomes XVIII Conference. San Diego, CA, USA; 2010.

126. Grubb P. Order Artiodactyla. In: Wilson DE, Reeder DM, editors. Mammal species of the world: a taxanomic and geographic reference. 3rd ed. Baltimore, Maryland: Johns Hopkins University Press; 2005. pp. 637–772.

127. Randi E, Lucchini V, Diong CH. Evolutionary genetics of the suiformes as reconstructed using mtDNA sequencing. J Mamm Evol. 1996;3: 163–194. Available: http://link.springer.com/10.1007/BF01454360

128. Ducrocq S, Chaimanee Y, Suteethorn V, Jaeger J-J. The earliest known pig from the upper Eocene of Thailand. Palaeontology. 1998;41: 147–156.

129. Groves CP, Grubb P. The suborder Suiformes. In: Oliver WLR, editor. Pigs, Peccaries and Hippos. Gland, Switzerland: IUCN/SSC; 1993. pp. 1–4.

130. Harris J, Liu L-P. Superfamily Suoidea. In: Prothero DR, Foss S, editors. The evolution of Artiodactyla. Baltimore, Maryland: John Hopkins University Press; 2007. pp. 130–150.

131. Mona S, Randi E, Tommaseo-Ponzetta M. Evolutionary history of the genus Sus inferred from cytochrome b sequences. Mol Phylogenet Evol. 2007;45: 757–62. Available: http://www.ncbi.nlm.nih.gov/pubmed/17629713

132. Lucchini V, Meijaard E, Diong CH, Groves CP, Randi E. New phylogenetic perspectives among species of South-east Asian wild pig (Sus sp.) based on mtDNA sequences and morphometric data. J Zool. Cambridge University Press; 2005;266: 25–35. Available: http://journals.cambridge.org/abstract_S0952836905006588

133. Frantz LAF, Schraiber JG, Madsen O, Megens H-J, Bosse M, Paudel Y, et al. Genome sequencing reveals fine scale diversification and reticulation history during speciation in Sus. Genome Biol. 2013;14: R107. Available: http://genomebiology.com/2013/14/9/R107

134. Larson G, Dobney K, Albarella U, Fang M, Matisoo-Smith E, Robins J, et al. Worldwide phylogeography of wild boar reveals multiple centers of pig domestication. Science. 2005;307: 1618–21. Available: http://www.ncbi.nlm.nih.gov/pubmed/15761152

135. Larson G, Albarella U, Dobney K, Rowley-Conwy P, Schibler J, Tresset A, et al. Ancient DNA, pig domestication, and the spread of the Neolithic into Europe. Proc Natl Acad Sci U S A. 2007;104: 15276–81. Available: http://www.pnas.org/content/104/39/15276.abstract

136. Pickford M. Synopsis of the biochronology of African Neogene and Quaternary Suiformes. Trans R Soc South Africa. Taylor & Francis Group; 2006;61: 51–62. Available: http://www.tandfonline.com/doi/abs/10.1080/00359190609519953#.VN4m8vnF9zg

137. Cooke HB. Suid evolution and correlation of African hominid localities: an alternative taxonomy. Science. 1978;201: 460–3. Available: http://www.ncbi.nlm.nih.gov/pubmed/96530

138. Gongora J, Cuddahee RE, Nascimento FF do, Palgrave CJ, Lowden S, Ho SYW, et al. Rethinking the evolution of extant sub-Saharan African suids (Suidae, Artiodactyla). Zool Scr. 2011;40: 327–335. Available: http://doi.wiley.com/10.1111/j.1463-6409.2011.00480.x

139. WU Gui-sheng PJZY. Molecular Phylogeny and Phylogeography of Suidae. Zool Res. Kunming Institute of Zoology, Chinese Academy of Sciences; 27: 197–201. Available: http://www.bioline.org.br/request?zr06027

140. Müller T, Hahn EC, Tottewitz F, Kramer M, Klupp BG, Mettenleiter TC, et al. Pseudorabies virus in wild swine: a global perspective. Arch Virol. 2011;156: 1691–705. Available: http://www.ncbi.nlm.nih.gov/pubmed/21837416

141. Pedersen A, Jones K, Nunn C, Altizer S. Infectious diseases and extinction risk in wild mammals [Internet]. CONSERVATION BIOLOGY. 2007. Available: http://discovery.ucl.ac.uk/1347949/

142. Paton D., Greiser-Wilke I. Classical swine fever – an update. Res Vet Sci. 2003;75: 169–178. Available: http://www.sciencedirect.com/science/article/pii/S0034528803000766

143. Jori F, Bastos ADS. Role of wild suids in the epidemiology of African swine fever. Ecohealth. 2009;6: 296–310. Available: http://www.ncbi.nlm.nih.gov/pubmed/19915917

144. Vicente J, Segalés J, Höfle U, Balasch M, Plana-Durán J, Domingo M, et al. Epidemiological study on porcine circovirus type 2 (PCV2) infection in the European wild boar (Sus scrofa). Vet Res. 35: 243–53. Available: http://www.ncbi.nlm.nih.gov/pubmed/15099500

145. Reiner G, Fresen C, Bronnert S, Willems H. Porcine Reproductive and Respiratory Syndrome Virus (PRRSV) infection in wild boars. Vet Microbiol. 2009;136: 250–8. Available: http://www.ncbi.nlm.nih.gov/pubmed/19131188

146. Novembre J, Di Rienzo A. Spatial patterns of variation due to natural selection in humans. Nat Rev Genet. Nature Publishing Group; 2009;10: 745–55. Available: http://dx.doi.org/10.1038/nrg2632

147. Spinage CA. Cattle plague: A history. New York, NY: Kluwer Academic; 2003.

148. Gifford-Gonzalez D. Animal Disease Challenges to the Emergence of Pastoralism in Sub-Saharan Africa. Diane Gifford-Gonzalez 2000 Anim Dis Challenges to Emerg Pastor Sub-Saharan Africa African Archaeol Rev 17 95-139. 2000;3: 95–139. Available: https://www.infona.pl/resource/bwmeta1.element.springer-f6bdf57b-b711-30d1-8209-442265e8153f

149. Nascimento FF, Gongora J, Charleston M, Tristem M, Lowden S, Moran C. Evolution of endogenous retroviruses in the Suidae: evidence for different viral subpopulations in African and Eurasian host species. BMC Evol Biol. 2011;11: 139. Available: http://www.biomedcentral.com/1471-2148/11/139

150. Fowler ME. Husbandry and diseases of captive wild swine and peccaries. Rev Sci Tech. 1996;15: 141–54. Available: http://www.ncbi.nlm.nih.gov/pubmed/8924700

151. Laddomada A. Incidence and control of CSF in wild boar in Europe. Vet Microbiol. 2000;73: 121–30. Available: http://www.ncbi.nlm.nih.gov/pubmed/10785322

152. Gers S, Vosloo W, Drew T, Lubisi AB, Pardini A, Williams M. Experimental Infection of Common Warthogs (Phacochoerus africanus) and Bushpigs (Potamochoerus larvatus) with Classical Swine Fever Virus II: A Comparative Histopathological Study. Transbound Emerg Dis. 2010; Available: http://www.ncbi.nlm.nih.gov/pubmed/21176120

153. Palgrave CJ, Gilmour L, Lowden CS, Lillico SG, Mellencamp MA, Whitelaw CBA. Species-Specific Variation in RELA Underlies Differences in NF- B Activity: a Potential Role in African Swine Fever Pathogenesis. J Virol. 2011;85: 6008–6014. Available: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3126315&tool=pmcentrez&rendertype=abstract

154. Ishii KJ, Koyama S, Nakagawa A, Coban C, Akira S. Host innate immune receptors and beyond: making sense of microbial infections. Cell Host Microbe. 2008;3: 352–63. Available: http://www.ncbi.nlm.nih.gov/pubmed/18541212

155. Tan RST, Ho B, Leung BP, Ding JL. TLR cross-talk confers specificity to innate immunity. Int Rev Immunol. 33: 443–53. Available: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=4266099&tool=pmcentrez&rendertype=abstract

Chapter 2: Adaptive evolution of Toll-like receptors (TLRs) in the family Suidae[1]

Abstract

Members of the family Suidae have diverged over extended evolutionary periods in diverse environments, suggesting that adaptation in response to endemic infectious agents may have occurred. Toll-like receptors (TLRs) comprise a multigene family that acts as the first line of defense against infectious microbes at the host-environment interface. We hypothesized that across the Suidae, positive selection mediated by infectious agents has contributed to the evolution of TLR diversity. Thus, we analyzed *Sus scrofa*, *Sus barbatus*, *Sus verrucosus*, *Sus celebensis*, *Sus scebifrons*, *Babyrousa babyrussa*, *Potamochoerus larvatus*, *Potamochoerus porcus* and *Phacochoerus africanus* genomes. Specifically, analyses were performed to identify evidence of positive selection using Maximum likelihood (ML) methods within a phylogenetic framework for bacterial and viral sensing Suidae TLR extracellular domains. Our analyses did not reveal evidence of positive selection for *TLR3* and *TLR7*, suggesting strong functional conservation among these two genes for members of the Suidae. Positive selection was inferred for Suidae *TLR1*, *TLR2*, *TLR6* and *TLR8* evolution. ML methods identified amino acid sites of the bacterial sensing *TLR1*, *TLR2*, *TLR6* and the viral sensing *TLR8* to be under persistent positive selection. Some of these sites are in close proximity to functionally relevant sites, further strengthening the case for pathogen mediated selection for these sites. The branch leading to the genus *Sus* demonstrated evidence of episodic positive selection for *TLR1*, indicating selection mediated by infectious agents encountered within the specific geographic origin of the *Sus.* These results indicate that species of the Suidae have positively selected residues within functional domains of TLRs reflective of prior infections. Thus, TLR genes represent candidates for experimental validation to determine their functional role in antibacterial and antiviral activity within members of the Suidae.

---

[1]This chapter has been accepted for publication in PLoS One

Introduction

Bacterial and viral infectious diseases constitute a significant threat to host survival. Host species have developed various strategies to combat these threats, including the development of innate and acquired immune defenses. The innate immune system provides an immediate defensive response against pathogenic infections while the acquired immune system response to pathogenic infections may require weeks to develop. As pathogens evolve to subvert the host immune response, host immune genes evolve in response. The arms race between hosts and microbial pathogens (host-parasite co-evolution) influences variation in the response to infectious disease agents at individuals, population, species levels and within higher order taxonomic units [1].

Adaptive evolution (positive selective) is selective pressure through a change in environment placed on a protein in order to improve the fitness of the organism in that environment [2]. With respect to vertebrate immune-related genes, studies on adaptive evolution have mainly focused on the major histocompatibility complex (MHC), cell surface glycoproteins of the acquired immune system that mediate presentation of peptides to T-cell receptors [3]. In humans, it has been shown that half of the genetic variability in immune response to infections is accounted for by non-MHC genes [4]. Most of these non-MHC genes seem to belong to the innate immune system [5], indicating that such genes may be under adaptive evolution. Phagocytic cells such as monocytes, macrophages and dendritic cells mediate the recognition of pathogens by the innate immune system through germline encoded receptors known as pattern recognition receptors (PRRs). These PRRs recognize conserved molecular features of microbes called pathogen-associated molecular patterns (PAMPs) [6,7]. Among the numerous PRRs, the Toll-like receptor family is the most widely studied. The Toll-like receptors (TLRs) are innate immunity receptors important during early phase of infections and also serve as a link between the innate and acquired immunity during host immune response [8]. Cell surface expressed TLRs (*TLR1*, *TLR2*, *TLR4*, *TLR5* and *TLR6*) recognize predominantly bacterial ligands and several fungal and parasite ligands while *TLR3*, *TLR7* and *TLR8* are expressed within the endosome and recognize single and double-stranded viral RNA [9]. TLRs are type I transmembrane glycoproteins composed of an extracellular domain characterized by

a leucine-rich repeat (LRR) motif responsible for binding infectious agents ligands, a transmembrane domain and an intracellular signaling domain.

Previous studies have documented purifying selection [10] and overdominant balancing selection [11] as the dominant selective pressures acting within innate immune genes including TLRs. TLRs might be under positive selection due to a co-evolutionary arms race with their microbial pathogens as they lie directly at the host-environment interface and target microbial molecules [12]. Studies at the interspecies level have found clear signatures of positive selection at codon positions across TLR genes from primate, avian and murinae species [12–15]. In the context of positive selection at the interspecies level, a distinction can be made between persistent positive selection, where selective pressure at codon positions within a gene remains constant throughout time across species and episodic positive selection where selective pressures act in a lineage specific manner [16]. In the case of persistent positive selection, the selective pressure affects most lineages within a phylogeny and is evident as codons rapidly evolving across the species in a phylogenetic tree. For episodic positive selection, codon positions under positive selective pressure within particular lineages may be neutrally or negatively evolving in other lineages. Regardless of the type of selective pressure (persistent or episodic positive selection), detection of evidence of selection of a gene region suggests a selective advantage in changing amino acid sequence in this region [2].

Members of the family Suidae have a widespread distribution. The natural occurrence of *Sus scrofa* (wild boar) is across most of Eurasia while all other species of the genus *Sus* are restricted to Southeast Asia [17]. The *Babyrussa babyrussa* (babyrussa) is also found in Southeast Asia and *Potamochoerus larvatus* (bush pig), *Potamochoerus porcus* (red river hog) and *Phacochoerus africanus* (common warthog) are restricted to sub-Saharan Africa [18]. Such diverse environments of members of the family Suidae suggests adaptation to endemic infectious disease agents may have occurred, that can be investigated as positive selection within TLR genes. However, information on how positive selection has influenced TLR genes within members of the Suidae is limited.

The aim of this study was to determine whether there is evidence of positive selective pressure in the family Suidae in a phylogenetic framework. We hypothesized that positive selection has contributed to the evolution of bacterial and viral sensing TLRs in the family Suidae. The specific aims of this study were to 1) identify evidence of persistent positive selection at TLRs across members of the family Suidae and 2) to determine whether restricted lineages within the Suidae demonstrate TLR positive selection. We focused on the bacterial sensing *TLR1*, *TLR2* and *TLR6* and viral sensing *TLR3*, *TLR7* and *TLR8* as viruses and bacteria are the dominant parasites threatening wild mammals [19]. Identifying positively selected residues within the TLR genes of members of the Suidae will yield vital information as to their adaptation to previous bacterial and viral infections. Our findings suggest that positive selection of TLRs amongst members of the Suidae has been mediated by infectious disease agents.

Materials and methods

Study animals

Ten animals representing 9 species of the family Suidae were utilized in this study. A range map showing the natural distribution of these species is shown in Fig. 2.1. The species *Sus scrofa* (wild boar) was represented by a European wild boar (*Sus scrofa* Europe) and a Asian wild boar (*Sus scrofa* Asia) to reflect the wide distribution of this species. Southeast Asian suids were represented by *Sus verrucosus* (javan warty pig), *Sus celebensis* (sulawesi warty pig), *Sus scebifrons* (visayan warty pig), *Sus barbatus* (bearded pig) and *Babyrousa babyrussa* (babirusa). Suidae species of African origin were represented by *Potamochoerus larvatus* (bush pig), *Potamochoerus porcus* (red river hog) and *Phacochoerus africanus* (common warthog).

Genes analyzed

TLR1, *TLR2* and *TLR6* encoding receptors for bacterial ligands and *TLR3*, *TLR7* and *TLR8* recognizing viral ligands were selected for this study. The extracellular domains were the focus since they encode the functional sites involved in pathogen ligand recognition.

DNA extraction and sequencing

DNA extraction, library preparation and sequencing were performed as previously described [20]. Briefly, DNA was extracted from whole blood by using the QIAamp DNA blood spin kit (Qiagen Sciences) and quantity and quality parameters were performed on the Qubit 2.0 fluorometer (Invitrogen) and run on a 1% agarose gel. Library construction and re-sequencing of individual members of the family Suidae were done with 1-3 ug of genomic DNA according to the Illumina library prepping protocols. The library insert size was 300-500 bp and sequencing was performed using a 100 paired-end sequencing kit [20]. All DNA were sequenced to approximately 8x depth. Quality trimmed reads (phred quality>20, minimum length of pairs of reads=40bp) were aligned to the *Sus scrofa* reference genome build 10.2 [21] using the unique alignment option of Mosaik Aligner (V.1.1.0017). The aligned reads from each of the animals together with the *Sus scrofa* reference genome were stored as bam files for each individual animal.

Orthologs identification and delineation of their extracellular domains

Porcine TLR mRNA sequences were obtained from Ensemble database (http://www.ensemble.org). The accession numbers of sequences obtained from the public databases were *TLR1*: NM_001031775, *TLR2*: NM_213761, *TLR3*: HQ412796, *TLR6*: NM_213760, *TLR7*: NM_001097434, *TLR8*: ENSSSCG00000012118. When a TLR gene was found to have more than one transcript, the longest transcript was chosen. The genomic coordinates of the porcine TLR mRNA sequences within the *Sus scrofa* genome assembly 10.2 were obtained from Ensemble. Based on these genomic coordinates, sequences of TLR gene orthologs were then retrieved from aligned bam files

(illumina resequencing data for family Suidae species aligned against *Sus scrofa* genome assembly 10.2) of *Sus scrofa* (*Sus scrofa* Europe and *Sus scrofa* Asia), *Sus verrucosus, Sus celebensis, Sus scebifrons, Sus barbatus Babyrousa babyrussa*, *Potamochoerus larvatus, Potamochoerus porcus* and *Phacochoerus africanus* to identify TLR gene orthologs. The resulting sequences for each species were then blast screened against the *Sus scrofa* genome to ensure similarity with the porcine TLR mRNA sequences. Exonic regions were then obtained from these sequences and concatenated to obtain coding sequences. The coding sequences were further trimmed to obtain sequences of the extracellular domain for each TLR in each species. Sequences were aligned using ClustalW 1.81 [22]. In this study, porcine TLR reference amino acid sequences were aligned to corresponding human and murine sequences in order to delineate the extracellular domains of porcine TLRs and their LRR modules and sub-domains [23–28] (Fig.2.2). The genomic coordinates of the TLR extracellular domains are provided in Table 2.1.

ML test for positive selection

Comparison of the non-synonymous substitutions per non-synonymous site (*dN*) with the number of synonymous substitutions per synonymous site (*dS*) in a maximum likelihood (ML) framework was used to test for positive selection for every codon, defining a *dN*/*dS* ratio (ω) > 1 in a codon as evidence of positive selection. First, we determined whether ω varied among codon sites for each TLR alignment by comparing CODEML program models in PAML version 4 [29,30] M0 which assumes that ω is constant across all sites in the alignment and M3 which allows ω to vary amongst sites.

Next, 4 site models were employed to detect sites under persistent positive selection. Two models each from the CODEML program in PAML version 4 [29,30] and the Datamonkey web server [31] were utilized. CODEML site model M1a, a nearly neutral evolution model where sites are assumed to be evolving under either purifying selection ($\omega < 1$) or neutral evolution ($\omega = 1$) was compared to model M2a that allows positive selection among sites. M7, which allows sites to evolve under either purifying selection or neutrally, was compared to model M8, which allows for positively selected sites. Models M7 and M8 differ from models M1a and M2a in that, the former assume that ω values are drawn from a beta

distribution [32]. Models were compared using a likelihood ratio test (LRT). In order to identify positive selection, twice the difference in log-likelihood values ($2\ln\Delta L$) between models would be significant by chi-square testing. The F3x4 model of codon frequencies was used for the analyses. Models were run in duplicates with ω of 0.5 and 1.5 to increase the probability of convergence of model parameters. The Bayes empirical Bayes (BEB) approach implemented in CODEML was used to identify codons under positive selection. BEB estimates the posterior probability of each site belonging to three selection classes: low, intermediate and high ω. Codon sites with $\omega>1$ and a posterior probability > 95% were inferred to be under positive selection. Fixed-effects likelihood (FEL) and Random-effects likelihood (REL) models implemented using the Datamonkey web server was also used to detect positive selection. The FEL model estimates synonymous and nonsynonymous rates directly at each codon site, without assuming an a priori distribution of rates across sites while REL model allows synonymous and nonsynonymous substitution rates to vary among codon sites. Codon sites were considered to be under positive selection at significant levels $p < 0.1$ for FEL and a Bayes factor > 50 for REL [33].

To test for episodic positive selection, Branch-Site REL and MEME (mixed effects model of evolution) implemented on the Datamonkey web server were utilized. The Branch-Site REL model estimates proportion of sites under selection along tree branches and allows evolutionary rates to simultaneously vary along phylogenetic branches and sites [16]. The MEME method identifies lineage-specific events of positive selection at sites, even though the same site is under purifying or neutral selection in other lineages [34]. A Suidae species tree (Fig. 2.3) derived from near complete genome sequences for members of the Suidae [17] [L. Frantz, personal communication, September 27, 2014] was used in all analyses.

Positively selected sites detected in this study were compared to human TLR Swiss-Prot database to determine their possible link to function. Sites under positive selection were also mapped to three dimensional (3D) protein structures using MuPIT Interactive [35] in order to examine their functional significance. We also determined the conservative or radical nature of amino acid changes at sites under positive selection within this study.

Results

The sequences (10 sequences within each TLR alignment) of the extracellular domains of bacterial sensing *TLR1*, *TLR2* and *TLR6* and viral sensing *TLR3*, *TLR7* and *TLR8* from species within the family Suidae were obtained. The length of the extracellular domains in terms of number of nucleotides of the TLRs ranged from 1668 bases for *TLR1* to 2445 bases for *TLR7*. Amino acid length ranged from 556 amino acids for *TLR1* to 792 amino acids for *TLR7*.

Heterogeneity of selective pressure along genes

To determine whether selective pressures varied amongst codon sites for each TLR gene, the M0 and M3 models of CODEML program was utilized. Comparison of M0 vs M3 indicated that *dN*/*dS* ratio (ω) of some TLR genes varied among codons, implying that selective constraints were heterogeneous between sites. We detected significant ($p < 0.01$ for $2\ln\Delta l$) heterogeneity of ω along *TLR1*, *TLR2*, *TLR6* and *TLR8* (Table 2.2). For these genes, we found that the proportion of sites with evidence of positive selection (p_2) is relatively smaller than the proportion of sites with evidence of purifying (p_0) or neutral (p_1) selection. Thus, the majority of sites within the proteins of *TLR1*, *TLR2*, *TLR6* and *TLR8* were functionally constrained. *TLR3* and *TLR7* sequences did not reveal heterogeneity of selection pressure ω among their codons and are thus functionally conserved along their entire extracellular domains within the members of Suidae involved in this study.

Detection of persistent positive selection across members of the Suidae

To detect positive selection pressure that have acted persistently and shared across most Suidae members regardless of their geographic origins, site models implemented in the CODEML program of the PAML package and on the Datamonkey web server were utilized. Site models permit detection of positive selection within gene codons. Site models detected positively selected codons in bacterial sensing *TLR1*, *TLR2* and *TLR6* and viral sensing *TLR8*. Specifically, comparisons of nested models available in CODEML program indicated that models including codons with $\omega > 1$ (M2a and M8) demonstrated a

better fit than did neutral models (M1a and M7) for all the four TLR genes (Table 2.3 and Table 2.4). Since detecting codons under positive selection using site-based methods have power limitations when analyzing a few closely related species [36], we defined sites under positive selection conservatively as those for which significant results were obtained by more than one site model. Such sites and the properties of their amino acids are shown in Table 2.5. Three codons were identified for *TLR1*, 2 codons for *TLR2*, 7 codons for *TLR6* and 2 codons for *TLR8* that showed evidence for persistent positive selection. The site based methods did not identify codons under positive selective pressure for *TLR3* and *TLR7*.

Detection of episodic positive selection in particular lineages

To detect signatures of episodic positive selection in specific lineages for each TLR gene, branch-site REL analysis available on the Datamonkey web server were performed. The branch-site REL identify lineages at which a proportion of sites have *dN*/*dS* ratios >1without making any assumptions as to which lineages should be analyzed for positive selection. With respect to *TLR1*, evidence for positive selection in the ancestral lineage of the genus *Sus* (internal branch leading to the *Sus* clade) and on *TLR2* species branch corresponding to *Sus verrucosus* (Table 2.6) were detected. Analyses also indicated that within the TLR2 gene, the species branch corresponding to *Potamochoerus porcus* is under positive selective pressure (Table 2.6). Thus, MEME was employed to identify sites under positive selection along branches. One codon position in *TLR1* (codon position 434) in the lineage leading to the genus *Sus* and the species branch corresponding to *Sus verrucosus* was identified as under positive selection. Another codon position (codon position 338 in *TLR2*) was found to be under positive selection in the species branch corresponding to *Potamochoerus porcus* (Table 2.6).

Functional significance of positively selected sites

To determine functional relevance of positively selected amino acid sites, sites determined to be under positive selection by more than one ML method were compared to human TLR Swiss-Prot entries (Table 2.7). First, human and porcine TLRs were aligned to determine the equivalent positions of positively selected sites in pigs within humans. Then analysis were performed to determine whether the sites in human TLRs have been implicated as having functional effects or are in close proximity to a functionally annotated site from human Swiss-Prot. Sites that were adjacent to residues and within regions known to affect TLR protein function (Table 2.7) were detected. Thus, amino acid sites under positive selection, as determined by more than one ML method were mapped onto TLR protein 3D crystallographic structures to gain further insight into their functional significance. Positively selected sites were within the following domains: *TLR1* 117 (LRR4), 434 (LRR16), 451 (LRR17), 559 (LRR carboxy termini (LRRCT)); *TLR2* 216 (LRR7), 338 (LRR12); *TLR6* 183 (LRR6), 334 (LRR12), 452 and 459 (LRR17), 501(LRR19), 554 and 560 (LRRCT); *TLR8* 178 (LRR5), 388 (LRR13). Positively selected sites which can be inferred to affect protein function based on their location within TLR protein 3D crystallographic structures are shown in Figure 2.4. Two of the positively selected codons (*TLR1* sites 434 and 451) are within *TLR1*/*TLR2* interface (Figure 2.4) and might have implications for *TLR1*/*TLR2* heterodimer formation. *TLR2* site 338 is in close proximity to a site that interacts with bacterial lipopeptides and may therefore have a role in ligand binding. The conservative or radical nature of amino acid changes occurring at positively selected sites was also determined. Radical amino acid changes have effects on protein function. Specific sites (*TLR1*:117, 559; *TLR2*:216, 338; *TLR6*:183, 334, 501, 554, 560; *TLR8*:388) have experienced radical amino acid changes (Table 2.5), suggesting a possible role of such sites in diverse protein functions.

Discussion

The important role of pathogen mediated positive selection pressure in shaping diversity in the TLRs of mammalian species has been documented elsewhere [12,13,37]. The adaptation of the members

of the Suidae to different environments presenting numerous bacterial and viral pathogenic challenges, make the family amenable to studies of pathogen mediated selection on immune genes. Results obtained in this study indicate that both persistent and episodic positive selection have shaped TLR evolution and diversity among the Suidae.

Our finding of small proportion of sites of *TLR1*, *TLR2*, *TLR6* and *TLR8* showing evidence of persistent positive selection agrees with the mostly accepted paradigm that purifying selection is the dominant force operating on TLRs [10,12]. As was the case with previous studies [12,15,38], more positively selected sites within bacterial-sensing TLRs than their viral-sensing counterparts were inferred. Viral infections are thought to exert stronger selective pressure than bacterial infections, constraining the evolution of viral-sensing TLRs [38]. In contrast to previous studies done in primates [12] and across rodents, carnivores, lagomorphs and primates [37], fewer sites under persistent positive selection within genes involved in this study were detected. Members of the Suidae represent closely related species and are therefore likely to be affected by fewer related bacteria and viruses than the diverse species involved in previous studies. *TLR6* stood out as the gene with the strongest evidence of selection, where more codons were under persistent positive selection. The dimerization interface in *TLR6*/*TLR2* is 80% larger than that of *TLR1*/*TLR2* [39]. Therefore one can speculate that the larger *TLR6*/*TLR2* dimerization surface exposes more codons of *TLR6* to positive selective pressure. The finding that among the bacterial sensing TLRs, *TLR2* had fewer sites under persistent positive selection despite having similar protein length as *TLR1* and *TLR6* is suggestive of a stronger selective constraint on *TLR2*. The *TLR2* gene product recognizes a myriad of ligands (microbial triacyl lipoproteins, diacyl lipoproteins found in mycoplasma, lipoteichoic acid of Gram-positive bacteria or Zymosan of yeast) pathogens through heterodimerization with *TLR1* and *TLR6* [40,41]. *TLR2* also been shown to affect IFN production, making the *TLR2* gene evolutionary constrained [42].

Apart from persistent pathogen mediated positive selection acting over long evolutionary time across members of the Suidae, the evolutionary histories of members of the Suidae may have been affected by periodic pathogenic infections confined to specific lineages within certain geographic

locations, leading to episodic positive selection within such lineages. Such a signal of adaptive evolution is usually masked by a background signal of purifying selection, which makes their identification difficult. Both Branch Site REL and the MEME methodology implemented in Datamonkey revealed the same lineages were evolving under episodic positive selection, suggesting that sites within these branches are under positive selection. MEME is a recently developed method [34] that allows the detection of episodic positive selection even when majority of lineages are evolving under purifying selection.

Our lineage specific analysis showed that the branch leading to the *Sus* clade was found to have undergone episodic positive selection at *TLR1* amino acid site 434 indicating that the ancestors of species within the genus *Sus* had to undergo adaptive changes at this site in response to their environment. With the exception of *Sus verrucosus* which had methionine at *TLR1* site 434, other *Sus* species had the leucine residue while the African suids and *Babyrousa babyrussa* had methionine. This finding suggests a possible selective advantage for leucine at *TLR1* site 434 in the environment in which ancestors of the *Sus* species originated. Indeed, methionine seems to be very rare at *TLR1* site 434 within the domesticated breeds of *Sus scrofa* [43], indicating leucine is preferred at this site. The substitutions of methionine with leucine within the interior of a protein increase protein stability [44] supporting a hypothesis that leucine within the *Sus* species stabilizes the *TLR1* protein prior to heterodimerization with *TLR2* for efficient recognition of diverse bacterial ligands (peptidoglycans and triacyl lipoproteins). This finding of positive selection on branches leading to *Sus verrucosus* for *TLR1* and *Potamochoerus porcus* for *TLR2* requires a cautious interpretation, since only one sequence from one animal is involved in each case. *Sus verrucosus* is thought to represent a distinct lineage following a deep split with other species of the genus *Sus* [17]. It is possible that bacterial pathogens restricted to *Sus verrucosus* may have exerted selective pressure on its *TLR1* gene. Related to *Potamochoerus porcus*, positive selection on *TLR2* gene could partly be due to adaptation to infectious agents within the African rain forest, a location outside of which they are rarely found [45].

The case for positive selection within TLR amino acid sites involved in this study is strengthened by the location of specific sites in close proximity to functionally relevant regions. Site 117 of *TLR1* is

within disulphide bonds region. Disulphide bonds are important to the overall function of proteins as they are associated with their folding and stability [46]. Site 434 of *TLR1* is adjacent to a glycosylation site. One conclusion would be that positive selection at this site is of consequence as glycosylation of TLRs is thought to influence receptor surface presentation, trafficking and ligand recognition [47]. The positive selection inferred at site 559 of *TLR1*, adjacent to a site that leads to impairment of NF-kB activation, suggests a role for this site in regulating inflammatory response to bacterial infection.

TLR1/*TLR2* heterodimer formation is required for ligand recognition and signal initiation [23]. Thus changes at sites 434 and 559 within *TLR1* suggest residues at these sites could be under selective pressure to improve the *TLR1/TLR2* heterodimer formation. As was the case with the study of [48] involving RIG-I-like pattern recognition receptors, in this study the majority (10/15 sites) of sites under positive selection involved radical amino acid residue changes across species of the Suidae. This is in agreement with positive selection favoring radical amino changes at sites within particular genes [49]. Such sites may be of functional significance.

Results obtained here have implications for present day domestic pigs. African wild suids are susceptible to some viral, bacterial and parasitic diseases of domestic pigs. As European and Asian wild boars are the progenitors of most domestic pigs, it is likely that species of the genus *Sus* are also susceptible to diseases of domestic swine [50]. Thus, residues that were under positive selection in the past could still be beneficial to domestic pigs in terms of disease resistance. Evidence for past positive selection influencing resistance or susceptibility to present day pathogens is seen in the Protein Kinase R (PKR) gene, where adaptive changes at important residues, most likely driven by old viruses [51], are important in the ability of PKR to fight infections from present-day poxyviruses [52].

Conclusion

In conclusion, residues within bacterial sensing *TLR1*, *TL2*, *TLR6* and viral sensing *TLR8* of members of the Suidae that have undergone persistent and episodic positive selection were identified. The evidence of positive selection on the TLR genes reveals that pathogen mediated selective pressure has

shaped Suidae TLR evolution. The case for positive selective at amino acid sites is strengthened by location of these sites in close proximity to functionally relevant sites and the radical changes in amino acids at some of these sites across members of the Suidae. Sites under positive selection may have aided in the adaptation of the Suidae to infectious agents that evolved rapidly or that were encountered in new environments.

Acknowledgements

We thank Dr. Laurie Rund of the department of animal sciences of the University of Illinois and Dr. Greger Larson of the School of Archaeology of Oxford University for reading the manuscript and making useful suggestions. We are also grateful to Laurent Frantz of the Animal Breeding and Genomics Centre, Wageningen University for providing the species tree of the Suidae used in this study.

Figures and Tables

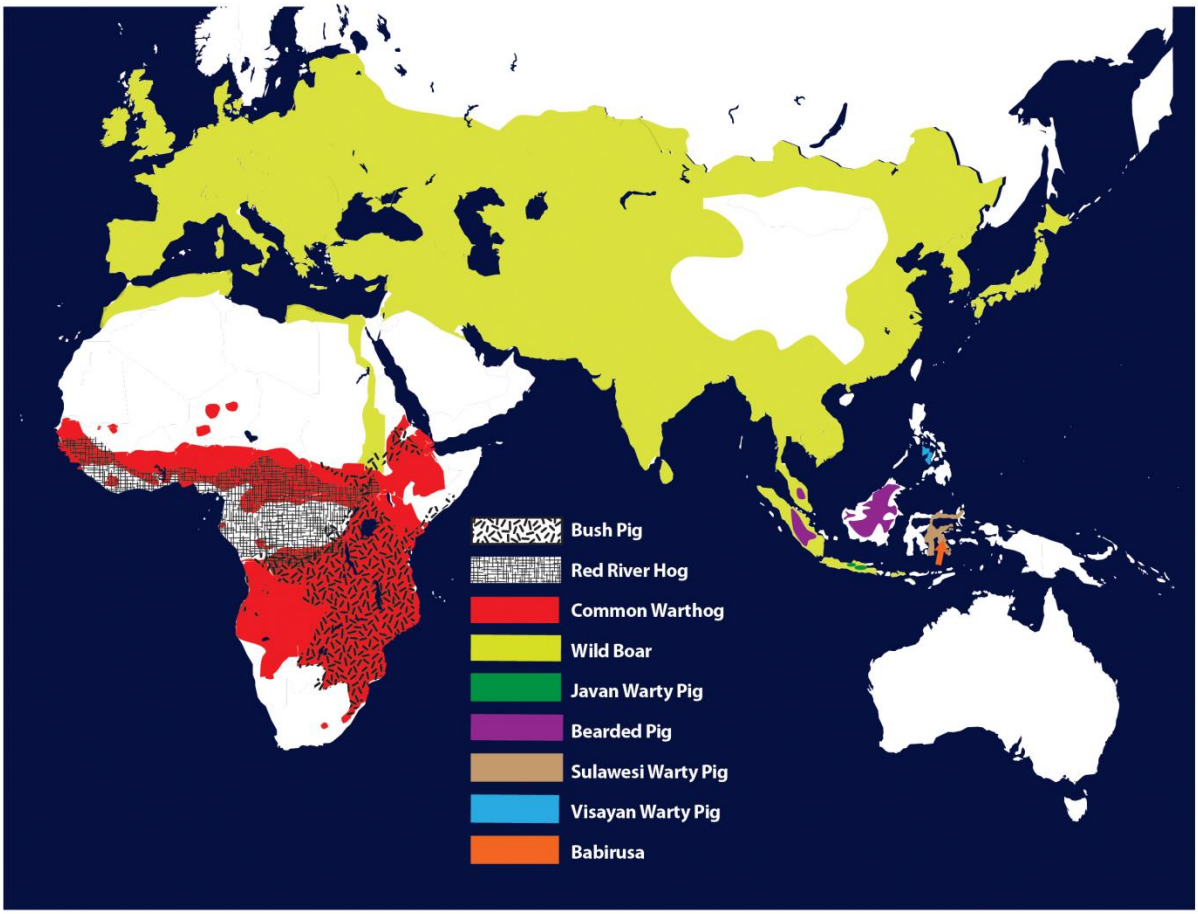

Fig. 2.1. Range map for members of the Suidae.

Fig. 2.2

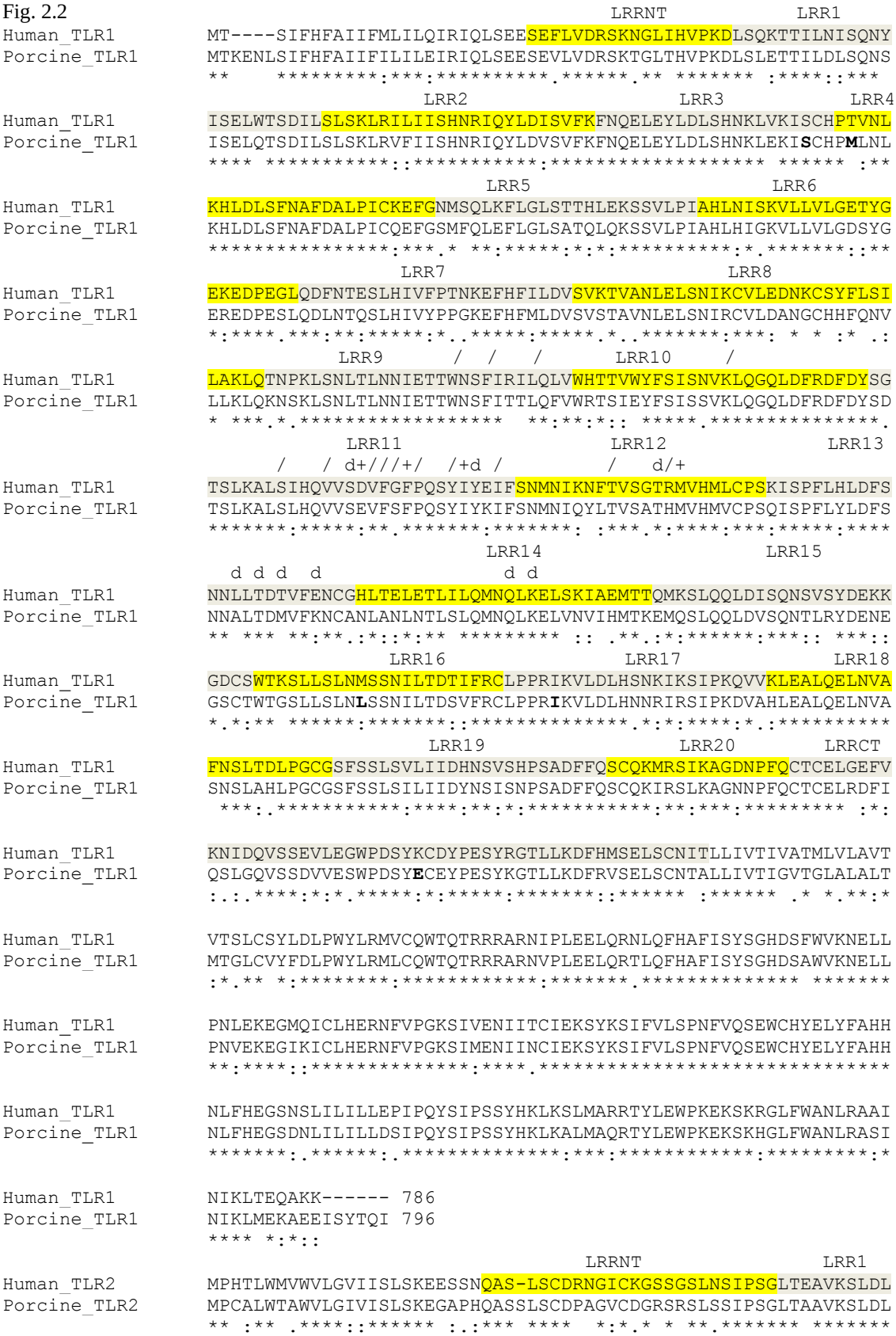
LRRNT
LRR1
Human_TLR1 MT----SIFHFAIIFMLILQIRIQLSEESEFLVDRSKNGLIHVPKDLSQKTTILNISQNY 56
Porcine_TLR1 MTKENLSIFHFAIIFILILEIRIQLSEESEVLVDRSKTGLTHVPKDLSLETTILDLSQNS 60
LRR2
LRR3
LRR4
Human_TLR1 ISELWTSDILSLSKLRILIISHNRIQYLDISVFKFNQELEYLDLSHNKLVKISCHPTVNL 116
Porcine_TLR1 ISELQTSDILSLSKLRVFIISHNRIQYLDVSVFKFNQELEYLDLSHNKLEKISCHPMLNL 120
LRR5
LRR6
Human_TLR1 KHLDLSFNAFDALPICKEFGNMSQLKFLGLSTTHLEKSSVLPIAHLNISKVLLVLGETYG 176
Porcine_TLR1 KHLDLSFNAFDALPICQEFGSMFQLEFLGLSATQLQKSSVLPIAHLHIGKVLLVLGDSYG 180
LRR7
LRR8
Human_TLR1 EKEDPEGLQDFNTESLHIVFPTNKEFHFILDVSVKTVANLELSNIKCVLEDNKCSYFLSI 236
Porcine_TLR1 EREDPESLQDLNTQSLHIVYPPGKEFHFMLDVSVSTAVNLELSNIRCVLDANGCHHFQNV 240
LRR9
LRR10
Human_TLR1 LAKLQTNPKLSNLTLNNIETTWNSFIRILQLVWHTTVWYFSISNVKLQGQLDFRDFDYSG 296
Porcine_TLR1 LLKLQKNSKLSNLTLNNIETTWNSFITTLQFVWRTSIEYFSISSVKLQGQLDFRDFDYSD 300
LRR11
LRR12
LRR13
Human_TLR1 TSLKALSIHQVVSDVFGFPQSYIYEIFSNMNIKNFTVSGTRMVHMLCPSKISPFLHLDFS 356
Porcine_TLR1 TSLKALSLHQVVSEVFSFPQSYIYKIFSNMNIQYLTVSATHMVHMVCPSQISPFLYLDFS 360
LRR14
LRR15
Human_TLR1 NNLLTDTVFENCGHLTELETLILQMNQLKELSKIAEMTTQMKSLQQLDISQNSVSYDEKK 416
Porcine_TLR1 NNALTDMVFKNCANLANLNTLSLQMNQLKELVNVIHMTKEMQSLQQLDVSQNTLRYDENE 420
LRR16
LRR17
LRR18
Human_TLR1 GDCSWTKSLLSLNMSSNILTDTIFRCLPPRIKVLDLHSNKIKSIPKQVVKLEALQELNVA 476
Porcine_TLR1 GSCTWTGSLLSLNLSSNILTDSVFRCLPPRIKVLDLHNNRIRSIPKDVAHLEALQELNVA 480
LRR19
LRR20
LRRCT
Human_TLR1 FNSLTDLPGCGSFSSLSVLIIDHNSVSHPSADFFQSCQKMRSIKAGDNPFQCTCELGEFV 536
Porcine_TLR1 SNSLAHLPGCGSFSSLSILIIDYNSISNPSADFFQSCQKIRSLKAGNNPFQCTCELRDFI 540
Human_TLR1 KNIDQVSSEVLEGWPDSYKCDYPESYRGTLLKDFHMSELSCNITLLIVTIVATMLVLAVT 596
Porcine_TLR1 QSLGQVSSDVVESWPDSYECEYPESYKGTLLKDFRVSELSCNTALLIVTIGVTGLALALT 600
Human_TLR1 VTSLCSYLDLPWYLRMVCQWTQTRRRARNIPLEELQRNLQFHAFISYSGHDSFWVKNELL 656
Porcine_TLR1 MTGLCVYFDLPWYLRMLCQWTQTRRRARNVPLEELQRTLQFHAFISYSGHDSAWVKNELL 660
Human_TLR1 PNLEKEGMQICLHERNFVPGKSIVENIITCIEKSYKSIFVLSPNFVQSEWCHYELYFAHH 716
Porcine_TLR1 PNVEKEGIKICLHERNFVPGKSIMENIINCIEKSYKSIFVLSPNFVQSEWCHYELYFAHH 720
Human_TLR1 NLFHEGSNSLILILLEPIPQYSIPSSYHKLKSLMARRTYLEWPKEKSKRGLFWANLRAAI 776
Porcine_TLR1 NLFHEGSDNLILILLDSIPQYSIPSSYHKLKALMAQRTYLEWPKEKSKHGLFWANLRASI 780
Human_TLR1 NIKLTEQAKK------ 786
Porcine_TLR1 NIKLMEKAEEISYTQI 796
LRRNT
LRR1
Human_TLR2 MPHTLWMVWVLGVIISLSKEESSNQAS-LSCDRNGICKGSSGSLNSIPSGLTEAVKSLDL 59
Porcine_TLR2 MPCALWTAWVLGIVISLSKEGAPHQASSLSCDPAGVCDGRSRSLSSIPSGLTAAVKSLDL 60

Fig. 2.2 (cont.)

```
                                    LRR2                        LRR3
Human_TLR2      SNNRITYISNSDLQRCVNLQALVLTSNGINTIEEDSFSSLGSLEHLDLSYNYLSNLSSSW 119
Porcine_TLR2    SNNRIAYVGSSDLRKCVNLRALRLGANSIHTVEEDSFSSLGSLEHLDLSYNHLSNLSSSW 120
                *****:*:..***::****:** * :*.*:*:********************:********

                          LRR4                      LRR5            LRR6
Human_TLR2      FKPLSSLTFLNLLGNPYKTLGETSLFSHLTKLQILRVGNMDTFTKIQRKDFAGLTFLEEL 179
Porcine_TLR2    FKSLSTLKFLNLLGNPYKTLGEAPLFSHLPNLRILKIGNNDTFPEIQAKDFQGLTFLQEL 180
                **.**:*.**************:.******.:*:**::** ***.:** *** *****:**

                                    LRR7                        LRR8
Human_TLR2      EIDASDLQSYEPKSLKSIQNVSHLILHMKQHILLLEIFVDVTSSVECLELRDTDLDTFHF 239
Porcine_TLR2    EIGASHLQRYAPKSLRSIQNISHLILHMRRPALLPKIFVDLLSSLEYLELRNTDFSTFNF 240
                **.**.** * ****:****:********::  ** :****: **:* ****:**:.**:*

                              LRR9   /    /    LRR10  / /    /    //
Human_TLR2      SELSTGETNSLIKKFTFRNVKITDESLFQVMKLLNQISGLLELEFDDCTLNGVGNFRASD 299
Porcine_TLR2    SDVSINEHCTVMKKFTFRKAEITDASFTEIVKLLNYVSGALEVEFDDCTLNGRGDLSTSA 300
                *::* .*  :::******:.:*** *: :::**** :** **:********* *:: :*

                         LRR11                    LRR12
                  /  / /  /d/ ddd+//       /   /  /  /d d/++/+  /
Human_TLR2      NDRVIDPGKVETLTIRRLHIPRFYLFYDLSTLYSLTERVKRITVENSKVFLVPCLLSQHL 359
Porcine_TLR2    LDTIKSLGNVETLTVRRLHIPQFFLFYDLRSIYSLTGAVKRITIENSKVFLVPCSLSQHL 360
                 * : . *:*****:******:*:***** ::****  *****:********** *****

                 LRR13   d d dddd d          LRR14     d             LRR15
Human_TLR2      KSLEYLDLSENLMVEEYLKNSACEDAWPSLQTLILRQNHLASLEKTGETLLTLKNLTNID 419
Porcine_TLR2    KSLEYLDLSENLMSEEYLKNSACEHAWPFLHTLILRQNHLKSLEKTGEVLVTLKNLTNLD 420
                ************* **********.*** *:********* *******.*:********:*

                                    LRR16                LRR17     LRR18
Human_TLR2      ISKNSFHSMPETCQWPEKMKYLNLSSTRIHSVTGCIPKTLEILDVSNNNLNLFSLNLPQL 479
Porcine_TLR2    ISKNNFDSMPETCQWPEKMKYLNLSSTRIHSLTHCLPQTLEVLDISNNNLNSFSLSLPQL 480
                ****.*.************************:* *:*:***:**:****** ***.****

                                   LRR19                LRR20
Human_TLR2      KELYISRNKLMTLPDASLLPMLLVLKISRNAITTFSKEQLDSFHTLKTLEAGGNNFICSC 539
Porcine_TLR2    KELYISRNKLKTLPDASFLPMLSVLRISRNTINTFSKEQLDSFQKLKTLEAGGNNFICSC 540
                ********** ******:**** **:****:*.**********:.***************

                              LRRCT
Human_TLR2      EFLSFTQEQQALAKVLIDWPANYLCDSPSHVRGQQVQDVRLSVSECHRTALVSGMCCALF 599
Porcine_TLR2    DFLSFTQGQQALAQVLSDWPENYLCDSPSHVRGQRVQDTRLSLTECHRVAVVSVVCCALF 600
                :****** *****:** *** ************:***.***::****.*:** :*****

Human_TLR2      LLILLTGVLCHRFHGLWYMKMMWAWLQAKRKPRKAPSRNICYDAFVSYSERDAYWVENLM 659
Porcine_TLR2    LLLLLTGALCHHFHGLWCMKMMWAWLQAKRKPRKAPRRDVCYDAFVSYSEQDSYWVENLM 660
                **:****.***:***** ***************** *::**********:*:*******

Human_TLR2      VQELENFNPPFKLCLHKRDFIPGKWIIDNIIDSIEKSHKTVFVLSENFVKSEWCKYELDF 719
Porcine_TLR2    VQELEHFQPPFKLCLHKRDFIPGKWIIDNIIDSIEKSQKTIFVLSENFVKSEWCKYELDF 720
                *****:*:*****************************:**:******************

Human_TLR2      SHFRLFDENNDAAILILLEPIEKKAIPQRFCKLRKIMNTKTYLEWPMDEAQREGFWVNLR 779
Porcine_TLR2    SHFRLFDENDDTAILILLEPIEKKTIPQRFCKLRKIMNTRTYLEWPADETQREGFWLNLR 780
                *********:*:************:**************:****** **:******:***

Human_TLR2      AAIKS 784
Porcine_TLR2    AAIKS 785


                                                  LRRNT/              LRR1
Human_TLR3      MRQTLPC-IYFWGGLLPFGMLCASSTTKCTVSHEVADCSHLKLTQVPDDLPTNITVLNLT 59
Porcine_TLR3    MSRSLPCHIYSFWVLLPFWILYTTSTNKCTVRHEIADCSHLKLTQIPDDLPANITVLNLT 60
                * ::*** ** :  **** :* ::**.**** **:**********:*****:********

                                    LRR2                    LRR3
                / / /                / /                    / / /
Human_TLR3      HNQLRRLPAANFTRYSQLTSLDVGFNTISKLEPELCQKLPMLKVLNLQHNELSQLSDKTF 119
Porcine_TLR3    HNQLRGLPPANFTIYSQLTTLDGGFNTIPKLEPELCQSLPLLDILNLQHNELSQLSDKTF 120
                ***** **.**** *****:** *****.********.**:*.:***************
```

Fig. 2.2 (cont.)

```
                                LRR4                    LRR5                LRR6
Human_TLR3        AFCTNLTELHLMSNSIQKIKNNPFVKQKNLITLDLSHNGLSSTKLGTQVQLENLQELLLS 179
Porcine_TLR3      IFCMNLIELHLMSNSIQKIQNNPFKNLKNLIKLDLSHNGLSSTKLGTQLQLENLQELLLA 180
                   ** ** ************:**** : ****.****************:**********:
                                          LRR7                      LRR8
Human_TLR3        NNKIQALKSEELDIFANSSLKKLELSSNQIKEFSPGCFHAIGRLFGLFLNNVQLGPSLTE 239
Porcine_TLR3      NNKISALKREELDFLGNSSLKRLELSSNQIQEFSPGCFHAIGKLFGLSLNNVKLSPSLTE 240
                  ****.*** ****::.*****:********:**********:**** ****:*.*****
                                    LRR9                  LRR10
Human_TLR3        KLCLELANTSIRNLSLSNSQLSTTSNTTFLGLKWTNLTMLDLSYNNLNVVGNDSFAWLPQ 299
Porcine_TLR3      KLCLELSNTSIENLSLSNIQLYKTSNTTFFGLKQTNLSMLDLSHNSLSVIGNDSFAWLPH 300
                  ******:****.****** ** .******:*** ***:*****:*.*.*:**********:
                              LRR11                       LRR12
Human_TLR3        LEYFFLEYNNIQHLFSHSLHGLFNVRYLNLKRSFTKQSISLASLPKIDDFSFQWLKCLEH 359
Porcine_TLR3      LKYFFLEYNNIERLSSRSLYGLSNVKYLNLRRSFTKQSISLASLPKIEDFSFQWLKSLEY 360
                  *:**********::* *:**:** **:****:****************:********.**:
                           LRR13                       LRR14                LRR15
Human_TLR3        LNMEDNDIPGIKSNMFTGLINLKYLSLSNSFTSLRTLTNETFVSLAHSPLHILNLTKNKI 419
Porcine_TLR3      LNMEDNNFPGIKRNTFTGLIKLKSLSLSNSFSSLRTLTNETFISLADSPLLILNLTKNKI 420
                  ******::**** * *****:** *******:**********:***.*** *********
                                       LRR16                    LRR17
Human_TLR3        SKIESDAFSWLGHLEVLDLGLNEIGQELTGQEWRGLENIFEIYLSYNKYLQLTRNSFALV 479
Porcine_TLR3      SKIESGAFSWLGHLKVLDLGLNEIGQELTGQEWRGLKNIVEVYLSYNRYLELTTDSFALV 480
                  *****.********:**********************:**.*:*****:**:** :*****
                              LRR18                     / /     LRR19    LRR20   /
Human_TLR3        PSLQRLMLRRVALKNVDSSPSPFQPLRNLTILDLSNNNIANINDDMLEGLEKLEILDLQH 539
Porcine_TLR3      PSLQQLMLRRVALRDMDCSPSPFHPLFNLTILDLSNNNIANINDELLKGLEKLQILDLQH 540
                  ****:********:::*.*****:** *****************::*:*****:******
                   /  /                            / LRR21                LRR22
Human_TLR3        NNLARLWKHANPGGPIYFLKGLSHLHILNLESNGFDEIPVEVFKDLFELKIIDLGLNNLN 599
Porcine_TLR3      NNLARLWKHANPGGPVQFLKGLSHLHILNLESNGFDEIPADAFRDLSELKSIDLGLNNLN 600
                  ***************: **********************.:.*:** *** *********
                                        /LRR23                        LRRCT
Human_TLR3        TLPASVFNNQVSLKSLNLQKNLITSVEKKVFGPAFRNLTELDMRFNPFDCTCESIAWFVN 659
Porcine_TLR3      ILPPSVFDNQVSLKSLSLQKNLITSVKKTVFGPAFQKLSNLDMRFNPFDCTCESIAWFVS 660
                   **.***:********.*********:*.******::*::*******************.

Human_TLR3        WINETHTNIPELSSHYLCNTPPHYHGFPVRLFDTSSCKDSAPFELFFMINTSILLIFIFI 719
Porcine_TLR3      WINSTHTNISELSSHYLCNTPPQYHGLPVILFDTSPCKDSAPFELFFMITASMLLIFIFI 720
                  ***.*****.************:***:** ******.*************.:*:*******

Human_TLR3        VLLIHFEGWRISFYWNVSVHRVLGFKEIDRQTEQFEYAAYIIHAYKDKDWVWEHFSSMEK 779
Porcine_TLR3      ILLIHFEGWRISFYWNVSVHRVLGFKEIDKQPEQFEYAAYIIHAYKDRDWVWEHFAPMEE 780
                  :*****************************:*.***************:*******:.**:

Human_TLR3        EDQSLKFCLEERDFEAGVFELEAIVNSIKRSRKIIFVITHHLLKDPLCKRFKVHHAVQQA 839
Porcine_TLR3      KDETLRFCLEERDFEAGALELEAIVNSIKRSRKIIFVITQHLLKDPLCKRFKVHHAVQQA 840
                  :*:::*:***********.:*********************:********************

Human_TLR3        IEQNLDSIILVFLEEIPDYKLNHALCLRRGMFKSHCILNWPVQKERIGAFRHKLQVALGS 899
Porcine_TLR3      IEQNLDSIILIFLEEIPDYKLNHALCLRRGMFKSHCILNWPVQKERINAFHHKLQVALGS 900
                  **********:************************************.**:*********

Human_TLR3        KNSVH 904
Porcine_TLR3      RNSVH 905
                  :****
```

Fig. 2.2 (cont.)

```
                                                    LRRNT             LRR1
Human_TLR6      MTKDKEPIVKSFHFVCLMIIIVGTRIQFSDGNEFAVDKSKRGLIHVPKDLPLKTKVLDMS 60
Porcine_TLR6    MTKDKKPTVISLHSVYVMTLVWGTLIQFSEESEFVVDKSKIGLTRVPKDLPPQTKVLDVS 60
                *****:* * *:* * :* :: ** ****: .**.***** ** :****** :*****:*
                                        LRR2                  LRR3
Human_TLR6      QNYIAELQVSDMSFLSELTVLRLSHNRIQLLDLSVFKFNQDLEYLDLSHNQLQKISCHPI 120
Porcine_TLR6    QNFITELHLSDISFLSQLTVLRLSQNRMQCLDISVFKFNQDLEYLDLSHNQLQTILCHPI 120
                **:*:**::**:****:*******:**:* **:********************.* ****
                           LRR4                 LRR5              LRR6
Human_TLR6      VSFRHLDLSFNDFKALPICKEFGNLSQLNFLGLSAMKLQKLDLLPIAHLHLSYILLDLRN 180
Porcine_TLR6    TSLKHLDLSFNDFEALPICKEFGNLTQLNFLGLSATKLQQLDLLPIAHLHLSCILLDLER 180
                .*::*********:**********:********* ***:************ *****..
                                      LRR7                        LRR8
Human_TLR6      YYIKENETESLQILNAKTLHLVFHPTSLFAIQVNISVNTLGCLQLTNIKLNDDNCQVFIK 240
Porcine_TLR6    YYMKENEKESLQILNTEKLHLVFHPNSFFSVQVNISVKSVGCLQLANIKLGDDNCQVFIT 240
                **:****.*******::.*******.*:*::******:::*****:****.********.
                                LRR9                  LRR10
Human_TLR6      FLSELTRGSTLLNFTLNHIETTWKCLVRVFQFLWPKPVEYLNIYNLTIIESIREEDFTYS 300
Porcine_TLR6    FLLELTQGPTLLNFTLNHVETTWKCLVGIFQFLWPKPVEYLSIYNLTIVESIDEEDFIYY 300
                ** ***:*.*********:******** :************.******:*** **** *
                                LRR11               LRR12            LRR13
                          d dd  d+/     d               d d d  d
Human_TLR6      KTTLKALTIEHITNQVFLFSQTALYTVFSEMNIMMLTISDTPFIHMLCPHAPSTFKFLNF 360
Porcine_TLR6    ETTLKGVKIEHITKRVFIFSQTALYRVFSDMNIRMLTIADTHFIHMLCPQVPSTFNFLNF 360
                :****.:.*****::**:******* ***:*** ****:** *******:.****:****
                 d d dd                        d  LRR14             LRR15
Human_TLR6      TQNVFTDSIFEKCSTLVKLETLILQKNGLKDLFKVGLMTKDMPSLEILDVSWNSLESGRH 420
Porcine_TLR6    TQNVFTDSVFQNCKTLARLETLILQKNKLEDLFKISLMTKDMLSLEILDVSSNSLEYDRH 420
                ********:*::*.**.:********* *:****:.****** ******** **** .**
                                LRR16                 LRR17          LRR18
Human_TLR6      KENCTWVESIVVLNLSSNMLTDSVFRCLPPRIKVLDLHSNKIKSVPKQVVKLEALQELNV 480
Porcine_TLR6    GENCTWVGSIVVLNLSSNILTDSVFRCLPPRIKVLDLHSNRIRSIPKDVAHLEALQELNV 480
                 ****** **********:**********************:*:*:**:*.:*********
                                LRR19                 LRR20        LRRCT
Human_TLR6      AFNSLTDLPGCGSFSSLSVLIIDHNSVSHPSADFFQSCQKMRSIKAGDNPFQCTCELREF 540
Porcine_TLR6    ASNSLAHLPGCGSFSSLSILSIDYNSISNPSADFFQSCQKIRSLKAGNNPFQCTCELRDF 540
                * ***:.***********:* **:**:*:***********:**:***:**********:*

Human_TLR6      VKNIDQVSSEVLEGWPDSYKCDYPESYRGSPLKDFHMSELSCNITLLIVTIGATMLVLAV 600
Porcine_TLR6    IQSLGQVSSDVVESWPDSYECEYPESYKGTLLKDFRVSELSCNTALLIVTIGVTGLALAL 600
                ::.:.****:*:*.*****:*:*****:*: ****::****** :*******.* *.**:

Human_TLR6      TVTSLCIYLDLPWYLRMVCQWTQTRRRARNIPLEELQRNLQFHAFISYSEHDSAWVKSEL 660
Porcine_TLR6    TMTGLCVYFDLPWYLRMLCQWTQTRRRARNVPLEELQRTLQFHAFISYSEHDSAWVKNEL 660
                *:*.**:*:********:************:*******.******************.**

Human_TLR6      VPYLEKEDIQICLHERNFVPGKSIVENIINCIEKSYKSIFVLSPNFVQSEWCHYELYFAH 720
Porcine_TLR6    VPCLEKEGIKICLHERNFVPGKSIMENIINCIEKSYKSIFVLSPNFVQSEWCHYELYFAH 720
                ** ****.*:***************:************************************

Human_TLR6      HNLFHEGSNNLILILLEPIPQNSIPNKYHKLKALMTQRTYLQWPKEKSKRGLFWANIRAA 780
Porcine_TLR6    HNLFHEGSDNLILILLDPIPQNSIPGKYHKLKALMAQRTYLEWPKEKSKHGPFWANIRAA 780
                ********:*******:********.*********:*****:*******:* ********

Human_TLR6      FNMKLTLVTENNDVKS 796
Porcine_TLR6    FNIKLKLVAEEDDVKT 796
                **:**.**:*::***:
```

Fig. 2.2 (cont.)

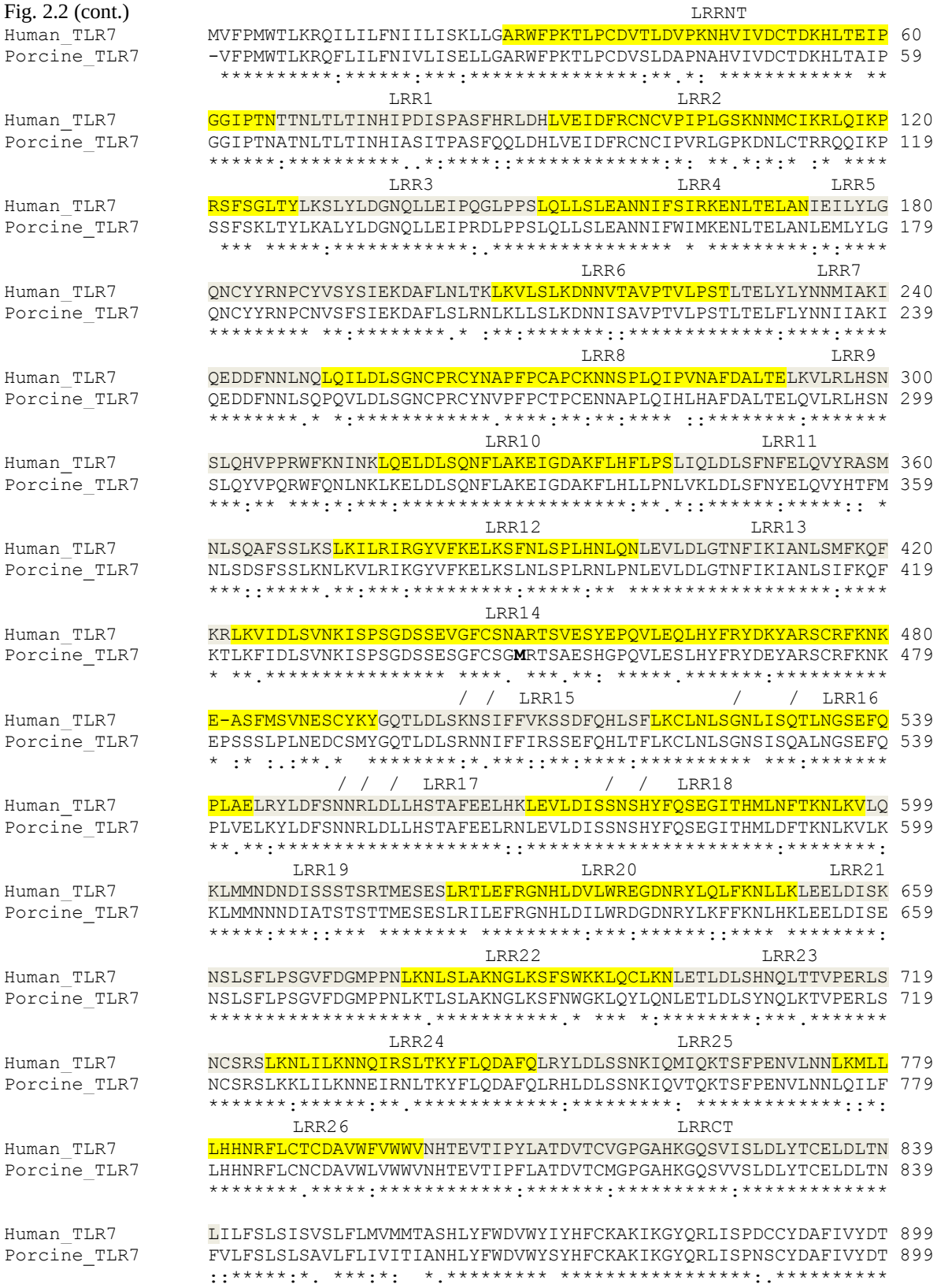
LRRNT
Human_TLR7 MVFPMWTLKRQILILFNIILISKLLGARWFPKTLPCDVTLDVPKNHVIVDCTDKHLTEIP 60
Porcine_TLR7 -VFPMWTLKRQFLILFNIVLISELLGARWFPKTLPCDVSLDAPNAHVIVDCTDKHLTAIP 59
LRR1
LRR2
Human_TLR7 GGIPTNTTNLTLTINHIPDISPASFHRLDHLVEIDFRCNCVPIPLGSKNNMCIKRLQIKP 120
Porcine_TLR7 GGIPTNATNLTLTINHIASITPASFQQLDHLVEIDFRCNCIPVRLGPKDNLCTRRQQIKP 119
LRR3
LRR4
LRR5
Human_TLR7 RSFSGLTYLKSLYLDGNQLLEIPQGLPPSLQLLSLEANNIFSIRKENLTELANIEILYLG 180
Porcine_TLR7 SSFSKLTYLKALYLDGNQLLEIPRDLPPSLQLLSLEANNIFWIMKENLTELANLEMLYLG 179
LRR6
LRR7
Human_TLR7 QNCYYRNPCYVSYSIEKDAFLNLTKLKVLSLKDNNVTAVPTVLPSTLTELYLYNNMIAKI 240
Porcine_TLR7 QNCYYRNPCNVSFSIEKDAFLSLRNLKLLSLKDNNISAVPTVLPSTLTELFLYNNIIAKI 239
LRR8
LRR9
Human_TLR7 QEDDFNNLNQLQILDLSGNCPRCYNAPFPCAPCKNNSPLQIPVNAFDALTELKVLRLHSN 300
Porcine_TLR7 QEDDFNNLSQPQVLDLSGNCPRCYNVPFPCTPCENNAPLQIHLHAFDALTELQVLRLHSN 299
LRR10
LRR11
Human_TLR7 SLQHVPPRWFKNINKLQELDLSQNFLAKEIGDAKFLHFLPSLIQLDLSFNFELQVYRASM 360
Porcine_TLR7 SLQYVPQRWFQNLNKLKELDLSQNFLAKEIGDAKFLHLLPNLVKLDLSFNYELQVYHTFM 359
LRR12
LRR13
Human_TLR7 NLSQAFSSLKSLKILRIRGYVFKELKSFNLSPLHNLQNLEVLDLGTNFIKIANLSMFKQF 420
Porcine_TLR7 NLSDSFSSLKNLKVLRIKGYVFKELKSLNLSPLRNLPNLEVLDLGTNFIKIANLSIFKQF 419
LRR14
Human_TLR7 KRLKVIDLSVNKISPSGDSSEVGFCSNARTSVESYEPQVLEQLHYFRYDKYARSCRFKNK 480
Porcine_TLR7 KTLKFIDLSVNKISPSGDSSESGFCSGMRTSAESHGPQVLESLHYFRYDEYARSCRFKNK 479
LRR15
LRR16
Human_TLR7 E-ASFMSVNESCYKYGQTLDLSKNSIFFVKSSDFQHLSFLKCLNLSGNLISQTLNGSEFQ 539
Porcine_TLR7 EPSSSLPLNEDCSMYGQTLDLSRNNIFFIRSSEFQHLTFLKCLNLSGNSISQALNGSEFQ 539
LRR17
LRR18
Human_TLR7 PLAELRYLDFSNNRLDLLHSTAFEELHKLEVLDISSNSHYFQSEGITHMLNFTKNLKVLQ 599
Porcine_TLR7 PLVELKYLDFSNNRLDLLHSTAFEELRNLEVLDISSNSHYFQSEGITHMLDFTKNLKVLK 599
LRR19
LRR20
LRR21
Human_TLR7 KLMMNDNDISSSTSRTMESESLRTLEFRGNHLDVLWREGDNRYLQLFKNLLKLEELDISK 659
Porcine_TLR7 KLMMNNNDIATSTSTTMESESLRILEFRGNHLDILWRDGDNRYLKFFKNLHKLEELDISE 659
LRR22
LRR23
Human_TLR7 NSLSFLPSGVFDGMPPNLKNLSLAKNGLKSFSWKKLQCLKNLETLDLSHNQLTTVPERLS 719
Porcine_TLR7 NSLSFLPSGVFDGMPPNLKTLSLAKNGLKSFNWGKLQYLQNLETLDLSYNQLKTVPERLS 719
LRR24
LRR25
Human_TLR7 NCSRSLKNLILKNNQIRSLTKYFLQDAFQLRYLDLSSNKIQMIQKTSFPENVLNNLKMLL 779
Porcine_TLR7 NCSRSLKKLILKNNEIRNLTKYFLQDAFQLRHLDLSSNKIQVTQKTSFPENVLNNLQILF 779
LRR26
LRRCT
Human_TLR7 LHHNRFLCTCDAVWFVWWVNHTEVTIPYLATDVTCVGPGAHKGQSVISLDLYTCELDLTN 839
Porcine_TLR7 LHHNRFLCNCDAVWLVWWVNHTEVTIPFLATDVTCMGPGAHKGQSVVSLDLYTCELDLTN 839
Human_TLR7 LILFSLSISVSLFLMVMMTASHLYFWDVWYIYHFCKAKIKGYQRLISPDCCYDAFIVYDT 899
Porcine_TLR7 FVLFSLSLSAVLFLIVITIANHLYFWDVWYSYHFCKAKIKGYQRLISPNSCYDAFIVYDT 899

Fig. 2.2 (cont.)

```
Human_TLR7        KDPAVTEWVLAELVAKLEDPREKHFNLCLEERDWLPGQPVLENLSQSIQLSKKTVFVMTD 959
Porcine_TLR7      KDPAVTEWVLDELVAKLEDPREKHFNLCLEERDWLPGQPVLENLSQSIQLSKKTVFVMTD 959
                  ********** ************************************************

Human_TLR7        KYAKTENFKIAFYLSHQRLMDEKVDVIILIFLEKPFQKSKFLQLRKRLCGSSVLEWPTNP 1019
Porcine_TLR7      KYAKTEKFKIAFYLSHQRLMDEKVDVIILIFLEKPLQKSKFFQLRKRLCGSSVLEWPTNP 1019
                  ******:****************************:*****:******************

Human_TLR7        QAHPYFWQCLKNALATDNHVAYSQVFKETV 1049
Porcine_TLR7      QAHPYFWQCLKNALATDNHVTYSQVFKETA 1049
                  ********************:********.
```

```
                                                                LRRNT
Human_TLR8        MENMFLQSSMLTCIFLLISGSCELCAEENFSRSYPCDEKKQNDSVIAECSNRRLQEVPQT 60
Porcine_TLR8      ---MTLHFLLLTCLFLLIPDSCEFFTGANYSRSYPCDERKENGSVIAECNNRQLQEVPRR 57
                     * *:  :***:****..***: :  *:********:*:*.******.**:*****:
                               LRR1                              LRR2
Human_TLR8        VGKYVTELDLSDNFITHITNESFQGLQNLTKINLNHNPNVQHQNGNPGIQSNGLNITDGA 120
Porcine_TLR8      VGNYVTELDLSDNFIRRITNESFQGLQNLTKINLNHNAKLWPQS------ENGMTITDGA 111
                  **:************ :*********************.::  *.      .**:.*****
                                 LRR3                    LRR4            LRR5
Human_TLR8        FLNLKNLRELLLEDNQLPQIPSGLPESLTELSLIQNNIYNITKEGISRLINLKNLYLAWN 180
Porcine_TLR8      FLNLHHLRELLLEDNQLREIPTGLPESLRELSLIQNKIILLNTKNMFGLRKLESLYLGWN 171
                  ****::************ :**:****** *******:*  :..:.:  * :*:.***.**
                   dddd                          LRR6             LRR7
Human_TLR8        CYFNKVCEKTNIEDGVFETLTNLELLSLSFNSLSHVPPKLPSSLRKLFLSNTQIKYISEE 240
Porcine_TLR8      CYFT-CNETFIIDEGAFENLTNLKVLSLSFNTLYRVPPKLPSSLTKLYLSNTKIRNINQE 230
                  ***.   *.  *::*.**.****::*******:* :********* **:****:*: *.:*
                                ddddd       LRR8                  LRR9
Human_TLR8        DFKGLINLTLLDLSGNCPRCFNAPFPCVPCDGGASINIDRFAFQNLTQLRYLNLSSTSLR 300
Porcine_TLR8      DFKGLENLRVLDLSGNCPRCFNAPFPCNPCPGDASIQIHPLAFRYLTELRYLNLSSTSLR 290
                  ***** ** :***************** ** *.***:*. :**: **:************
                                        LRR10                 / /    /  LRR11
Human_TLR8        KINAAWFKNMPHLKVLDLEFNYLVGEIASGAFLTMLPRLEILDLSFNYIKGSYPQHINIS 360
Porcine_TLR8      RIPATWFENLHHLKVLHLEFNYLMDEIASGEFLAKLPSLEILDLSYNYEQKKYPQYINIS 350
                  :* *:**:*: ***** .******:.***** **: ** *******:** : .***:****
                                 / /   LRR12                  / / LRR13
Human_TLR8        RNFSKLLSLRALHLRGYVFQELREDDFQPLMQLPNLSTINLGINFIKQIDFKLFQNFSNL 420
Porcine_TLR8      HYFANLTSLQILHLRAYVFQELRKEDFQPLRNLLHL**KF**INLGINFIKQIDFTIF**S**EFSNL 410
                  : *::* **: ****.*******::***** :* :*. *************.:*.:****

                        dd/                                ddd    LRR14
Human_TLR8        EIIYLSENRISPLVKDTRQSYANSSSFQRHIRKRRSTDFEFDPHSNFYHFTRPLIKPQCA 480
Porcine_TLR8      SIIYLSENRISPLVNNTGQKNGDRPSFQSHVLKPRSATPKFDPHSNFYHNTKPLIKPQCS 470
                  .*************::* *. .: .*** *: * **:  :********* *:*******:
                             ddd d    LRR15     LRR16 ddd   /             LRR17
Human_TLR8        AYGKALDLSLNSIFFIGPNQFENLPDIACLNLSANSNAQVLSGTEFSAIPHVKYLDLTNN 540
Porcine_TLR8      RYGKALDLSLNSIFFIGPNQFEAFKDIACLNLSSNGNGQVLHGSEFSHLPGIKYLDLTNN 530
                   ******************** : ********:*.*.*** *:*** :* :*********
                  d / /                    ddd ddd///  LRR18    LRR19      dddd
Human_TLR8        RLDFDNASALTELSDLEVLDLSYNSHYFRIAGVTHHLEFIQNFTNLKVLNLSHNNIYTLT 600
Porcine_TLR8      RLDFDDDAAFSELPLLEVLDLSYNSHYFRIAGVTHRLGFIQNLPQLRVLNLSHNSIFTLT 590
                  *****: :*::**. ********************:* ****:.:*:*******.*:***
```

Fig. 2.2 (cont.)

```
                        d          LRR20     d   dddddd          LRR21      d
Human_TLR8      DKYNLESKSLVELVFSGNRLDILWNDDDNRYISIFKGLKNLTRLDLSLNRLKHIPNEAFL 660
Porcine_TLR8    ETY-LKSTSLKELVFSGNRLDLLWNAQDDRYWQIFKNLSTLTHLDLSSNNLQHIPSEAFL 649
                :.* *:*.** **********:*** :*:** .***.*..**:**** *.*:***.****
                                  LRR22               LRR23            LRR24
Human_TLR8      NLPASLTELHINDNMLKFFNWTLLQQFPRLELLDLRGNKLLFLTDSLSDFTSSLRTLLLS 720
Porcine_TLR8    NLPQTLTELYISDNRLNFFNWSLLQQFPNLTLLDLSGNELSFLTDSLSKFTTSLQTLILR 709
                *** :****:*.** *:****:******.* **** **:* *******.**:**:**:*
                                       LRR25                              LRR26
Human_TLR8      HNRISHLPSGFLSEVSSLKHLDLSSNLLKTINKSALETKTTTKLSMLELHGNPFECTCDI 780
Porcine_TLR8    QNRISYLPSGLLSEASSLTHLDLSSNQLKMVNISKLHAKTTTNLAILKLDRNPFDCTCDI 769
                :****:****:***.***.******* ** :* * *.:****:*::*:*. ***:*****
                                                  LRRCT
Human_TLR8      GDFRRWMDEHLNVKIPRLVDVICASPGDQRGKSIVSLELTTCVSDVTAVILFFFTFFITT 840
Porcine_TLR8    RDFRKWMDENLKVTIPRLTDVICASPGDQRGRSIVSLELTTCVSDTIAAIICFFTFFVTS 829
                 ***:****:*:*.****.************:**************. *.*: *****:*:

Human_TLR8      MVMLAALAHHLFYWDVWFIYNVCLAKVKGYRSLSTSQTFYDAYISYDTKDASVTDWVINE 900
Porcine_TLR8    TVMLAALAHHWFYWDAWFIYHVCLAKVKGYRSLPTSQTFYDAYVSYDTKDASVTDWVMNE 889
                 ********* ****.****:************.*********:*************:**
```

Fig. 2.2 (cont.)

```
Human_TLR8      LRYHLEESRDKNVLLCLEERDWDPGLAIIDNLMQSINQSKKTVFVLTKKYAKSWNFKTAF 960
Porcine_TLR8    LRFHLEESEGKNVLLCLEERDWDPGLAIIDNLMQSINQSKKTIFVLTKKYAKNWNFKTAF 949
                **:*****..*********************************:*********.*******

Human_TLR8      YLALQRLMDENMDVIIFILLEPVLQHSQYLRLRQRICKSSILQWPDNPKAEGLFWQTLRN 1020
Porcine_TLR8    YLALQRLMDENMDVIVFILLEPVLQHSQYLRLRQRICKSSILQWPDNPKAEGLFWQSLKN 1009
                ***************:*****************************************:*:*

Human_TLR8      VVLTENDSRYNNMYVDSIKQY 1041
Porcine_TLR8    VVLTENDSRYNSLYVNSIK-- 1028
                ***********.:**:***
```

Fig. 2.2. Domain characterization of TLRs. / ligand binding residues, d residues involved in dimerization, + residues involved in both ligand binding and dimerization. LRR represents leucine rich repeat. LRRNT represents LRR amino termini. LRRCT represents LRR carboxy termini. Human TLR sequences accession numbers: *TLR1*:Q15399; *TLR2*:O60603; *TLR3*: NP_003256.1; *TLR6*:Q9Y2C9; *TLR7*:NP_057646; *TLR8*:NP_619542. Porcine TLR sequences accession numbers: *TLR1*:NP_001026945; *TLR2*:NP_998926.1; *TLR3*:DQ266435.1 *TLR6*:NP_998925.1; *TLR7*: DQ647699; *TLR8*: NP_999352.1. For *TLR6*, human LRRs were determined from their alignment with murine *TLR6*. Asterisks, colons and periods under aligned the aligned sequences indicate complete match, strong conservation and weaker conservation of amino acid respectively.

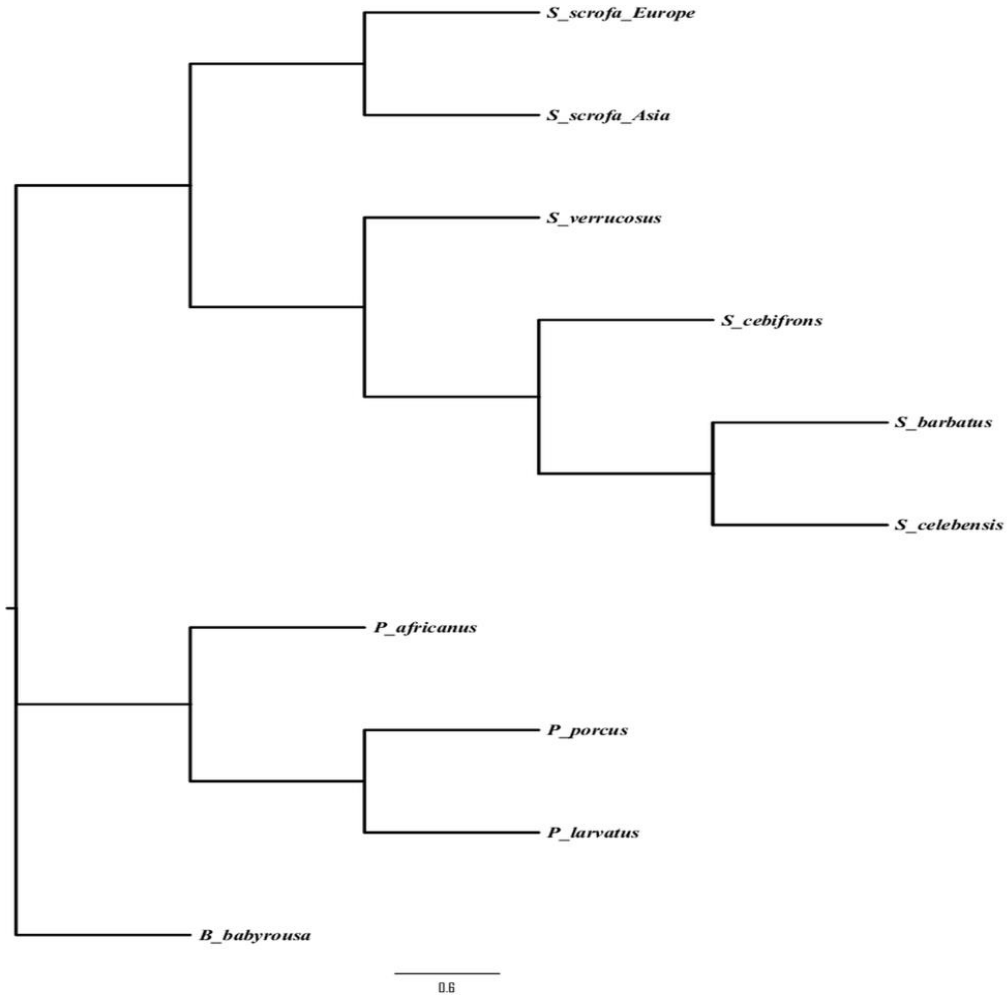

Fig. 2.3. Species phylogeny of the Suidae. Shown here is a representation of the relationships among members of the Suidae used in analyses. The relationships were derived from near complete genome data of each species. The posterior probability at each node is 1.

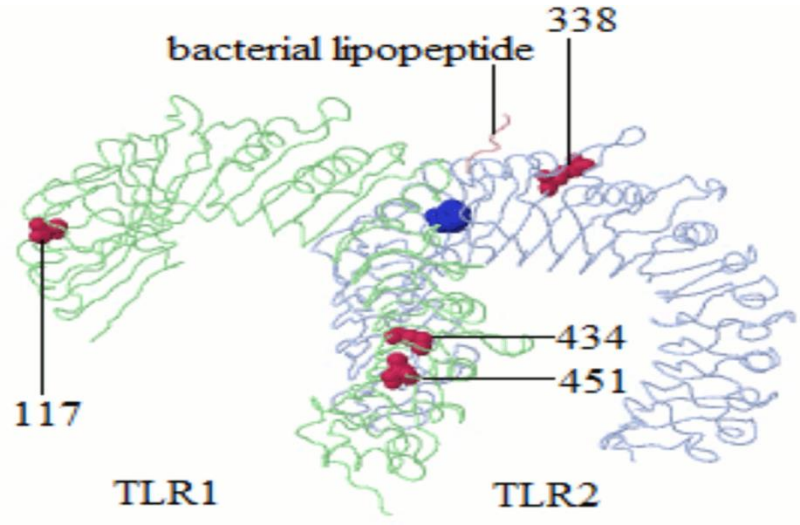

Fig. 2.4. Positively selected sites in the three dimensional structure of TLR1/TLR2 heterodimer. Positively selected sites are colored in red. A site known to interact with bacterial lipopeptide is colored in blue. Only sites likely to affection protein function based on their location within the structure are shown.

Table 2.1

Gene	Exon no[a]	Genomic coordinates of extracellular domain[b]	Aligned length (bp) of extracellular domain sequences[c]
TLR1	3	8:31628613-31630280:-1	1668
TLR2	2	8:79825324-79827018:-1	1695
TLR3	4	15:53849144-53849505:-1	366
TLR3	5	15:53848182-53848373:-1	192
TLR3	6	15:53841845-53843311:-1	1467
TLR6	2	8:31642930-31644612:-1	1683
TLR7	2	X:10472926-10475370:1	2445
TLR8	4	X:10509874-10512249:1	2376

Table 2.1. Summary of TLR extracellular domains studied

[a]Exon encoding the extracellular domain

[b]Genomic coordinates were determined by blat search of porcine TLR mRNA sequences against *Sus scrofa* genome assembly build 10.2.

[c]Sample origin for species for which sequences were obtained are indicated in bracket as follows: *Potomochoerus larvatus* (San diego zoo, USA), *Potomochoerus porcus* (San diego zoo, USA), *Phacochoerus africanus* (Omaha's zoo, USA), *Babyrousa Babyrousa* (San diego zoo, USA), *Sus verucossus* (Surabaya zoo, Indonesia), *Sus celebensis* (San diego zoo, USA), *Sus barbatus* (Omaha's zoo, USA), *Sus cebifrons* (San diego zoo, USA), *Sus scrofa* ,Europe, (Meinweg, Roerdalen, Netherlands), *Sus scrofa*, Asia (North China).

Table 2.2

Gene	Model	Parameter estimates	Log likelihood (L)	$2\ln\Delta L^{a}$
TLR1	M0	$\omega = 0.3788$	-2648.2197	35.9222**
	M3	$p_0 = 0.2608$, $\omega_0 = 0.2002$, $p_1 = 0.7235$, $\omega_1 = 0.2002$, $p_2 = 0.0157$, $\omega_2 = 20.3741$	-2630.2586	
TLR2	M0	$\omega = 0.4383$	-2762.6134	17.2620**
	M3	$p_0 = 0.0695$, $\omega_0 = 0.2757$, $p_1 = 0.9142$, $\omega_1 = 0.2757$, $p_2 = 0.0163$, $\omega_2 = 11.1174$	-2753.9824	
TLR3	M0	$\omega = 0.2649$	-2952.9019	0.0000^{NS}
	M3	$p_0 = 0.2240$, $\omega_0 = 0.2649$, $p_1 = 0.3124$, $\omega_1 = 0.2650$, $p_2 = 0.4635$, $\omega_2 = 0.2650$	-2952.9019	
TLR6	M0	$\omega = 0.4692$	-2695.5056	30.8540**
	M3	$p_0 = 0.3864$, $\omega_0 = 0.0000$, $p_1 = 0.5357$, $\omega_1 = 0.0000$, $p_2 = 0.0780$, $\omega_2 = 6.7044$	-2680.0786	
TLR7	M0	$\omega = 0.1087$	-3980.3407	1.2288^{NS}
	M3	$p_0 = 0.4460$, $\omega_0 = 0.0000$, $p_1 = 0.2772$, $\omega_1 = 0.0000$, $p_2 = 0.2768$, $\omega_2 = 0.3977$	-3979.7263	
TLR8	M0	$\omega = 0.3085$	-3985.5001	21.3048**
	M3	$p_0 = 0.7187$, $\omega_0 = 0.0000$, $p_1 = 0.1417$, $\omega_1 = 0.0000$, $p_2 = 0.1397$, $\omega_2 = 2.3685$	-3974.8477	

Table 2.2. Test of heterogeneity of selective pressure along genes

p_0 is proportion of sites where $\omega < 1(\omega_0)$, p_1 is proportion of sites where $\omega = 1(\omega_1)$

p_2 is proportion of sites where $\omega > 1(\omega_2)$

[a] Twice the difference in log-likelihood values between models M0 and M3

**$p < 0.01$

[NS]Not significant

Table 2.3

Gene	**Codons**	**Tests of selection**					
		PAML 2a		**PAML M8**		**FEL**	**REL**
		pp > 0.95	$2\ln\Delta L^a$	pp > 0.95	$2\ln\Delta L^b$	**p < 0.1**	**BF > 50**
TLR1	117	0.9850	21.3890**	0.9940	21.6814**		167.5190
	434	0.9970		0.9990			5232.9900
	451			0.9610			58.204
	559	0.9930		0.9970			8417.3400
TLR2	216	0.9580	7.8286**	0.9800	8.0464**		354.448
	338	0.9620		0.9900			370.5550
TLR6	49						98.3187
	63						107.5270
	79						84.4861
	121						80.5618
	180						84.6772
	183		16.1352	0.9520	16.1358**		742.6930
	187						97.7689
	213						92.1804
	269						98.6309
	307						137.7390
	334			0.9660			1210.1800
	356						104.0040
	386						102.0440
	394						98.9399
	452	0.9880		0.9960		0.0704	1844.8500
	459			0.9540			922.0600
	467						102.7000
	470						79.1197
	501			0.9610			1592.2300
	536						92.1069
	554			0.9560			1087.2700
	560			0.9650			1104.8600

Table 2.3 (cont.)

Gene	Codons	Tests of selection					
		PAML 2a		**PAML M8**		**FEL p < 0.1**	**REL BF > 50**
TLR8	178		4.6024		5.0782		226.4190
	236						232.4120
	387						260.8170
	388			0.9590			711.3420
	405						300.8590
	412						258.9650
	740						270.1570
	778						233.6280

Table 2.3. Results from site models at codons under persistent positive selection in members of the family Suidae
pp represents posterior probability in the BEB analysis
Sites identified by more than one ML method are underlined
[a]Twice the difference in log-likelihood values between models M1a and M2a
[b]Twice the difference in log-likelihood values between models M7 and M8
**$p < 0.01$

Table 2.4

Gene	**Model**	**Parameters**	**Log likelihood**
TLR1	M1	$p_0 = 0.7411$, $\omega_0 = 0.0000$, $p_1 = 0.2589$, $\omega_1 = 1.0000$	-2640.9531
	M2	$p_0 = 0.9843$, $\omega_0 = 0.2002$, $p_1 = 0.0000$, $\omega_1 = 1.0000$	-2630.2586
		$p_2 = 0.0157$, $\omega_2 = 20.3745$	
	M7	$p = 0.0050$, $q = 0.0118$	-2641. 0997
	M8	$p_0 = 0.9843$, $p = 24.8444$, $q = 99.0000$, $p_1 = 0.0157$, $\omega = 20.3864$	-2630.2590
TLR2	M1	$p_0 = 0.6628$, $\omega_0 = 0.0000$, $p_1 = 0.3372$, $\omega_1 = 1.0000$	-2757.8967
	M2	$p_0 = 0.9837$, $\omega_0 = 0.2757$, $p_1 = 0.0000$, $\omega_1 = 1.0000$	-2753.9824
		$p_2 = 0.0163$, $\omega_2 = 11.1175$	
	M7	$p = 0.0050$, $q = 0.0115$	-2758. 0063
	M8	$p_0 = 0.9837$, $p = 37.7613$, $q = 99.0000$, $p_1 = 0.0163$, $\omega = 11.1332$	-2753.9831
TLR3	M1	$p_0 = 0.7411$, $\omega_0 = 0.0000$, $p_1 = 0.2589$, $\omega_1 = 1.0000$	-2640.9531
	M2	$p_0 = 0.9843$, $\omega_0 = 0.2002$, $p_1 = 0.0000$, $\omega_1 = 1.0000$	-2630.2586
		$p_2 = 0.0157$, $\omega_2 = 20.3745$	
	M7	$p = 0.0050$, $q = 0.0118$	-2641. 0997
	M8	$p_0 = 0.9843$, $p = 24.8444$, $q = 99.0000$, $p_1 = 0.0157$, $\omega = 20.3864$	-2630.2590
TLR6	M1	$p_0 = 0.7018$, $\omega_0 = 0.0000$, $p_1 = 0.2982$, $\omega_1 = 1.0000$	-2688.1462
	M2	$p_0 = 0.9221$, $\omega_0 = 0.0000$, $p_1 = 0.0000$, $\omega_1 = 1.0000$	-2680.0786
		$p_2 = 0.0780$, $\omega_2 = 6.7045$	
	M7	$p = 0.0050$, $q = 0.0117$	-2688. 1465
	M8	$p_0 = 0.9221$, $p = 0.0050$, $q = 1.9056$, $p_1 = 0.0780$, $\omega = 6.7045$	-2680.0786
TLR7	M1	$p_0 = 0.9554$, $\omega_0 = 0.0692$, $p_1 = 0.0446$, $\omega_1 = 1.0000$	-3979.8010
	M2	$p_0 = 0.9554$, $\omega_0 = 0.0692$, $p_1 = 0.0247$, $\omega_1 = 1.0000$	-3979.8010
		$p_2 = 0.0200$, $\omega_2 = 1.000$	
	M7	$p = 0.1901$, $q = 1.4764$	-3979.7424
	M8	$p_0 = 1.0000$, $p = 0.1901$, $q = 1.4765$, $p_1 = 0.0000$, $\omega = 1.0000$	-3979.7424
TLR8	M1	$p_0 = 0.7371$, $\omega_0 = 0.0000$, $p_1 = 0.2629$, $\omega_1 = 1.0000$	-3977.1489
	M2	$p_0 = 0.8604$, $\omega_0 = 0.0000$, $p_1 = 0.0000$, $\omega_1 = 1.0000$	-3974.8477
		$p_2 = 0.1397$, $\omega_2 = 2.3685$	
	M7	$p = 0.0050$, $q = 0.0119$	-3977.3868
	M8	$p_0 = 0.8604$, $p = 0.0050$, $q = 2.8051$, $p_1 = 0.1397$, $\omega = 2.3686$	-3974.8477

Table 2.4. Parameter estimates for PAML models used in detecting persistent positive selection in members of the family Suidae

p_0 represents proportion of sites with $\omega < 1$

p_1 represents proportion of sites with $\omega = 1$

p_2 represents proportion of sites with $\omega > 1$

p and q represent parameters of the beta distribution

Table 2.5

Species	**Origin**	**Genes**														
		TLR1				***TLR2***		***TLR6***							***TLR8***	
		Codons														
		117	434*	451	559	216	338*	183	334	452	459	501	554	560	178	388
S. scrofa (E)	Eurasia	Met	Leu	Ile	Glu	Lys	Ala	Met	Arg	Ile	Ser	Thr	Ser	Glu	Asp	Phe
S. scrofa (A)		Met	Leu	Ile	Glu	Lys	Ala	Met	Arg	Ile	Asn	Ile	Gly	Lys	Glu	Phe
S. barbatus	Southeast Asia	Thr	Leu	Val	Glu	Lys	Thr	X	Arg	Ile	Asn	X	Gly	Glu	Asp	Phe
S. celebensis		Thr	Leu	Val	Lys	Glu	Thr	Met	Gly	Val	X	Ile	Gly	Glu	Glu	Phe
S. cebifrons		Thr	Leu	X	X	Lys	Thr	Thr	Arg	Val	Asn	Thr	Gly	Glu	Glu	Phe
S. verrucosus		Thr	Met	Ile	X	Glu	Thr	Met	Gly	Ile	X	Ile	Ser	Glu	Glu	Phe
B. babyrussa		Ile	Met	Val	Lys	Glu	Ala	Val	Arg	Val	Ser	Ile	Gly	Lys	Glu	Phe
P. larvatus	Africa	Thr	Met	Val	Glu	Glu	Ala	Val	Arg	Val	Asn	Ile	Gly	Glu	Glu	Leu
P. porcus		Thr	Met	Val	Glu	Glu	Lys	Val	Arg	Val	Asn	Ile	Gly	Glu	Glu	Phe
P. africanus		Thr	Met	Val	Lys	Glu	Ala	Val	Arg	Val	Asn	Ile	Gly	Glu	Glu	Val

Table 2.5. Positively selected codons within the extracellular domains of TLRs

*Codon site 434 in TLR1 is under episodic positive selection and codon site 338 in TLR2 is under both persistent and episodic positive selection. Other codon sites are under persistent positive selection**.** Amino acid properties: Met, Ile, Leu, Val, Ala are non-polar aliphatic; Thr, Ser, Asn are polar neutral; Lys, Arg are polar positive; Glu, Asp are polar negative; Gly is non-polar neutral; Phe is non-polar aromatic. (E) represents Europe, (A) represents Asia. X represents an undetermined amino acid.

Table 2.6

Branch-site REL analysis					MEME analysis	
Gene	**Branch**	**ω^+**	**Pr[$\omega = \omega^+$]**	**p value**	**Codons**	**p value**
TLR1	Ancestral lineage of *Sus*	1012.60	0.0002	0.003	434	0.090
	S.verrucosus	785.40	0.0002	0.003	434	0.090
TLR2	*P. porcus*	3334.61	0.0001	0.001	338	0.002

Table 2.6. Branches and codons under lineage specific positive selection in family Suidae
ω+ represents the ω value inferred for positively selected sites along branch; Pr[ω = ω+] represents the proportion of sites inferred to be evolving at ω+ along branch; p represents the p-value for episodic selection at branch corrected for multiple testing using the Holm-Bonferroni method.

Table 2.7

Gene	Codon	Equivalent codon in human	Functional Information
TLR1	117	113	In the region (110-132) of cysteine residues participating in disulphide bonds[a]
	434	430	Adjacent to a glycosylation site (429)[a]
	559	555	Adjacent to site of SNP (Tyr554Cys) that leads to marked impairment of NF-kB activation[b]
TLR2	338	337	LRR12 – involved in ligand binding[a]
TLR6	334	334	LRR12 – involved in TLR2-TLR6 heterodimerization[c]

Table 2.7. Positively selected sites predicted to affect TLR function based on human Swiss Prot entries and site location in three dimensional TLR structures

[a, b, c] represent references:

[a] [23]

[b] [42]

[c] [25]

References

1. Dawson HD, Loveland JE, Pascal G, Gilbert JGR, Uenishi H, Mann KM, et al. Structural and functional annotation of the porcine immunome. BMC Genomics. 2013;14: 332. Available: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3658956&tool=pmcentrez&rendertype=abstract

2. Morgan CC, Loughran NB, Walsh TA, Harrison AJ, O'Connell MJ. Positive selection neighboring functionally essential sites and disease-implicated regions of mammalian reproductive proteins. BMC Evol Biol. 2010;10: 39. Available: http://www.biomedcentral.com/1471-2148/10/39

3. Klein J, Sato A. The HLA system. First of two parts. N Engl J Med. 2000;343: 702–9. Available: http://www.ncbi.nlm.nih.gov/pubmed/10974135

4. Jepson A, Banya W, Sisay-Joof F, Hassan-King M, Nunes C, Bennett S, et al. Quantification of the relative contribution of major histocompatibility complex (MHC) and non-MHC genes to human immune responses to foreign antigens. Infect Immun. 1997;65: 872–6. Available: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=175062&tool=pmcentrez&rendertype=abstract

5. Fornuskova A, Bryja J, Vinkler M, Macholán M, Piálek J. Contrasting patterns of polymorphism and selection in bacterial-sensing toll-like receptor 4 in two house mouse subspecies. Ecol Evol. 2014;4: 2931–44. Available: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=4130449&tool=pmcentrez&rendertype=abstract

6. Medzhitov R. Recognition of microorganisms and activation of the immune response. Nature. Nature Publishing Group; 2007;449: 819–26. Available: http://dx.doi.org/10.1038/nature06246

7. Kawai T, Akira S. The role of pattern-recognition receptors in innate immunity: update on Toll-like receptors. Nat Immunol. Nature Publishing Group; 2010;11: 373–84. Available: http://dx.doi.org/10.1038/ni.1863

8. Werling D, Jungi TW. TOLL-like receptors linking innate and adaptive immune response. Vet Immunol Immunopathol. 2003;91: 1–12. Available: http://www.ncbi.nlm.nih.gov/pubmed/12507844

9. Akira S, Uematsu S, Takeuchi O. Pathogen recognition and innate immunity. Cell. 2006;124: 783–801. Available: http://www.ncbi.nlm.nih.gov/pubmed/16497588

10. Mukherjee S, Sarkar-Roy N, Wagener DK, Majumder PP. Signatures of natural selection are not uniform across genes of innate immune system, but purifying selection is the dominant signature. Proc Natl Acad Sci U S A. 2009;106: 7073–8. Available: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=2678448&tool=pmcentrez&rendertype=abstract

11. Ferrer-Admetlla A, Bosch E, Sikora M, Marquès-Bonet T, Ramírez-Soriano A, Muntasell A, et al. Balancing selection is the main force shaping the evolution of innate immunity genes. J Immunol. 2008;181: 1315–22. Available: http://www.ncbi.nlm.nih.gov/pubmed/18606686

12. Wlasiuk G, Nachman MW. Adaptation and constraint at Toll-like receptors in primates. Mol Biol Evol. 2010;27: 2172–86. Available: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3107592&tool=pmcentrez&rendertype=abstract

13. Alcaide M, Edwards S V. Molecular evolution of the toll-like receptor multigene family in birds. Mol Biol Evol. 2011;28: 1703–15. Available: http://www.ncbi.nlm.nih.gov/pubmed/21239391

14. Grueber CE, Wallis GP, Jamieson IG. Episodic positive selection in the evolution of avian toll-like receptor innate immunity genes. Ellegren H, editor. PLoS One. Public Library of Science; 2014;9: e89632. Available: http://dx.plos.org/10.1371/journal.pone.0089632

15. Fornůsková A, Vinkler M, Pagès M, Galan M, Jousselin E, Cerqueira F, et al. Contrasted evolutionary histories of two Toll-like receptors (Tlr4 and Tlr7) in wild rodents (MURINAE). BMC Evol Biol. 2013;13: 194. Available: http://www.biomedcentral.com/1471-2148/13/194

16. Kosakovsky Pond SL, Murrell B, Fourment M, Frost SDW, Delport W, Scheffler K. A random effects branch-site model for detecting episodic diversifying selection. Mol Biol Evol. 2011;28: 3033–43. Available: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3247808&tool=pmcentrez&rendertype=abstract

17. Frantz LAF, Schraiber JG, Madsen O, Megens H-J, Bosse M, Paudel Y, et al. Genome sequencing reveals fine scale diversification and reticulation history during speciation in Sus. Genome Biol. 2013;14: R107. Available: http://genomebiology.com/2013/14/9/R107

18. Grubb P. Order Artiodactyla. In: Wilson DE, Reeder DM, editors. Mammal species of the world: a taxanomic and geographic reference. 3rd ed. Baltimore, Maryland: Johns Hopkins University Press; 2005. pp. 637–772.

19. Pedersen A, Jones K, Nunn C, Altizer S. Infectious diseases and extinction risk in wild mammals [Internet]. CONSERVATION BIOLOGY. 2007. Available: http://discovery.ucl.ac.uk/1347949/

20. Bosse M, Megens H-J, Madsen O, Paudel Y, Frantz LAF, Schook LB, et al. Regions of homozygosity in the porcine genome: consequence of demography and the recombination landscape. PLoS Genet. 2012;8: e1003100. Available: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3510040&tool=pmcentrez&rendertype=abstract

21. Groenen MAM, Archibald AL, Uenishi H, Tuggle CK, Takeuchi Y, Rothschild MF, et al. Analyses of pig genomes provide insight into porcine demography and evolution. Nature. 2012;491: 393–8. Available: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3566564&tool=pmcentrez&rendertype=abstract

22. Thompson JD, Gibson TJ, Plewniak F, Jeanmougin F, Higgins DG. The CLUSTAL_X windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. Nucleic Acids Res. 1997;25: 4876–82. Available: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=147148&tool=pmcentrez&rendertype=abstract

23. Jin MS, Kim SE, Heo JY, Lee ME, Kim HM, Paik S-G, et al. Crystal structure of the TLR1-TLR2 heterodimer induced by binding of a tri-acylated lipopeptide. Cell. 2007;130: 1071–82. Available: http://www.ncbi.nlm.nih.gov/pubmed/17889651

24. Liu L, Botos I, Wang Y, Leonard JN, Shiloach J, Segal DM, et al. Structural basis of toll-like receptor 3 signaling with double-stranded RNA. Science. 2008;320: 379–81. Available: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=2761030&tool=pmcentrez&rendertype=abstract

25. Kang JY, Nan X, Jin MS, Youn S-J, Ryu YH, Mah S, et al. Recognition of lipopeptide patterns by Toll-like receptor 2-Toll-like receptor 6 heterodimer. Immunity. 2009;31: 873–84. Available: http://www.ncbi.nlm.nih.gov/pubmed/19931471

26. Wei T, Gong J, Jamitzky F, Heckl WM, Stark RW, Rössle SC. Homology modeling of human Toll-like receptors TLR7, 8, and 9 ligand-binding domains. Protein Sci. 2009;18: 1684–91. Available: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=2776956&tool=pmcentrez&rendertype=abstract

27. Tanji H, Ohto U, Shibata T, Miyake K, Shimizu T. Structural reorganization of the Toll-like receptor 8 dimer induced by agonistic ligands. Science. 2013;339: 1426–9. Available: http://www.ncbi.nlm.nih.gov/pubmed/23520111

28. Govindaraj RG, Manavalan B, Basith S, Choi S. Comparative analysis of species-specific ligand recognition in Toll-like receptor 8 signaling: a hypothesis. PLoS One. 2011;6: e25118. Available: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3176813&tool=pmcentrez&rendertype=abstract

29. Yang Z. PAML: a program package for phylogenetic analysis by maximum likelihood. Comput Appl Biosci. 1997;13: 555–6. Available: http://www.ncbi.nlm.nih.gov/pubmed/9367129

30. Yang Z. PAML 4: phylogenetic analysis by maximum likelihood. Mol Biol Evol. 2007;24: 1586–91. Available: http://www.ncbi.nlm.nih.gov/pubmed/17483113

31. Delport W, Poon AFY, Frost SDW, Kosakovsky Pond SL. Datamonkey 2010: a suite of phylogenetic analysis tools for evolutionary biology. Bioinformatics. 2010;26: 2455–7. Available: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=2944195&tool=pmcentrez&rendertype=abstract

32. Nielsen R, Yang Z. Likelihood models for detecting positively selected amino acid sites and applications to the HIV-1 envelope gene. Genetics. 1998;148: 929–36. Available: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=1460041&tool=pmcentrez&rendertype=abstract

33. Kosakovsky Pond SL, Frost SDW. Not so different after all: a comparison of methods for detecting amino acid sites under selection. Mol Biol Evol. 2005;22: 1208–22. Available: http://www.ncbi.nlm.nih.gov/pubmed/15703242

34. Murrell B, Wertheim JO, Moola S, Weighill T, Scheffler K, Kosakovsky Pond SL. Detecting individual sites subject to episodic diversifying selection. PLoS Genet. 2012;8: e1002764.

Available: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3395634&tool=pmcentrez&rendertype=abstract

35. Niknafs N, Kim D, Kim R, Diekhans M, Ryan M, Stenson PD, et al. MuPIT interactive: webserver for mapping variant positions to annotated, interactive 3D structures. Hum Genet. 2013;132: 1235–43. Available: http://www.ncbi.nlm.nih.gov/pubmed/23793516

36. Aguileta G, Lengelle J, Marthey S, Chiapello H, Rodolphe F, Gendrault A, et al. Finding candidate genes under positive selection in Non-model species: examples of genes involved in host specialization in pathogens. Mol Ecol. 2010;19: 292–306. Available: http://www.ncbi.nlm.nih.gov/pubmed/20041992

37. Areal H, Abrantes J, Esteves PJ. Signatures of positive selection in Toll-like receptor (TLR) genes in mammals. BMC Evol Biol. 2011;11: 368. Available: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3276489&tool=pmcentrez&rendertype=abstract

38. Barreiro LB, Ben-Ali M, Quach H, Laval G, Patin E, Pickrell JK, et al. Evolutionary dynamics of human Toll-like receptors and their different contributions to host defense. PLoS Genet. 2009;5: e1000562. doi:10.1371/journal.pgen.1000562

39. Kim HM, Park BS, Kim J-I, Kim SE, Lee J, Oh SC, et al. Crystal structure of the TLR4-MD-2 complex with bound endotoxin antagonist Eritoran. Cell. 2007;130: 906–17. Available: http://www.ncbi.nlm.nih.gov/pubmed/17803912

40. Ozinsky A, Underhill DM, Fontenot JD, Hajjar AM, Smith KD, Wilson CB, et al. The repertoire for pattern recognition of pathogens by the innate immune system is defined by cooperation between toll-like receptors. Proc Natl Acad Sci U S A. 2000;97: 13766–71. Available: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=17650&tool=pmcentrez&rendertype=abstract

41. Hajjar AM, O'Mahony DS, Ozinsky A, Underhill DM, Aderem A, Klebanoff SJ, et al. Cutting edge: functional interactions between toll-like receptor (TLR) 2 and TLR1 or TLR6 in response to phenol-soluble modulin. J Immunol. 2001;166: 15–9. Available: http://www.ncbi.nlm.nih.gov/pubmed/11123271

42. Ben-Ali M, Corre B, Manry J, Barreiro LB, Quach H, Boniotto M, et al. Functional characterization of naturally occurring genetic variants in the human TLR1-2-6 gene family. Hum Mutat. 2011;32: 643–52. Available: http://www.ncbi.nlm.nih.gov/pubmed/21618349

43. Shinkai H, Tanaka M, Morozumi T, Eguchi-Ogawa T, Okumura N, Muneta Y, et al. Biased distribution of single nucleotide polymorphisms (SNPs) in porcine Toll-like receptor 1 (TLR1), TLR2, TLR4, TLR5, and TLR6 genes. Immunogenetics. 2006;58: 324–30. Available: http://www.ncbi.nlm.nih.gov/pubmed/16604477

44. Lipscomb LA, Gassner NC, Snow SD, Eldridge AM, Baase WA, Drew DL, et al. Context-dependent protein stabilization by methionine-to-leucine substitution shown in T4 lysozyme. Protein Sci. 1998;7: 765–73. Available:

http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=2143956&tool=pmcentrez&rendertype=abstract

45. Kingdon J. The Kingdon field guide to African mammals. San Diego: Academic Press; 1997.

46. Rietsch A, Beckwith J. The genetics of disulfide bond metabolism. Annu Rev Genet. 1998;32: 163–84. Available: http://www.ncbi.nlm.nih.gov/pubmed/9928478

47. Weber ANR, Morse MA, Gay NJ. Four N-linked glycosylation sites in human toll-like receptor 2 cooperate to direct efficient biosynthesis and secretion. J Biol Chem. 2004;279: 34589–94. Available: http://www.ncbi.nlm.nih.gov/pubmed/15173186

48. Lemos de Matos A, McFadden G, Esteves PJ. Positive evolutionary selection on the RIG-I-like receptor genes in mammals. PLoS One. 2013;8: e81864. Available: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3842351&tool=pmcentrez&rendertype=abstract

49. Smith NGC. Are radical and conservative substitution rates useful statistics in molecular evolution? J Mol Evol. 2003;57: 467–78. Available: http://www.ncbi.nlm.nih.gov/pubmed/14708579

50. Fowler ME. Husbandry and diseases of captive wild swine and peccaries. Rev Sci Tech. 1996;15: 141–54. Available: http://www.ncbi.nlm.nih.gov/pubmed/8924700

51. Patel MR, Loo Y-M, Horner SM, Gale M, Malik HS. Convergent evolution of escape from hepaciviral antagonism in primates. PLoS Biol. 2012;10: e1001282. Available: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3302847&tool=pmcentrez&rendertype=abstract

52. Rothenburg S, Seo EJ, Gibbs JS, Dever TE, Dittmar K. Rapid evolution of protein kinase PKR alters sensitivity to viral inhibitors. Nat Struct Mol Biol. 2009;16: 63–70. Available: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3142916&tool=pmcentrez&rendertype=abstract

Chapter 3: Evidence for adaptation of porcine Toll-like receptors

Abstract

Naturally endemic infectious diseases provide selective pressures for pig populations. Toll-like receptors (*TLRs*) represent the first line of immune defense against pathogens and are likely to play a crucial adaptive role for pig populations. This study was done to determine whether wild and domestic pig populations representing diverse global environments demonstrate local *TLR* adaptation. The extracellular genomic sequence domains, responsible for interacting with pathogen ligands of bacterial (*TLR1*, *TLR2* and *TLR6*) and viral (*TLR3*, *TLR7* and *TLR8*) receptors were obtained. Mitochondrial D-loop region sequences were obtained and a phylogenetic analysis using these sequences revealed a clear separation of animals into Asian (n=27) and European (n=40) clades. The *TLR* sequences were then analyzed for population specific positive selection signatures within wild boars and domesticated pig populations derived from Asian and European clades. Using within population and between population tests for positive selection, a *TLR2* derived variant 376A (126Thr), estimated to have arisen in 163,000 years ago with a frequency of 83.33% within European wild boars, 98.00% within domestic pig breeds of European origin, 40.00% within Asian wild boars and 11.36% within Asian domestic pigs was identified to be under positive selection in pigs of European origin. The variant is located within the N terminal domain of the *TLR2* protein 3D crystal structure and could affect ligand binding. This study suggests the *TLR2* gene contributing to responses to bacterial pathogens has been crucial in adaptation of pigs to pathogens.

Introduction

European and Asian wild boars diverged 1.6-0.8 million years (Myr) ago [1] with domestication events occurring approximately 9,000 years ago [2]. European wild boars and European domesticated pigs and their Asian counterparts may have experienced different selective pressures given that they have originated from different continents with different endemic infectious diseases [3,4]. Such differences in selective pressures can be exploited to detect immune related genes that have been of adaptive value in terms of disease resistance within pig populations.

The vertebrate immune system is comprised of the innate and acquired immune components. The innate immune component is ancient and characterized by rapid and non-specific responses to pathogenic microbes while the acquired immune component is characterized by highly specific memory responses. The Toll-like receptor (*TLR*) family represents a class of innate immunity receptors important during early phases of infections that also serve as a link between the innate and acquired immune responses [5], consequently triggering inflammatory responses to prevent bacterial and viral spread. In vertebrates, 10-12 *TLRs* have been identified [6] and are classified based on their cellular location and the type of pathogenic ligands they recognize. Cell surface expressed *TLRs* (*TLR1*, *TLR2*, *TLR4*, *TLR5* and *TLR6*) recognize predominantly bacterial ligands and several fungal and parasite ligands while *TLR3*, *TLR7*, *TLR8* and *TLR9* are expressed within endosomes and recognize single and double-stranded RNA and CpG DNA [7]. Single nucleotide polymorphisms (SNPs) and differences in expression of *TLR* molecules affect immune responses to numerous pathogens and are associated with host resistance to infectious diseases [8]. Several non-synonymous substitutions have been detected in porcine *TLR* genes, especially in the extracellular domain of cell surface expressed *TLRs* [9], suggesting functional relevance of the extracellular domain and specific SNPs residing within these regions.

Naturally occurring porcine diseases endemic to particular geographic locations include African swine fever prevalent in sub-Saharan Africa; the ancient Foot and Mouth disease endemic in l parts of

Asia, most of Africa and the Middle East; Brucellosis observed in the Middle East, the Mediterranean region, sub-Saharan Africa, China, India, Peru, and Mexico; and Swine influenza which is endemic to North and South America, Europe and parts of Asia [10]. The occurrence of such locally or regionally endemic diseases suggests that long-term adaptation may have occurred between the host and pathogen [11]. Thus, such adaptation could be detected as genomic signatures across populations. Recent studies have shown that approximately 7% of the pig genome has been influenced by selection either in the form of positive or balancing selection [12]. These signals were observed in genomic regions involved in coat color, brain and muscle development, growth, metabolism, olfaction and immunity and were specific to certain pig populations [12]. Differences in amino acid substitutions and heterozygosity of *TLR* SNPs in European and Asian pigs have been observed and might be due to differences in pathogens encountered in the domestication and initial breed development environments of the pigs [13]. Porcine *TLRs* therefore represent candidate immune genes for investigating pathogen driven selective pressures specific to geographically defined populations.

The aim of this study was to determine whether adaptation to local environments of wild and domestic European and Asian pig populations resulted in *TLR* gene genomic signatures of selection. We hypothesized that geographically restricted selective pressures resulted in adaptive genomic signatures within porcine *TLRs*. To test this hypothesis, the extracellular domain of bacterial recognizing (*TLR1*, *TLR2* and *TLR6*) and viral sensing (*TLR3*, *TLR7* and *TLR8*) receptors were sequenced and the following specific questions were addressed: 1) whether adaptive selection pressures on pigs from both European and Asian environments was evident and 2) whether different selective pressures for each geographic environment were identified.

Materials and Methods

Study populations

Sixty seven (67) individuals consisting of wild boars and domestic pigs representing populations with origins from diverse geographic environments within Asia and Europe were utilized in this study.

European wild boars were represented by 11 unrelated individuals from the Netherlands, 2 from Greece and 1 each from Switzerland and Italy. European domestic breeds were sampled from unrelated pigs including Angler Sattleschwein (1), Mangalitsa (1), Durocs (4), Hampshires (3), Landraces (6), Pietrains (4), Charto Murciano (2), Cassertana (2), Retinto (1) and Cinta Senese (1). Asian wild boars were represented by individuals from Northern China (3), from Southern China (1) and from Japan (1). Finally Asian domestic pigs included Meishan pigs (10), Jiangquahai pigs (3), Xiang pigs (2), Wannan spotted pigs (2), Jinhua pigs (2), Leping spotted pigs (2) and Zang pig (1). Animals within a group were not related (no shared grandparents). In addition, one species individual from the family Suidae (*Phacochoerus africanus* (Common warthog), *Sus verrucosus* (Javan warty pig) or *Sus celebensis* (Sulawesi warty pig) were also analyzed. Fig. 3.1 shows the geographic sampling regions.

Loci Analyzed

Six *TLR* genes chosen for this study are described in Table 3.1. The extracellular domains of *TLRs* were examined since they encode the functional sites involved in pathogen ligand recognition. The D-loop region of mtDNA, a neutral marker was used to demonstrate the independent domestication origins of European and Asian pigs [14]. The D–loop region of the mtDNA of each animal was obtained by direct sequencing and analyzed to obtain clustering patterns among the sampled animals. On each of the 18 autosomal chromosomes of pigs, 1 noncoding genomic region (2 kb in size) (Table 3.2) was chosen to represent neutrally evolving regions. These noncoding genomic regions were at least 50 kb from any known or predicted porcine gene, as determined by inspecting the *Sus scrofa* 10.2 genome assembly on the UCSC genome browser. The 50 kb buffer was chosen to minimize the effect of linkage disequilibrium with genic SNPs [15].

DNA Sequencing

DNA extraction, library preparation and sequencing have been described previously [16]. Briefly, DNA was extracted from whole blood by using the QIAamp DNA blood spin kit (Qiagen Sciences) and

quantity and quality parameters were performed on the Qubit 2.0 fluorometer (Invitrogen) and run on a 1% agarose gel. Library construction and re-sequencing of individual pig samples utilized 1-3 ug of genomic DNA following Illumina library preparation protocols. The library insert size was 300-500 bp and sequencing was performed using the 100 paired-end sequencing kit. All DNA were sequenced to approximately 8x depth. Quality trimmed reads (phred quality>20, minimum length of pairs of reads = 40 bp) were aligned to the *Sus scrofa* reference genome build 10.2 using the unique alignment option of Mosaik Aligner (V.1.1.0017). The aligned reads from each of the animals together with the *Sus scrofa* reference genome were stored as bam files for each individual animal.

The coordinates of genes within the bam files were obtained by querying the *Sus scrofa* reference genome build 10.2 with coding sequences of porcine TLRs from public databases. The accession numbers of sequences obtained from the public databases for TLRs were *TLR1*: NM_001031775, *TLR2*: NM_213761, *TLR3*: HQ412796, *TLR6*: NM_213760, *TLR7*: NM_001097434, *TLR8*: ENSSSCG00000012118. Perl scripts were used to extract the TLR coding sequences and the noncoding genomic regions. Sequences were aligned using ClustalW 1.81 [17]. The D-loop region of mtDNA was obtained by direct sequencing.

A 715 bp fragment of the D-loop region was amplified by polymerase chain reaction (PCR). The nucleotide sequences of the primers used were: forward 5´CTCCGCCATCAGCACCCAAAG3´ and reverse 5´GCACCTTGTTTGGATTRTCG3´ [18]. Final reactions were made to a final volume of 12 µL containing each reaction PCR master mix (Thermo Fisher Scientific, USA) which supplies 1.5 mM $MgCl_2$, 0.625 U of Taq DNA polymerase and 0.2 mM dNTPs. Then, 5 pmolar of each primer and ~40 ng DNA were added to the final reaction. Amplification protocol was performed as follows: 35 cycles, each consisting of 95ºC for 30s, 55ºC for 45s and 72º for 90s and then a final extension at 72ºC for 10 min. The amplicon was confirmed using 2 µL PCR product by electrophoresis on a 1.5% agarose gel stained with ethidium bromate visualized under UV light . The remaining PCR products were purified using a Multiscreen PCR 96 cleanup vacuum system (Millipore). Bidirectional sequencing reactions were carried out using Big-Dye Terminator Cycle Sequencing Ready Reaction kit v3.1 (Applied Biosystem, USA) in

an ABI3730 DNA Analyzer sequencer (Applied Biosystem, USA). In order to correct sequencing errors and obtain the consensus sequence for each individual amplicon, GAP 4 was utilized [19], using the mtDNA pig sequence GenBank AJ002189 [20] as a reference sequence.

Data analysis

Porcine TLR reference amino acid sequences were aligned to the corresponding human sequences in order to delineate the extracellular domains, the leucine rich repeat (LRR) modules, ligand-binding and dimerization domains, and other sub-domains within the extracellular domain (Fig. 3.2). Haplotype reconstructions from the aligned sequences for all loci were carried out with PHASE 2.1.1 software [21] using the SNiPLAY web-based tool for SNP and polymorphism analysis [22]. For the D-loop region sequences of mtDNA, the Neighbor-Joining method (using p-distance) implemented in MEGA version 5 [23] was used for phylogenetic analysis (Fig. 3.3). A bootstrap of 1000 replicates was conducted. To test for departure from the standard neutral model of evolution, Tajima's D, Fu and Li's D* and Fu and Li's F* were conducted using DnaSP [24].

Derived allele under positive selection was determined using the DIND test. The DIND test was applied by plotting for all SNPs within the extracellular domain of genes within groups that showed deviation from neutrality, the ratio between the ancestral and derived internal nucleotide diversity (diversity among haplotypes carrying alleles) against the frequencies of derived alleles. An elevated ratio associated with a high derived allele frequency was used as an indication of positive selection of the derived allelic state. To define statistical significance, the values estimated for *TLRs* were then compared against the background neutral distribution obtained by means of 10,000 simulations of the extracellular domain conditional on the number of segregating sites and the recombination rate of the extracellular domain and integrating a simplified version of a wild boar demographic model (initial effective population sizes of European and Asian wild boars = ~25000, a bottleneck at 20000 yrs ago and an effective population size 10000 yrs ago of 3000 for European wild boars and 13500 for Asian wild boars) [25]. Simulations were carried out using Fastsimcoal2 [26]. *TLR* SNPs that fell beyond 90^{th} and 95^{th}

percentiles of the neutral distribution were considered to be under positive selection. Singletons were excluded from this analysis.

Interpopulation differention (F_{ST}) and expected loci heterozygosity were calculated for *TLR* extracellar domain SNPs that were polymorphic in each of two populations being compared and intergenic region sequences SNPs using Arlequin ver 3.5 [27]. *TLR* SNPs showing high levels of population differention and thus, the target of positive selection were indentified by comparison of *TLR* SNP F_{ST} values and the 90^{th} and 95^{th} percentiles for F_{ST} distribution (estimated using heterozygosity sliding windows of size 0.025 with increasing step of 0.01) computed for SNPs of the noncoding genomic regions. The p value for a SNP was estimated [28] where first, F_{ST} values for a *TLR* SNP was compared with F_{ST} values from the noncoding genomic region sequences SNPs with an expected heterozygosity value of ±0.025 with respect to that observed for the *TLR* SNP. Then among the noncoding genomic region SNPs, the proportion of SNPs with F_{ST} values higher than that observed for the *TLR* SNP was used as the p value. Ancestral and derived states of *TLR* alleles, were determined by a strategy [25] where an allele is assumed to be ancestral when one of the alleles in *Sus scrofa* was observed in *Phacochoerus africanus* (Common warthog), *Sus verrucosus* (Javan warty pig) or *Sus celebensis* (Sulawesi warty pig) in that order, respectively. A maximum likelihood approach implemented in GENETREE version 9 [29] was used to estimate theta ($\Theta = 4N_e\mu$) and age of mutations. The default mutation rate ($\mu = 2.5 \times 10^{-8}$) of humans was used as there is no known mutation rates for pigs [25]. Time estimated in generations (T) were converted into years (t) using a 5-year generation time (g) with the formula $t=2*N_e*T*g$ as stated in the GENETREE manual. Median-joining phylogenetic haplotype networks were constructed based on SNPs within extracellular domains of the TLRs using Network 4.6.1.1 (www.fluxus-technology.com). Only haplotypes present in a mimimum of two animals were considered. MuPIT Interactive [30] was used to map variants under selection on to three dimensional (3D) protein structures. Swiss prot and Ensemble genome browsers were utilized to determine the functional consequences of TLR sites under positive selection.

Results

Sequences (67 sequences for each TLR alignment) of the extracellular domains of bacterial sensing *TLR1*, *TLR2* and *TLR6* and viral sensing *TLR3, TLR7* and *TLR8* from wild boars and domestic pigs of European and Asian origins were obtained. The length of the extracellular domains in terms of number of nucleotides of the TLRs ranged from 1668 bases for *TLR1* to 2445 bases for *TLR7*. Amino acid length ranged from 556 amino acids for *TLR1* to 792 amino acids for *TLR7*. A total of 136 SNPs were identified within the TLR extracellular domains (Table 3.3).

Evidence of positive selection pressure within population

To determine whether there is evidence of positive selective pressure mediated by infectious agents of endemic diseases on pig populations, wild boars and domestic pigs from previously defined European and Asian lineages were chosen [31]. The geographic origins of these populations were validated by sequencing the mitochondrial D-loop regions of these animals and constructing a phylogenetic tree. The Neighbor Joining tree (Fig. 3.3) obtained from analysis of the D-loop region sequences revealed two clades of animals representing animals of European and Asian origins. For analysis, animals were grouped by geographic origins and domestication status. The groups considered were therefore all European animals (wild boars and domestic pigs combined, N = 40), all Asian animals (wild boars and domestic pigs combined, N = 27), Asian wild boars (N = 5), European wild boars (N = 15), Asian domesic pigs (N = 22) and European domestic pigs (N = 25). Gene sequences of bacterial sensing *TLR1*, *TLR2*, *TLR6* and viral sensing *TLR3*, *TLR7* and *TLR8* for each animal were extracted from whole genome resequenced data from each animal. Analysis of positive selection focused on the extracellular domains (Table 3.1) involved in pathogen recognition.

The following tests were performed to determine evidence of adaptive selection pressure for European and Asian pig populations: 1) test of deviation from neutrality (sliding window analysis of

Tajima's D, Fu and Li's D* and Fu and Li's F*) due to shift to a low frequency spectrum polymorphism; and 2) test for derived alleles under recent positive selection due to high frequency of the allele in a population [32] using the derived intrallelic nucleotide diversity (DIND) test [28]. Given the relatively small population sample sizes, genes under selection were defined conservatively as those for which both neutrality and DIND test were significant in the same population [33]. Using these stringent criteria, *TLRs* (*TLR1*, *3*, *6*, *7* and *8*) investigated in this study did not show signatures of adaptive selection (data not shown) in any population. Each of the three neutrality tests detected significant ($p<0.05$) excess of rare alleles within the bacterial sensing *TLR2* exon 2 (encoding the extracellular domain) of the European (wild boars and domestic pigs combined) population (Tajima's D = -1.80; Fu & Li's D* = -3.74; Fu & Li's F* = -3.67, Fig. 3.4) consistent with positive selection or population expansion. Based on the DIND test involving SNPs within the extracellular domain of *TLR2* of the European population, the derived allele *TLR2* SNP 376A (126Thr), located on exon 2 of the *TLR2* gene with a frequency of 92.5% (83.33% within European wild boars and 98.00% within Asian wild boars) was detected as showing evidence of positive selection ($\pi A/\pi D = 6.88$; $p = 0.055$, Fig. 3.5). Three European wild boars were heterozygous, 1 wild boar was homozygous for the ancestral allele whereas the 11 remaining wild boars were homozygous for the derived allele. One European landrace pig was heterozygous whereas all remaining European domestic pigs (N=24) were homozygous for the derived allele indicating that the different artificial selective pressures on the various breeds have not resulted in variation at this locus. The derived allele is at a frequency of 40.00% in Asian wild boars and 11.36% in Asian domestic pigs population. Of the 5 Asian wild boars used in this study, 1 northern Chinese wild boar was homozygous for the derived allele, 1 northern Chinese wild boar and 1 Japanese wild boar were homozygous for the ancestral allele whereas 1 southern Chinses wild boar and 1 northern Chinese wild boar were heterozygous. Seventeen Asian domestic pigs were homozygous for the ancestral allele. The remaining 5 Asian domestic pig breeds were heterozygous. Details on frequencies for TLR SNPs are as shown in Table 3.3. Determination of the ancestral or derived state of an allele is described in the materials and methods section. Fig. 3.6. shows a conservation of the G allele at *TLR2* position 376 within three wild pig

relatives. The G allele is the ancestral allele and the A allele is derived. The change from the ancestral to the derived allele at *TLR2* SNP 376 is a nonsynonymous change (*TLR2* SNP G376A, Ala126Thr) and is likely to affect protein function.

Selective pressure differences between populations

To determine whether selective pressures were heterogenous between populations, Fst values for SNPs within the extracellular domains of *TLRs* were compared to the empirical distribution of F_{ST} from SNPs obtained from noncoding genomic region sequences chosen to represent neutrally evolving regions (see Materials and methods). The rationale behind this approach was that differences in selective pressure between populations could lead to elevated levels of population differentiation at immune genes relative to neutrally evolving loci [28]. The nonsynonymous variant *TLR2* SNP G376A showed the highest level of population differentiation (Fst between European domestic pigs and Asian domestic pigs = 0.86, p = 0.02 (Fig. 3.7); Fst between European pigs (wild boars and domestic pigs combined) and Asian pigs (wild boars and domestic pigs combined) = 0.74, p = 0.08 (Fig. 3.7). A comparison of European wild boars and European domestic pigs indicated that *TLR8* SNPs were highly differentiated relative to other *TLR* SNPs (Fig. 3.7). However the location of TLR8 gene on the X chromosome means it is prone to higher genetic drift which may result in elevated levels of population differentiation for *TLR8* SNPs [28]. Fst values between European wild boars and European domestic pigs were low relative to those between pigs of European and Asian origins, indicating weak differentiation between pig populations from the same geographic origin. Asian wild boar population was not compared to any other population in terms of Fst given the small number of Asian wild boars involved in this study.

To determine whether positive selection drives the prevalence of certain *TLR* haplotypes in European and Asian pig populations, haplotype reconstruction involving SNPs within the extracellular domain of each *TLR* was done using the PHASE 2.1.1 software [21] and the evolutionary relationships between haplotypes for each *TLR* were determined using a median-joining network. For *TLR2*, two high frequency haplotypes were observed and differed only at *TLR2* SNP 376 (Fig. 3.8, Table 3.4), where the

high frequency haplotype (H_1) dominated by the European individuals carried the derived SNP 376A allele and the high frequency haplotype (H_8) dominated by Asian individuals carried the ancestral SNP G376 allele. The high frequency haplotype within the European population was detected in all the European domestic pigs and the majority (10/15) of European wild boars. In order to estimate the divergence time of *TLR2* extracellular domain region and the age of *TLR2* SNP G376A (Ala126Thr), maximum coalescence analysis by GENETREE [29] was utilized. Using all populations, the estimated time to most common ancestor (T_{MRCA}) for the entire *TLR2* extracellular domain geneology was 0.900 ± 0.28Mya, which is close to the ~ 1Mya since the split of the Asian and European wild boar, and the age of the 126Thr variant assuming neutrality was estimated to be 0.163 ± 0.08 Mya. The age of the derived allele and the presence of the haplotype carrying the derived allele in most European wild boars and all European domestic pigs involved in this study indicate that the allele arose within the wild boars, prior to the domestication process.

Functional relevance of SNPs under positive selection

To determine whether variants under selection are within functional domains of TLR receptors, variants were mapped onto *TLR* protein 3D structure. The *TLR2* variant 126Thr is located within the N-terminal (on the fourth leucine rich repeat (Fig. 3.2)) and alpha helices of the *TLR2* protein and is surface exposed (Fig. 3.9), indicating a likely role in protein-protein interactions. Porcine *TLR2* amino acid sequence were also compared to human *TLR2* amino acid sequence from Swiss-Prot to examine whether the site under positive selection was within functionally relevant domains of the protein. *TLR2* variant 126Thr did not fall within any known region of functional region. Lastly, the *TLR2* variation table in Ensembl was examined to determine the effects of substitutions at *TLR2* amino acid site 126 on protein function. SIFT predicts whether an amino acid substitution affects protein function based on sequence homology and the physico-chemical similarity between alternate amino acids [34]. Substitutions with scores < 0.05 are called ‘deleterious’ and those with scores > 0.05 are called ‘tolerated’. Within Ensembl, the substitution from Threonine to Alanine at porcine *TLR2* 126 (dbSNP identifier rs81218810) is

predicted by SIFT to be deleterious (score = 0.03), indicating that substitution at the *TLR2* amino acid site 126 affects protein function.

Discussion

This study supports the hypothesis that geographically restricted selective pressures on European and Asian wild boars that have diverged over 1mya have resulted in genomic signatures of adaptation in porcine *TLRs*. Specifically, *TLR2* SNP 376A (126Thr) of the extracellular domain of European pig (wild boars and domestic pigs combined) and not the Asian pig population showed evidence of positive selection, consistent with previous studies that have demonstrated that certain polymorphisms in *TLR2* extracellular domains of primates [35], wild rodents [36] and cattle [37] involved in pathogen recognition have evolved adaptively confering selective advantage. Porcine *TLR2*, is found at the distal end of the q arm of Chromosome 8, in a region with identified QTLs for some immune related traits [38] and is therefore a potential target for positive selection. Furthermore, the association of *TLR2* with a wider panel of ligands and the need for heterodimerization (*TLR2* forms heterodimers with *TLR1* and *TLR6* to recognize lipopetide components of gram-positive and gram-negative bacterial cell walls) indicates that the *TLR2* region experiences contrasting evolutionary actions, including adaptive evolution to the environment and pathogens [37]. A comparison of the genomic coordinates [8:79,824,541-79,834,592] of porcine *TLR2* to recombination maps computed for 4 different pedigrees [39] indicate that Porcine *TLR2* lies in a region with a relatively low recombination rate on Chromosome 8. Regions with low recombination rate have been shown to be prone to positive selection [40].

Within population based tests for positive selection utilized here have previously been employed to detect SNPs under positive selection in immune-related genes (interferons and Toll-like receptors) within European and Asian human populations [28,33]. The positive selection of the derived allele *TLR2* SNP 376A (126Thr) in European pig population based on within population tests is likely due to selective pressure mediated by bacterial infectious agents encountered by the European wild boars following

divergence from their Asian counterparts. Furthermore, the conservation of the ancestral allele across the three wild pig relatives, none of which have origins in the European continent, suggests that the derived allele has arisen in response to selective pressure pertaining to Europe. In humans, clinical genetic studies have indicated a role of *TLR2* coding region polymorphisms in immune response to bacteria [41,42], demonstrating the action of selective pressure of bacterial origin on *TLR2* from a clinical perspective.

The estimated age of the *TLR2* 126Thr variant, its presence in European wild boars, European domestic pigs, Asian wild boars and Asian domestic in this study supports a situation where the selective pressure may have been of an ancient nature and present prior to the domestication process. The variant is likely to be of adaptive value to both European wild boars and their domesticated counterparts as they share some common bacterial agents [43], some of which may have persisted over extended periods. In this regard, one can expect that the selective pressure experienced by the European wild boars will persist in the domestic pigs. This may partly explain the high frequency of the derived allele in both European wild boars and domestic pigs. Consistent with our observation of high frequency of a derived allele in both European wild boar and domestic pig probably in response to pathogen mediated selective pressure, genetic variability in wild boar populations have been detected to be preserved in local domestic breeds at genomic sites with potential phenotypic effects [44]. The European domestic pig breeds involved in this study included both local and commercial breeds that have experienced artificial selective pressures of different intensities. For example, Charto Murciano, Cinta Senese, Cassertana and Mangalitsa are local breeds and are not subjected to intense artificial selective pressures as experienced by commercial pigs like Pietrain, Duroc and Landrace [44]. These differences have however not resulted in much variation at the locus under positive selection as apart from one landrace breed that was heterozygous, all other European domestic pigs were homozygous for the derived allele. Therefore, as mentioned earlier, the selective pressure responsible for the positive selective might have been in place before the domestication process. The estimated age of 163,000 ± 80,000 years of the derived allele under positive selection coincides with the late middle Pleistocene periods when there was an initiation of European wild boar expansion (about 190,000 years ago [45]) and when *Sus scrofa* was spreading from southeastern to

northern eastern area of the Asian continent (about 140,000 to 253,000 years ago [46]). This would have provided the necessary environment for the spread of infectious diseases that would lead to adaptation at host genes. As the derived allele under selection is found in both European and Asian pig populations, determining its origin is a bit of a challenge. Our analysis revealed high Fst values for certain *TLR* alleles between European and Asian pig populations with the highest genetic differentiation detected for *TLR2* SNP G376A (Ala126Thr). This may be due to different selective pressures associated with each population given that historically, different continental populations have have been exposed to different infectious agents [3,4]. *TLR2* mediates host immune response to gram positive bacteria and in the case of pigs, gram positive bacteria challenges peculiar to specific continents have been documented. For example, the methiclin-resistant *staphylococcus aureus* sequence type (ST) 398 have been detected to be highly prevalent among pigs in Europe and North America whereas ST9 is predominant in Asia [47–49]. Such heterogenous selectives pressures across populations can result in positive selection for resistance alleles in certain populations. A similar approach of genetic differentiation (Fst) between populations has been used to detect geographically restricted adaptation at type III interferons in European and Asian human populations [33]. In a comparison of European and Chinese pig populations utilizing Fst outlier tests, the *TLR4* C7485C have been shown to be under positive selection [50]. Two haplotypes with highest frequencies are differentiated at *TLR2* SNP 376, (Ala126 for major Asian haplotype and 126Thr for the major European haplotype), further supporting a possible role of positive selection at this site.

Even though the Swiss Prot database did not indicate that *TLR2* site 126 is within a functionally relevant region of the *TLR2* protein, the LRR4 within which *TLR2* site 126 is located may be of functional relevance given that it contains *TLR2* site 136, where amino acid substitution (Pro136Ala) is associated with the prevalence of pneumonia in pigs [9]. Therefore one can speculate that *TLR2* Ala126Thr can be of medical relevance to porcine diseases. Non-synonymous SNPs in LRR domains have been suggested to dramatically alter the ability of the molecule to identify extracellular pathogens [51]. The nonsynonymous nature of the *TLR2* SNP G376A substitution, which causes a change of amino acid property from a non-polar to a polar amino acid (Ala126Thr) suggests that the substitution may be important in adaptation of

European pigs. The location of *TLR2* 126Thr within the alpha helices of the 3D structure of *TLR1/TLR2* complex and at the N-terminal domain of the *TLR2* protein suggests it it is important for ligand detection for a variety of ligands including lipoteichoic acid and peptidoglycan [52,53]. The 'deleterious' nature of the threonine to alanine substitution at *TLR2* site 126 as predicted by SIFT further suggests this site is functionally relevant.

Results presented here suggest a role of pathogen mediated selective pressures among pig populations in driving the differentiation at *TLR2* SNP G376A (Ala126Thr). Future experimental functional analyses are required to determine how such SNP variant affects porcine immune response. A recent study [36] involving a wild rodent population has identified an association between *Borrelia* infection and haplotypes carrying the variants Ala and Thr (*TLR2* Thr276Ala) located within the extracellular domain [36]. The study of Tschirren et al. [36] thus has demonstrated a role of alanine-threonine substitutions within *TLR2* in infectious diseases.

Conclusions

In conclusion, this study provides evidence, based on within and between population tests, that European wild boars and domestic pigs show evidence of adaptation which is reflected in *TLR2* as signatures of selection, whereas no such evidence was observed in Asian wild boars and domestic breeds. Thus, our study suggests that *TLR2* 126Thr present in European wild boars, European domestic pigs, Asian wild boars and Asian domestic pigs has evolved under positive selection within the European pigs involved in this study, probably in response to pathogen mediated selective pressures. Experimental studies designed to investigate the role of the *TLR2* 126Thr in ligand binding and subsequent immune response are needed.

Acknowledgements

We are thankful to Dr. Laurie Rund of the Animal Sciences department of University of Illinois and Dr. Greger Larson of the School of Archaeology of Oxford University for reading the manuscript and making useful suggestions. We are grateful to J. M. Herrero-Medrano of the Wageningen University for providing mitochondrial D-loop sequences. We are also thankful to Dr. Guillaume Laval of the Unit of Human Evolutionary Genetics, Centre National de la Recherche Scientific, Paris, France for providing scripts for the detection of Fst outliers and DIND test. This work was funded by US Department of Agriculture (USDA) Agriculture Research Service (ARS) Grant 58-5438-2-307 to LBS.

Figures and Tables

Fig. 3.1. Geographic locations from where animal samples were obtained.

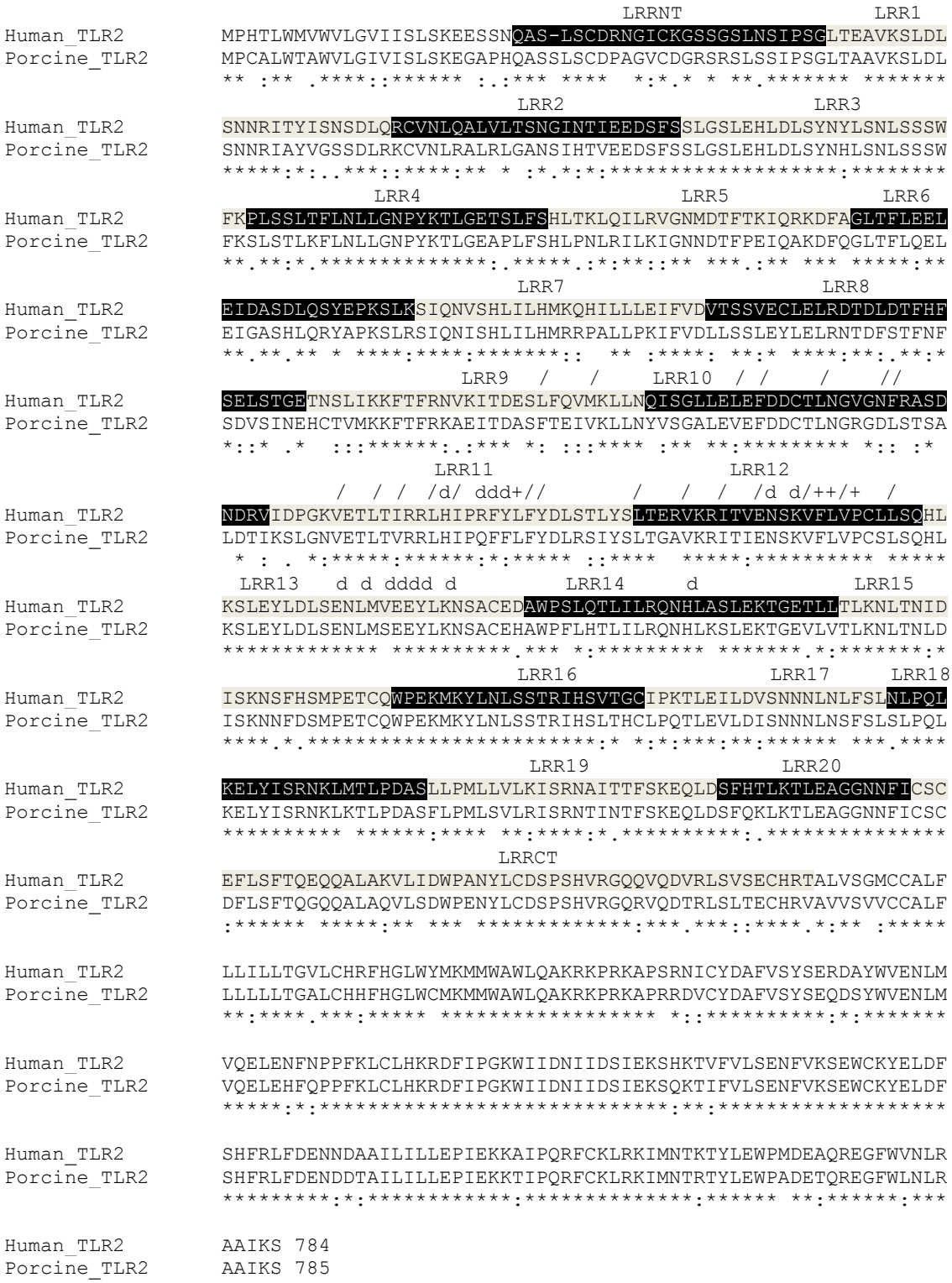

Fig. 3.2. Alignment of porcine *TLR2* amino acid sequences and human *TLR2* amino acid sequences to delineate LRRs, and functional domains of porcine TLRs. / ligand binding residues, d residues involved in dimerization, + residues involved in both ligand binding and dimerization. Asterisks, colons and periods under the aligned sequences indicate complete match, strong conservation and weaker conservation of amino acids respectively.

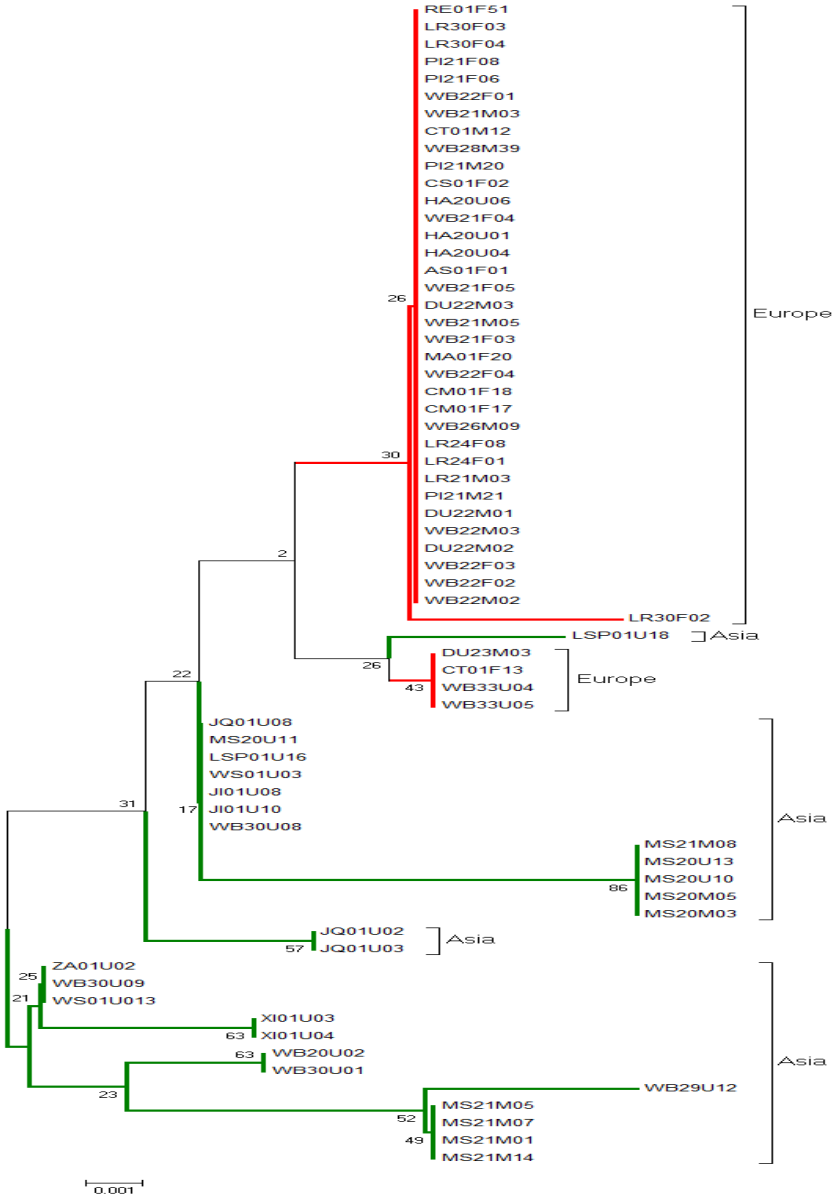

Fig. 3.3. Neighbor-joining phylogeny of the partial D-loop region sequences of the mitochondria DNA. Red branches represent pigs of European origin and green branches represent pigs of Asian origin.

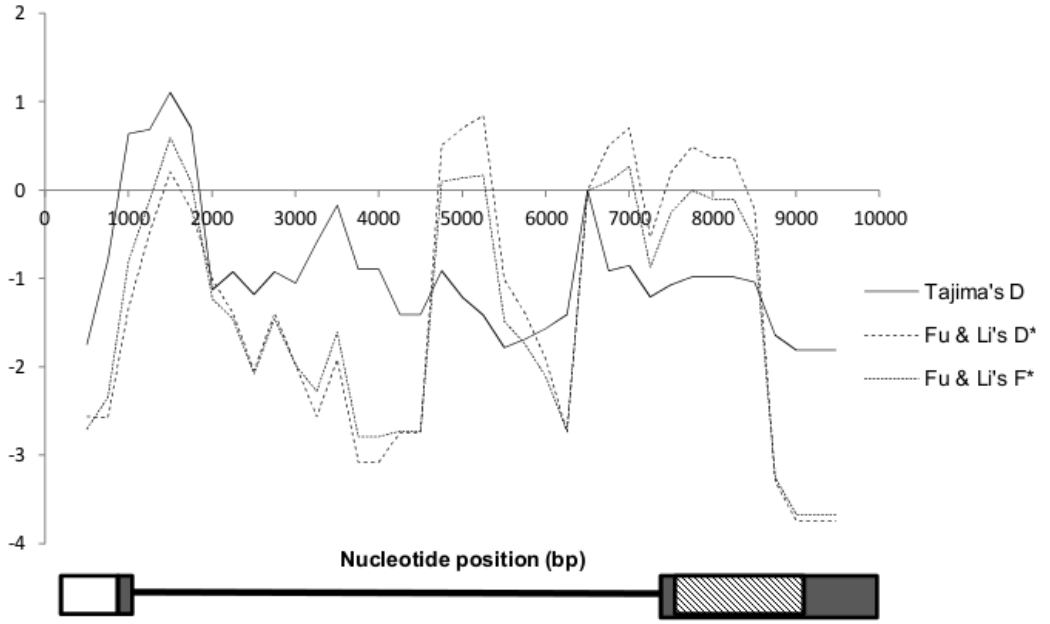

Fig. 3.4. Sliding window plots for *TLR2* gene. Sliding window plots of Tajima's D, Fu & Li's F* and Fu & Li's D* tests within the European (wild boars and domestic pigs combined) porcine population using window length of 1000bp and step size of 250bp. The white box represents the 5 upstream region, the thick black line represent the intronic region, the grey boxes show the exonic regions and the extracellular domain is represented by the box with light down diagonal lines. The distal part of the extracellular domain show significant values for neutrality indices.

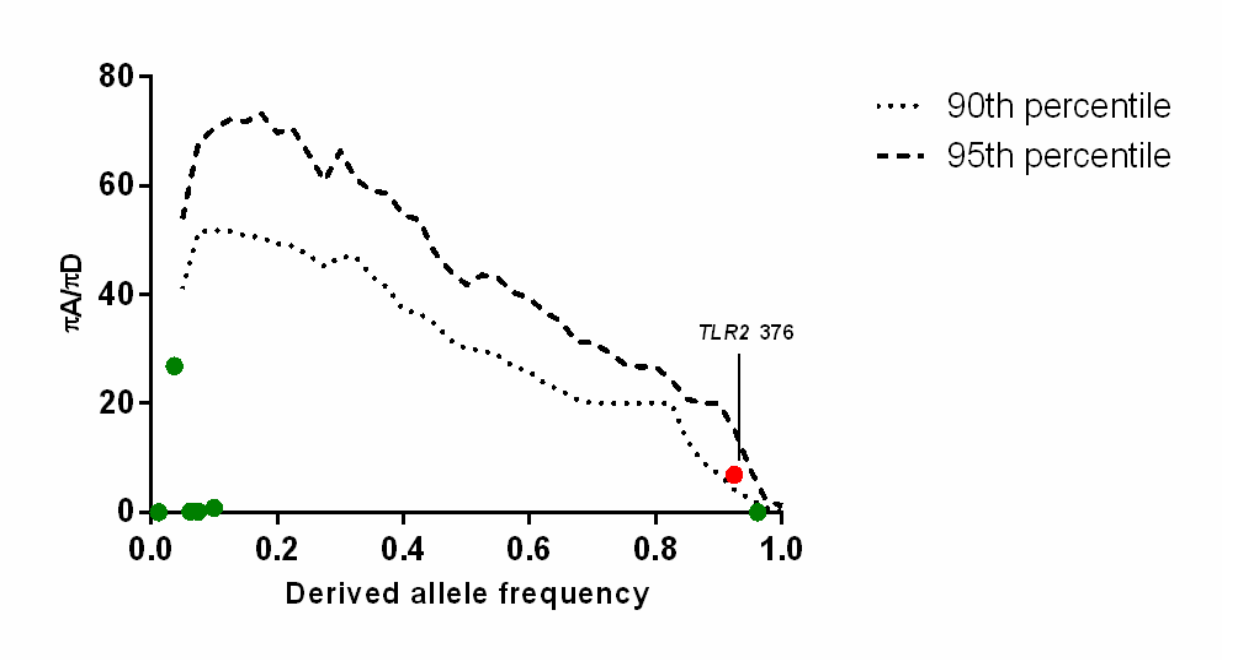

Fig. 3.5. Positive selection for *TLR2*. Detection of positive selection in European (wild boars and domestic pigs combined) porcine population *TLR2* extracellular domain using the DIND test. The short and long dashed lines represent the 90th and 95th percentiles of the empirical distribution of expected πA/πD values obtained from 10,000 simulations of the the extracellular domain conditional on the number of segregating sites and the recombination rate of the extracellular domain and integrating the wild boar demographic model previously described [25]. Green dots represent *TLR2* SNPs under neutrality and red dot represents *TLR2* SNPs under positive selection.

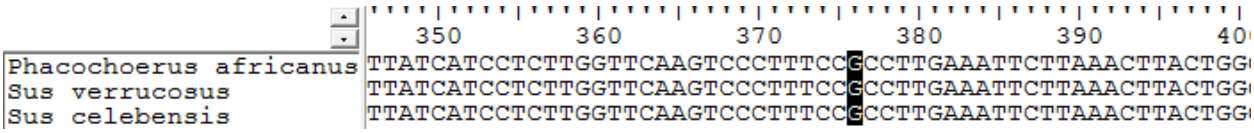

Fig. 3.6. Alignment of *TLR2* sequences of Suidae species. Illustrated is the G allele at nucleotide position 376 within *TLR2* as the ancestral allele.

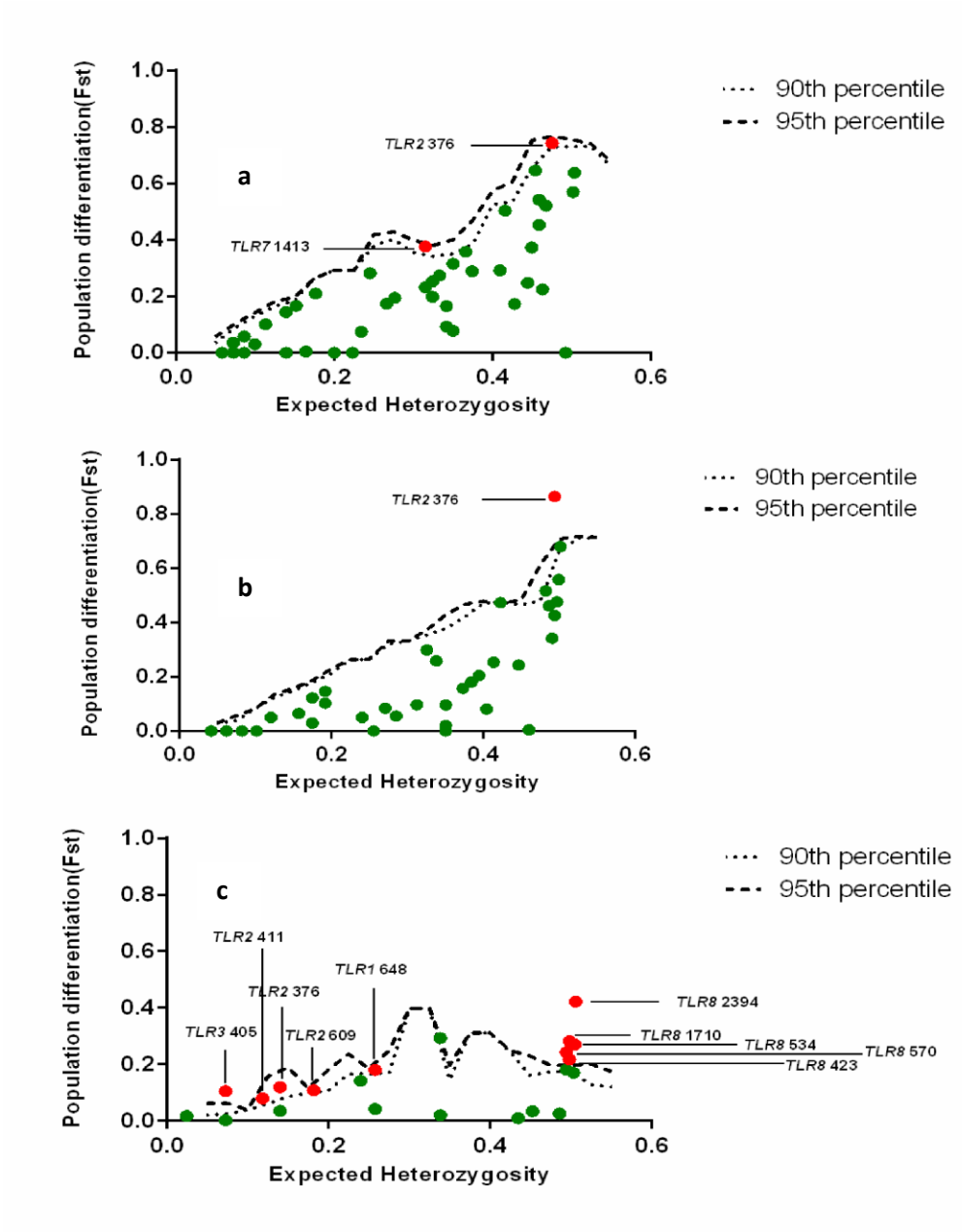

Fig. 3.7. Positive selection detection at *TLR* SNPs on the basis of population differentiation. F_{ST} is plotted as a function of expected heterozygosity for every SNP between (a); Asian (wild boars and domestic pigs combined) versus European (wild boars and domestic pigs combined) porcine populations (b); Asian domestic pig versus European domestic pig populations; and (c) European wild boar versus European domestic pig populations. Dots represent *TLR* SNPs. The short and long dashed lines represent the 90^{th} and 95^{th} percentiles of the empirical SNP distribution of F_{ST} of The noncoding genomic region sequences.

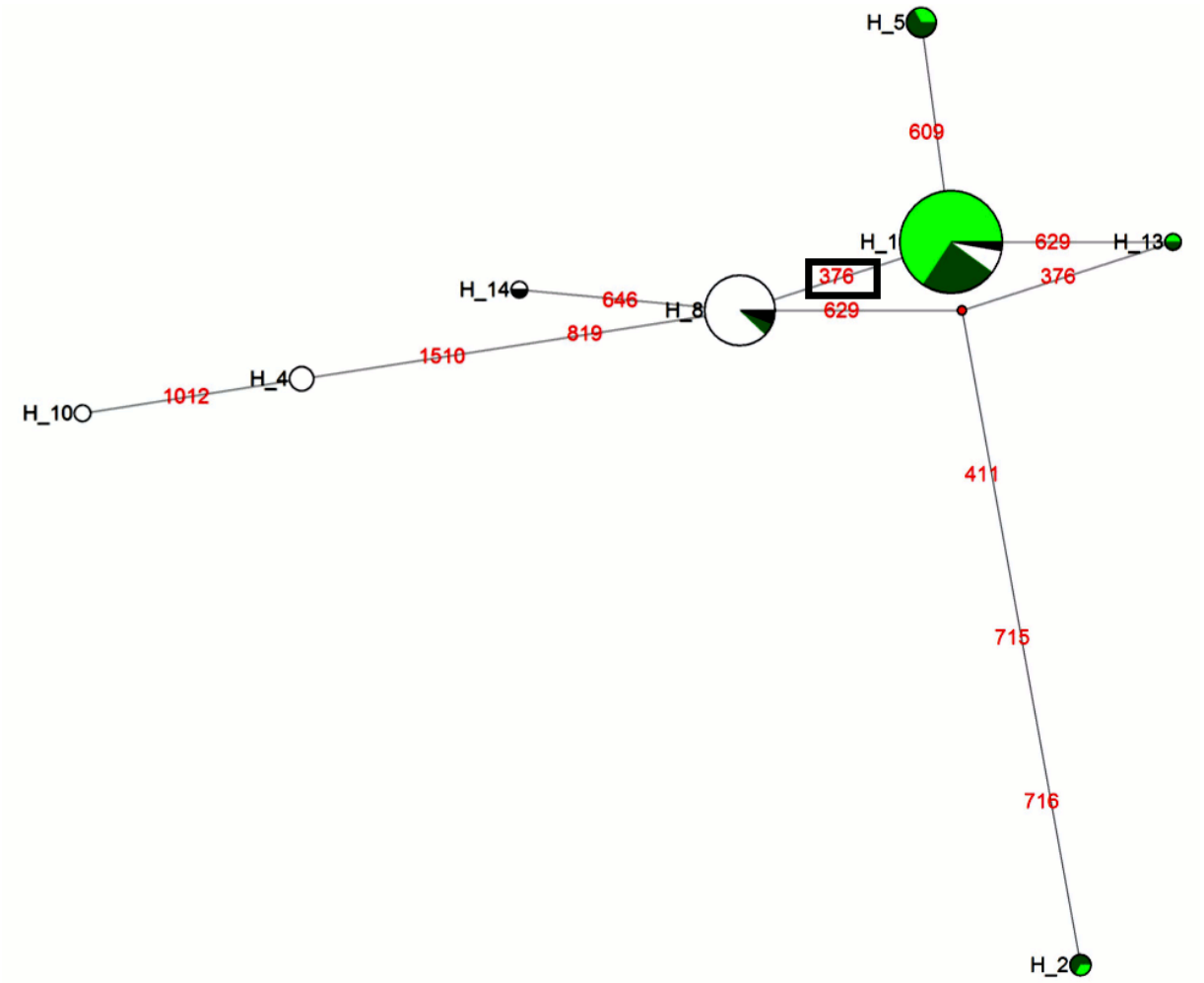

Fig. 3.8. Median-joining network for haplotypes at the *TLR2* extracellular domain. The circles represent haplotypes and the size of the circles are proportional to the frequency of the haplotypes. The mutation positions are shown along branches as red numbers. Asian wild boars, Asian domestic pigs, European wild boars and European domestic pigs populations are shown in black, white, deep green and light green colors respectively. SNP 376, under positive selection is boxed and differentiates the Asian and European populations.

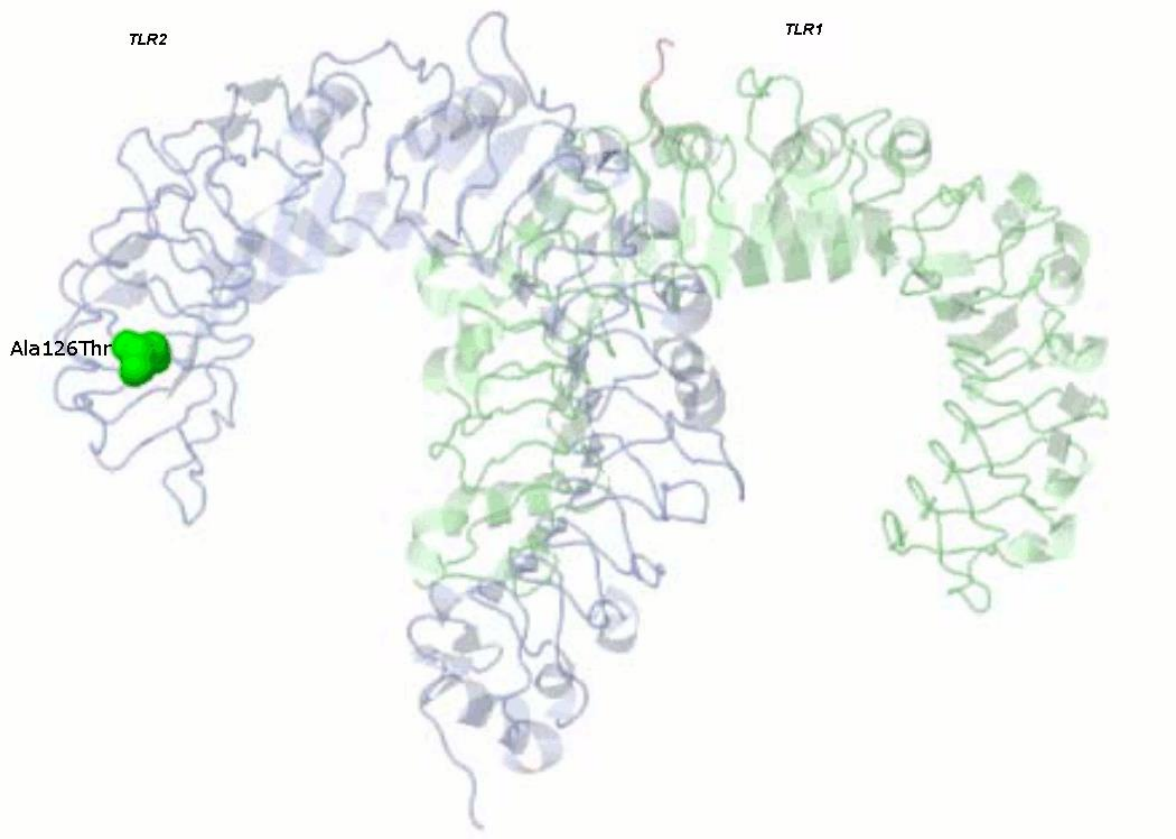

Fig. 3.9. Location of amino acid residues under selection in *TLR2* 3D protein structure. Amino acid residue *TLR2* 126Thr under positive selection within the European (wild boars and domestic pigs combined) porcine population is shown within the crystal structure of *TLR1/TLR2* heterodimer (PDB ID 2Z7X) induced by binding of a tri-acylated lipopeptide.

Table 3.1

Gene	**Exon no**[a]	**Genomic coordinates of extracellular domain**	**Aligned length (bp) of extracellular domain sequences**
TLR1	3	8:31628613-31630280:-1	1668
TLR2	2	8:79825324-79827018:-1	1695
TLR3	4	15:53849144-53849505:-1	366
TLR3	5	15:53848182-53848373:-1	192
TLR3	6	15:53841845-53843311:-1	1467
TLR6	2	8:31642930-31644612:-1	1683
TLR7	2	X:10472926-10475370:1	2445
TLR8	4	X:10509874-10512249:1	2376
D-loop region of mtDNA			598

Table 3.1. Summary of TLR extracellular domains studied

[a]Exon encoding extracellular domain

Table 3.2

Control sequence	**Genomic coordinates***
Chr1	1: 196113933-196115932
Chr2	2: 24343289-24345288
Chr3	3: 22621121-22623120
Chr4	4: 25952645-25954644
Chr5	5: 56514331-56516330
Chr6	6: 106599095-106601094
Chr7	7: 75443317-75445316
Chr8	8: 101852934--101854933
Chr9	9: 98181590-98183589
Chr10	10: 41491552-41493551
Chr11	11: 32898992-32900991
Chr12	12: 30205604-30207603
Chr13	13: 100433088-100435087
Chr14	14: 45367685-45369684
Chr15	15: 10096184-10098183
Chr16	16: 31646612-31648611
Chr17	17: 7579504-7581503
Chr18	18: 38554557-38556556

Table 3.2. Genomic coordinates of noncoding genomic region
*****The genomic coordinates are on the positive strand.

Table 3.3

Gene	SNP name	Allele frequency[a]			
		Wild boars		Domestic pigs	
		EWB (N=15)	AWB (N=5)	ED (N=25)	AD (N=22)
TLR1	A156G	100.00	60.00	88.00	81.82
	C204T	100.00	90.00	100.00	90.91
	A329T	100.00	100.00	98.00	100.00
	C338T	100.00	100.00	100.00	68.18
	T350C	100.00	50.00	90.00	75.00
	G391A	100.00	90.00	100.00	97.73
	A403G	100.00	100.00	100.00	77.27
	C489T	73.33	100.00	94.00	100.00
	G505A	100.00	90.00	100.00	97.73
	T519C	100.00	90.00	100.00	97.73
	T532G	100.00	90.00	100.00	97.73
	C537T	100.00	100.00	100.00	95.46
	C648T	70.00	90.00	94.00	61.36
	G649A	100.00	90.00	100.00	97.73
	G697A	100.00	100.00	100.00	95.46
	T798C	30.00	100.00	46.00	97.73
	C840T	100.00	60.00	88.00	59.09
	G855A	100.00	60.00	88.00	61.36
	C900T	100.00	100.00	100.00	61.36
	C966T	100.00	50.00	88.00	52.27
	A1095G	100.00	100.00	100.00	95.46
	C1278C	100.00	50.00	88.00	56.82
	T1305C	100.00	100.00	100.00	95.46
	C1319T	100.00	100.00	100.00	70.45
	A1351G	100.00	70.00	88.00	61.36
	A1373G	100.00	60.00	98.00	81.82
	T1499C	100.00	80.00	86.00	97.73
	A1579G	100.00	100.00	94.00	95.46
	G1636C	100.00	80.00	100.00	100.00
	G1641A	100.00	80.00	96.00	93.18
	C1647T	100.00	100.00	100.00	72.73
	A1657G	100.00	100.00	96.00	61.36
	T1669A	100.00	100.00	96.00	95.46
	G1675A	100.00	90.00	100.00	38.64
	C1695T	100.00	60.00	90.00	70.45
TLR2	A159G	100.00	100.00	100.00	97.73
	C198A	100.00	80.00	100.00	100.00
	C375T	100.00	80.00	100.00	100.00
	A376G	83.33	40.00	98.00	11.36
	C406G	100.00	70.00	100.00	97.73
	C411T	86.67	100.00	98.00	100.00
	C570T	100.00	100.00	100.00	97.73
	T609C	80.00	100.00	96.00	100.00
	G629C	86.67	100.00	96.00	100.00
	A646G	100.00	70.00	100.00	95.46

Table 3.3 (cont.)

Gene	SNP name	Allele frequency[a]			
		Wild boars		Domestic pigs	
		EWB (N=15)	AWB (N=5)	ED (N=25)	AD (N=22)
TLR2	A715C	93.33	100.00	98.00	100.00
	A716G	93.33	80.00	98.00	97.73
	G819C	100.00	90.00	100.00	84.09
	G1012A	100.00	90.00	100.00	93.18
	A1170G	96.67	100.00	100.00	100.00
	A1294G	100.00	80.00	100.00	97.73
	C1475T	100.00	100.00	100.00	97.73
	T1494C	100.00	90.00	100.00	97.73
	G1510C	100.00	100.00	100.00	84.09
	A1549C	100.00	90.00	100.00	100.00
	G1739T	96.67	100.00	100.00	100.00
	G1746A	100.00	100.00	100.00	97.73
TLR3	G95A	100.00	100.00	100.00	97.73
	G153A	100.00	80.00	100.00	100.00
	T159C	100.00	80.00	100.00	100.00
	A405T	90.00	100.00	100.00	100.00
	C798A	97.73	100.00	100.00	100.00
	C800T	97.73	100.00	100.00	100.00
	A1116T	100.00	70.00	82.00	61.36
	T1479C	100.00	100.00	98.00	100.00
	C1647T	100.00	30.00	82.00	29.55
	C1722A	100.00	100.00	98.00	100.00
	G1857A	100.00	100.00	98.00	100.00
	G1872A	100.00	80.00	84.00	47.73
TLR6	C133T	100.00	80.00	100.00	97.73
	G228A	100.00	80.00	76.00	93.18
	A266G	100.00	30.00	66.00	11.36
	C341T	100.00	100.00	100.00	95.46
	G663T	56.67	100.00	92.00	100.00
	A882G	100.00	80.00	98.00	95.46
	G919C	100.00	80.00	100.00	100.00
	C931T	100.00	100.00	98.00	100.00
	G977A	100.00	100.00	98.00	100.00
	C1061T	100.00	90.00	76.00	90.91
	C1124G	96.67	100.00	100.00	100.00
	A1259G	100.00	70.00	88.00	15.91
	G1284A	100.00	80.00	80.00	93.18
	T1329C	100.00	80.00	100.00	100.00
	A1354G	100.00	80.00	82.00	97.73
	G1376A	100.00	20.00	78.00	77.27
	G1391A	100.00	100.00	98.00	100.00
	G1438C	100.00	90.00	86.00	95.46
	C1502T	86.67	0.00	74.00	11.36
	A1660G	100.00	80.00	98.00	95.46
	G1678A	100.00	80.00	88.00	29.55
	T1698C	93.33	60.00	80.00	34.09
	G1729A	100.00	80.00	100.00	100.00

Table 3.3 (cont.)

		Allele frequency[a]			
		Wild boars		Domestic pigs	
Gene	**SNP name**	**EWB (N=15)**	**AWB (N=5)**	**ED (N=25)**	**AD (N=22)**
TLR7	G129A	100.00	100.00	100.00	94.12
	A357G	100.00	85.71	100.00	73.53
	C465T	100.00	100.00	100.00	97.06
	C520T	100.00	71.43	100.00	100.00
	T663G	100.00	100.00	97.56	100.00
	T792C	100.00	85.71	100.00	97.06
	C936T	100.00	100.00	100.00	97.06
	C1019A	78.26	71.43	65.85	100.00
	A1031G	100.00	100.00	100.00	94.12
	A1129C	100.00	100.00	100.00	97.06
	G1319A	100.00	100.00	100.00	82.35
	T1413C	91.30	57.14	100.00	52.94
	C1479A	100.00	85.71	100.00	100.00
	C1633T	100.00	100.00	100.00	88.24
	C1914T	100.00	100.00	100.00	82.35
	T1917C	100.00	100.00	100.00	97.06
	G2034A	100.00	85.71	97.56	76.47
	C2160T	100.00	85.71	97.56	79.41
	A2232G	78.26	100.00	63.41	100.00
TLR8	C99T	65.22	100.00	31.71	100.00
	G124A	100.00	100.00	100.00	79.41
	G156T	100.00	85.71	100.00	88.24
	C177A	100.00	100.00	100.00	79.41
	T199C	34.78	57.14	73.17	100.00
	A273C	100.00	100.00	97.56	100.00
	C309T	100.00	100.00	100.00	88.24
	A423G	34.78	100.00	73.17	100.00
	A534C	26.09	85.71	70.73	88.24
	A570T	30.43	100.00	75.61	85.29
	C636T	100.00	100.00	100.00	64.71
	G864T	100.00	71.43	97.56	94.12
	T907C	39.13	42.86	73.17	61.76
	T1150A	100.00	85.71	100.00	100.00
	C1281T	100.00	57.14	100.00	100.00
	A1593G	100.00	100.00	100.00	88.24
	T1605C	100.00	71.43	97.56	88.24
	G1647A	100.00	85.71	100.00	100.00
	T1710A	69.57	85.71	26.83	100.00
	C1740T	100.00	85.71	100.00	52.94
	A2144G	95.65	100.00	100.00	100.00
	C2338T	100.00	85.71	100.00	82.35
	A2394G	82.61	71.43	31.71	88.24
	C2397T	100.00	85.71	100.00	82.35
	G2439A	100.00	100.00	90.24	100.00

Table 3.3. Polymorphic positions in TLR1, TLR2, TLR3, TLR6, TLR7 and TLR8 in wild boars and domestic pigs
[a]The frequency of the first allele. The number in SNP name indicates nucleotide position of SNPs within TLR coding sequences. EWB European wild boar, AWB Asian wild boar, ED European domestic pig, AD Asian domestic pigN number of animals

Table 3.4

Haplotype	**SNPs**									
	376	411	609	629	646	715	716	819	1012	1510
H_1 (70)	A	C	T	G	A	A	A	G	G	G
H_2 (3)	G	T	T	C	A	C	G	G	G	G
H_4 (4)	G	C	T	G	A	A	A	C	G	C
H_5 (6)	A	C	C	G	A	A	A	G	G	G
H_8 (33)	G	C	T	G	A	A	A	G	G	G
H_10 (2)	G	C	T	G	A	A	A	C	A	C
H_13 (2)	A	C	T	C	A	A	A	G	G	G
H_14 (2)	G	C	T	G	G	A	A	G	G	G

Table 3.4. Haplotypes of TLR2 extracellular region
Numbers in brackets are frequencies of the haplotypes

References

1. Frantz LAF, Schraiber JG, Madsen O, Megens H-J, Bosse M, Paudel Y, et al. Genome sequencing reveals fine scale diversification and reticulation history during speciation in Sus. Genome Biol. 2013;14: R107. Available: http://genomebiology.com/2013/14/9/R107

2. Giuffra E, Kijas JM, Amarger V, Carlborg O, Jeon JT, Andersson L. The origin of the domestic pig: independent domestication and subsequent introgression. Genetics. 2000;154: 1785–91. Available: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=1461048&tool=pmcentrez&rendertype=abstract

3. Troy CS, MacHugh DE, Bailey JF, Magee DA, Loftus RT, Cunningham P, et al. Genetic evidence for Near-Eastern origins of European cattle. Nature. 2001;410: 1088–91. Available: http://www.ncbi.nlm.nih.gov/pubmed/11323670

4. Novembre J, Di Rienzo A. Spatial patterns of variation due to natural selection in humans. Nat Rev Genet. Nature Publishing Group; 2009;10: 745–55. Available: http://dx.doi.org/10.1038/nrg2632

5. Werling D, Jungi TW. TOLL-like receptors linking innate and adaptive immune response. Vet Immunol Immunopathol. 2003;91: 1–12. Available: http://www.ncbi.nlm.nih.gov/pubmed/12507844

6. Leulier F, Lemaitre B. Toll-like receptors--taking an evolutionary approach. Nat Rev Genet. 2008;9: 165–78. Available: http://www.ncbi.nlm.nih.gov/pubmed/18227810

7. Akira S, Uematsu S, Takeuchi O. Pathogen recognition and innate immunity. Cell. 2006;124: 783–801. Available: http://www.ncbi.nlm.nih.gov/pubmed/16497588

8. Lazarus R, Vercelli D, Palmer LJ, Klimecki WJ, Silverman EK, Richter B, et al. Single nucleotide polymorphisms in innate immunity genes: abundant variation and potential role in complex human disease. Immunol Rev. 2002;190: 9–25. Available: http://www.ncbi.nlm.nih.gov/pubmed/12493003

9. Uenishi H, Shinkai H, Morozumi T, Muneta Y, Jozaki K, Kojima-Shibata C, et al. Polymorphisms in pattern recognition receptors and their relationship to infectious disease susceptibility in pigs. BMC Proc. 2011;5 Suppl 4: S27. Available: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3108222&tool=pmcentrez&rendertype=abstract

10. OIE W organisation for animal health. General disease information sheets [Internet]. Available: http://www.oie.int/fileadmin/Home/eng/Media_Center/docs/pdf/Disease_cards/

11. Zimmerman J, Karriker L, Ramirez A, Schwartz K, Stevenson G, editors. Disease transmission and biosecurity. Diseases of Swine. 10th ed. Hoboken, NJ: Wiley-Blackwell; 2012. pp. 141–164.

12. Amaral AJ, Ferretti L, Megens H-J, Crooijmans RPMA, Nie H, Ramos-Onsins SE, et al. Genome-wide footprints of pig domestication and selection revealed through massive parallel sequencing of

pooled DNA. Ellegren H, editor. PLoS One. Public Library of Science; 2011;6: e14782. Available: http://dx.plos.org/10.1371/journal.pone.0014782

13. Shinkai H, Tanaka M, Morozumi T, Eguchi-Ogawa T, Okumura N, Muneta Y, et al. Biased distribution of single nucleotide polymorphisms (SNPs) in porcine Toll-like receptor 1 (TLR1), TLR2, TLR4, TLR5, and TLR6 genes. Immunogenetics. 2006;58: 324–30. Available: http://www.ncbi.nlm.nih.gov/pubmed/16604477

14. Larson G, Dobney K, Albarella U, Fang M, Matisoo-Smith E, Robins J, et al. Worldwide phylogeography of wild boar reveals multiple centers of pig domestication. Science. 2005;307: 1618–21. Available: http://www.ncbi.nlm.nih.gov/pubmed/15761152

15. Wagh K, Bhatia A, Alexe G, Reddy A, Ravikumar V, Seiler M, et al. Lactase persistence and lipid pathway selection in the Maasai. PLoS One. 2012;7: e44751. Available: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3461017&tool=pmcentrez&rendertype=abstract

16. Bosse M, Megens H-J, Madsen O, Paudel Y, Frantz LAF, Schook LB, et al. Regions of homozygosity in the porcine genome: consequence of demography and the recombination landscape. PLoS Genet. 2012;8: e1003100. Available: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3510040&tool=pmcentrez&rendertype=abstract

17. Thompson JD, Gibson TJ, Plewniak F, Jeanmougin F, Higgins DG. The CLUSTAL_X windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. Nucleic Acids Res. 1997;25: 4876–82. Available: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=147148&tool=pmcentrez&rendertype=abstract

18. Luetkemeier ES, Sodhi M, Schook LB, Malhi RS. Multiple Asian pig origins revealed through genomic analyses. Mol Phylogenet Evol. 2010;54: 680–6. Available: http://www.ncbi.nlm.nih.gov/pubmed/19900568

19. Bonfield JK, Smith K f, Staden R. A new DNA sequence assembly program. Nucleic Acids Res. 1995;23: 4992–9. Available: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=307504&tool=pmcentrez&rendertype=abstract

20. Ursing BM, Arnason U. Analyses of mitochondrial genomes strongly support a hippopotamus-whale clade. Proc Biol Sci. 1998;265: 2251–5. Available: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=1689531&tool=pmcentrez&rendertype=abstract

21. Stephens M, Scheet P. Accounting for decay of linkage disequilibrium in haplotype inference and missing-data imputation. Am J Hum Genet. 2005;76: 449–62. Available: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=1196397&tool=pmcentrez&rendertype=abstract

22. Dereeper A, Nicolas S, Le Cunff L, Bacilieri R, Doligez A, Peros J-P, et al. SNiPlay: a web-based tool for detection, management and analysis of SNPs. Application to grapevine diversity projects.

BMC Bioinformatics. 2011;12: 134. Available: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3102043&tool=pmcentrez&rendertype=abstract

23. Tamura K, Peterson D, Peterson N, Stecher G, Nei M, Kumar S. MEGA5: molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. Mol Biol Evol. 2011;28: 2731–9. Available: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3203626&tool=pmcentrez&rendertype=abstract

24. Librado P, Rozas J. DnaSP v5: a software for comprehensive analysis of DNA polymorphism data. Bioinformatics. 2009;25: 1451–2. Available: http://www.ncbi.nlm.nih.gov/pubmed/19346325

25. Groenen MAM, Archibald AL, Uenishi H, Tuggle CK, Takeuchi Y, Rothschild MF, et al. Analyses of pig genomes provide insight into porcine demography and evolution. Nature. 2012;491: 393–8. Available: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3566564&tool=pmcentrez&rendertype=abstract

26. Excoffier L, Dupanloup I, Huerta-Sánchez E, Sousa VC, Foll M. Robust demographic inference from genomic and SNP data. PLoS Genet. 2013;9: e1003905. Available: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3812088&tool=pmcentrez&rendertype=abstract

27. Excoffier L, Lischer HEL. Arlequin suite ver 3.5: a new series of programs to perform population genetics analyses under Linux and Windows. Mol Ecol Resour. 2010;10: 564–7. Available: http://www.ncbi.nlm.nih.gov/pubmed/21565059

28. Barreiro LB, Ben-Ali M, Quach H, Laval G, Patin E, Pickrell JK, et al. Evolutionary dynamics of human Toll-like receptors and their different contributions to host defense. PLoS Genet. 2009;5: e1000562. doi:10.1371/journal.pgen.1000562

29. Griffiths RC, Tavare S. Ancestral Inference in Population Genetics. Stat Sci. Institute of Mathematical Statistics; 1994;9: 307–319. Available: http://projecteuclid.org/euclid.ss/1177010378

30. Niknafs N, Kim D, Kim R, Diekhans M, Ryan M, Stenson PD, et al. MuPIT interactive: webserver for mapping variant positions to annotated, interactive 3D structures. Hum Genet. 2013;132: 1235–43. Available: http://www.ncbi.nlm.nih.gov/pubmed/23793516

31. Megens H-J, Crooijmans RPMA, San Cristobal M, Hui X, Li N, Groenen MAM. Biodiversity of pig breeds from China and Europe estimated from pooled DNA samples: differences in microsatellite variation between two areas of domestication. Genet Sel Evol. 40: 103–28. Available: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=2674914&tool=pmcentrez&rendertype=abstract

32. Rubin C-J, Megens H-J, Martinez Barrio A, Maqbool K, Sayyab S, Schwochow D, et al. Strong signatures of selection in the domestic pig genome. Proc Natl Acad Sci U S A. 2012;109: 19529–36. Available: http://www.pnas.org/content/109/48/19529.abstract

33. Manry J, Laval G, Patin E, Fornarino S, Itan Y, Fumagalli M, et al. Evolutionary genetic dissection of human interferons. J Exp Med. 2011;208: 2747–59. Available: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3244034&tool=pmcentrez&rendertype=abstract

34. Kumar P, Henikoff S, Ng PC. Predicting the effects of coding non-synonymous variants on protein function using the SIFT algorithm. Nat Protoc. Nature Publishing Group; 2009;4: 1073–81. Available: http://dx.doi.org/10.1038/nprot.2009.86

35. Takaki A, Yamazaki A, Maekawa T, Shibata H, Hirayama K, Kimura A, et al. Positive selection of Toll-like receptor 2 polymorphisms in two closely related old world monkey species, rhesus and Japanese macaques. Immunogenetics. 2012;64: 15–29. Available: http://www.ncbi.nlm.nih.gov/pubmed/21744114

36. Tschirren B, Andersson M, Scherman K, Westerdahl H, Mittl PRE, Råberg L. Polymorphisms at the innate immune receptor TLR2 are associated with Borrelia infection in a wild rodent population. Proc Biol Sci. 2013;280: 20130364. Available: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3619520&tool=pmcentrez&rendertype=abstract

37. Jann OC, Werling D, Chang J-S, Haig D, Glass EJ. Molecular evolution of bovine Toll-like receptor 2 suggests substitutions of functional relevance. BMC Evol Biol. 2008;8: 288. Available: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=2588590&tool=pmcentrez&rendertype=abstract

38. Bergman I-M, Rosengren JK, Edman K, Edfors I. European wild boars and domestic pigs display different polymorphic patterns in the Toll-like receptor (TLR) 1, TLR2, and TLR6 genes. Immunogenetics. 2010;62: 49–58. Available: http://www.ncbi.nlm.nih.gov/pubmed/19953243

39. Tortereau F, Servin B, Frantz L, Megens H-J, Milan D, Rohrer G, et al. A high density recombination map of the pig reveals a correlation between sex-specific recombination and GC content. BMC Genomics. 2012;13: 586. Available: http://www.biomedcentral.com/1471-2164/13/586

40. Barton NH. Genetic hitchhiking. Philos Trans R Soc Lond B Biol Sci. 2000;355: 1553–62. Available: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=1692896&tool=pmcentrez&rendertype=abstract

41. Kang T-J, Chae G-T. Detection of Toll-like receptor 2 (TLR2) mutation in the lepromatous leprosy patients. FEMS Immunol Med Microbiol. 2001;31: 53–58. Available: http://doi.wiley.com/10.1111/j.1574-695X.2001.tb01586.x

42. Schröder NWJ, Schumann RR. Single nucleotide polymorphisms of Toll-like receptors and susceptibility to infectious disease. Lancet Infect Dis. 2005;5: 156–64. Available: http://www.ncbi.nlm.nih.gov/pubmed/15766650

43. Meng XJ, Lindsay DS, Sriranganathan N. Wild boars as sources for infectious diseases in livestock and humans. Philos Trans R Soc Lond B Biol Sci. 2009;364: 2697–707. Available: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=2865094&tool=pmcentrez&rendertype=abstract

44. Herrero-Medrano JM, Megens H-J, Groenen MAM, Bosse M, Pérez-Enciso M, Crooijmans RPMA. Whole-genome sequence analysis reveals differences in population management and selection of European low-input pig breeds. BMC Genomics. 2014;15: 601. Available: http://www.biomedcentral.com/1471-2164/15/601

45. Fang M, Andersson L. Mitochondrial diversity in European and Chinese pigs is consistent with population expansions that occurred prior to domestication. Proc Biol Sci. 2006;273: 1803–10. Available: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=1634785&tool=pmcentrez&rendertype=abstract

46. Watanobe T, Ishiguro N, Nakano M. Phylogeography and Population Structure of the Japanese Wild Boar Sus scrofa leucomystax: Mitochondrial DNA Variation. Zoolog Sci. Zoological Society of Japan; 2003;20: 1477–1489. Available: http://dx.doi.org/10.2108/zsj.20.1477

47. Hasman H, Moodley A, Guardabassi L, Stegger M, Skov RL, Aarestrup FM. Spa type distribution in Staphylococcus aureus originating from pigs, cattle and poultry. Vet Microbiol. 2010;141: 326–31. Available: http://www.ncbi.nlm.nih.gov/pubmed/19833458

48. Asai T, Hiki M, Baba K, Usui M, Ishihara K, Tamura Y. Presence of Staphylococcus aureus ST398 and ST9 in Swine in Japan. Jpn J Infect Dis. 2012;65: 551–2. Available: http://www.ncbi.nlm.nih.gov/pubmed/23183212

49. Jamrozy DM, Fielder MD, Butaye P, Coldham NG. Comparative genotypic and phenotypic characterisation of methicillin-resistant Staphylococcus aureus ST398 isolated from animals and humans. Otto M, editor. PLoS One. Public Library of Science; 2012;7: e40458. Available: http://dx.plos.org/10.1371/journal.pone.0040458

50. Chen S, Gomes R, Costa V, Santos P, Charneca R, Zhang Y, et al. How immunogenetically different are domestic pigs from wild boars: a perspective from single-nucleotide polymorphisms of 19 immunity-related candidate genes. Immunogenetics. 2013;65: 737–48. Available: http://www.ncbi.nlm.nih.gov/pubmed/23846851

51. Fujita M, Into T, Yasuda M, Okusawa T, Hamahira S, Kuroki Y, et al. Involvement of leucine residues at positions 107, 112, and 115 in a leucine-rich repeat motif of human Toll-like receptor 2 in the recognition of diacylated lipoproteins and lipopeptides and Staphylococcus aureus peptidoglycans. J Immunol. 2003;171: 3675–83. Available: http://www.ncbi.nlm.nih.gov/pubmed/14500665

52. Meng G, Grabiec A, Vallon M, Ebe B, Hampel S, Bessler W, et al. 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. J Biol Chem. 2003;278: 39822–9. Available: http://www.jbc.org/content/278/41/39822

53. Mitsuzawa H, Wada I, Sano H, Iwaki D, Murakami S, Himi T, et al. Extracellular Toll-like receptor 2 region containing Ser40-Ile64 but not Cys30-Ser39 is critical for the recognition of Staphylococcus aureus peptidoglycan. J Biol Chem. 2001;276: 41350–6. Available: http://www.jbc.org/content/276/44/41350/F3.expansion

Chapter 4: Evolutionary patterns of Suidae genes involved in Toll-like receptor signaling pathway

Abstract

The Toll-like receptor (TLR) signaling pathway constitutes an essential component of the innate immune system. The pathway is characterized by highly conserved proteins, indicative of a critical role in host survival. Selective constraints on genes therefore might vary depending on the gene's position within the pathway. Thus, we investigated whether gene position might influence protein evolution in TLR signaling pathway across the genomes of the family Suidae. The members include the European *Sus scrofa*, Asian *Sus scrofa*, *Sus verrucosus*, *Sus celebensis*, *Sus scebifrons*, *Sus barbatus*, *Babyrousa babyrussa*, *Potamochoerus larvatus*, *Potamochoerus porcus and Phacochoerus africanus*. A total of 33 TLR signaling pathway genes were retrieved from resequencing data of these members of the family Suidae. The evolutionary parameter ω (dn/ds) had an overall mean of 0.1668 across genes, indicating high functional conservation within the signaling pathway. A significant relationship was inferred between the network parameters gene position, number of protein-protein interactions, protein length and the evolutionary parameter dn (nonsynonymous substitutions) such that downstream genes had lower nonsynonymous substitution rates, more interactors and shorter protein length than upstream genes. Gene position was also significantly correlated with the number of protein-protein interactions and protein length. Thus, the polarity in the selective constraint along the TLR signaling pathway was due to the number of molecules a protein interacted with and the protein's length. Results here indicate that the level of selective constraints on genes within the TLR signaling pathway of the family Suidae is dependent on the gene's position and other network parameters. In particular, downstream genes evolve more slowly as a result of being highly connected and having shorter protein lengths. Our findings highlight the critical role of gene network parameters in gene evolution.

Introduction

Proteins carry out their biological function by working within intricate networks of interacting molecules. High throughput techniques such as whole genome sequencing have led to accurate representation of gene networks [1]. Key insights regarding the influence of natural selection on genes can be obtained by taking into account the network topology [2].

The evolutionary rate of a protein (ω), represented by the ratio of the rate of its nonsynonymous substitutions (dn) to the rate of its synonymous substitutions (ds), is used as an indicator of the selective constraints acting on proteins. Studies have indicated that many factors affect the evolutionary rate of genes within gene pathways and networks. For example, gene position [3–5], protein length [6], the number of protein-protein interactions [7,8] and codon bias [9,10] have influence on gene evolution. Signaling pathways mediate the sensing and processing of both extracellular and intracellular stimuli. They rely on receptors that recognize a signaling molecule and trigger a series of events leading to the transmission of signals to the downstream region of the pathway. Within the *Drosophila* Toll and Imd signaling pathway, downstream genes were more conserved, indicating a relatively stronger evolutionary constraints than upstream genes [11]. This is consistent with trends reported for the yeast HOG-signaling pathway [12] and the *Caenorhabditis elegans* and *Drosophila* insulin/TOR pathway [13,14]. In contrast, in analysis of evolutionary patterns of the Insulin/FOXO signaling pathway across metazoan species genomes, components within the middle of the pathway were rather under stronger purifying selection [10]. Moreover, within the *Drosophila* Ras signaling pathway, upstream genes have lower evolutionary rate than upstream genes [15]. Obviously, different patterns of selective constraints exist for different signaling pathways. This is expected, as signaling inputs and network architecture vary among the pathways [16].

The TLR signaling pathway represents the best characterized component of the innate immune system [17]. TLR signaling involves the dimerization of receptors in response to pathogenic microbial products. This is followed by the recruitment of various adaptor molecules including *MyD88* that is used by all TLRs except *TLR3* and *TRIF* which is required by only *TLR3* and *TLR4*. The recruitment of these

adaptor molecules leads to the recruitment of downstream signaling molecules that activate transcription factors leading to the release of inflammatory cytokines and type I interferons. The TLR signaling pathway genes therefore demonstrate sequentiality. Thus, the effect of pathway parameters such as gene position on the evolutionary rate of genes can be investigated.

The TLR signaling pathway has been selectively constrained overtime, indicating their essential role for host survival [18]. In humans, population genetic studies have indicated that purifying selection is stronger within TLR adaptors relative to receptors [18]. A recent study on the TLR signaling pathway-related genes from eight vertebrate genomes showed that the selective constraints of genes was negatively correlated with gene position along the TLR signaling pathway [19]. Thus, different components of the TLR signaling pathway appear to differ in their evolutionary rates. Moreover, taxon specific differences among pathway components have been demonstrated for the TLR pathway [17]. New insights relating to the influence of network parameters on the evolutionary rate of genes within the TLR signaling pathway can be gained through studies on closely related species not yet investigated in detail.

Members of the family Suidae inhabit broad geographic locations including Eurasia, Mainland Southeast Asia, Island Southeast Asia [20] and the African continent [21]. Given the numerous pathogenic challenges across these diverse environments, the TLR signaling pathway may have been critical in the survival of Suidae members. The availability of a complete porcine genome and resequencing libraries of the family Suidae species provide an opportunity to investigate the evolution of the TLR signaling pathway genes within the context of gene position and other network parameters. Species of the Suidae family have also evolved over a relative short time span of 1-10 million years [22]. The low divergence among these species indicate that synonymous substitutions are not likely to be saturated, making the estimation and comparison of selective pressure (dn/ds) among genes less prone to bias. This low divergence among the species also ensures reliability in aligning TLR sequences for subsequent analysis.

In this study, our goal was to investigate the evolutionary constraints of genes within the TLR signaling pathway of the family Suidae members. We sought to investigate the extent to which patterns of

protein evolution observed in other organisms extend to members of the family Suidae. We hypothesized that contrasting patterns of selective constraints have acted among positions of the TLR signaling pathway within the family Suidae and tested our hypothesis by answering the following questions 1) is there a relationship between the strength of purifying selection and gene position in the TLR signaling pathway; and 2) are there any network parameters that contribute to the polarity in the strength of purifying selection?

Methods

DNA extraction and sequencing

DNA was extracted from whole blood by using the QIAamp DNA blood spin kit (Qiagen Sciences) and quantity and quality parameters were performed on the Qubit 2.0 fluorometer (Invitrogen) and run on a 1% agarose gel. Library construction and re-sequencing of individual members of the family Suidae were done with 1-3 ug of genomic DNA according to the Illumina library prepping protocols. The library insert size was 300-500 bp and sequencing was performed using a 100 paired-end sequencing kit. All DNA were sequenced to approximately 8x depth. Quality trimmed reads (phred quality>20, minimum length of pairs of reads=40bp) were aligned to the *Sus scrofa* reference genome build 10.2 using the unique alignment option of Mosaik Aligner (V.1.1.0017). The aligned reads from each of the animals together with the *Sus scrofa* reference genome (resequencing libraries) were stored as bam files for each individual animal.

Orthologs identification

Genes involved in the TLR signaling pathway (KEGG database pathway: ssc04620) were utilized in this study (Fig. 4.1). In addition *TLR10* which is not found in the KEGG database TLR signaling pathway was included as it cooperates with *TLR2* and senses lipopeptides [23] and activates the TLR signaling pathway through association with *MyD88* [24]. The genomic coordinates of *Sus scrofa* genes within the TLR signaling pathway (Table 4.1) were obtained from Ensemble database (http://www.ensemble.org). Based on these genomic coordinates, sequences of gene orthologs were then retrieved from aligned bam files (Illumina resequencing data for family Suidae species aligned against *Sus scrofa* genome assembly 10.2) of *Sus scrofa* (*Sus scrofa* Europe and *Sus scrofa* Asia), *Sus verrucosus*, *Sus celebensis*, *Sus scebifrons*, *Sus barbatus*, *Babyrousa babyrussa*, *Potamochoerus larvatus*, *Potamochoerus porcus* and *Phacochoerus africanus* to identify gene orthologs. The resulting sequences for each species were then blast screened against the *Sus scrofa* genome to ensure similarity with the *Sus scrofa* genes. To obtain the coding sequence of genes for each species, exonic regions were retrieved based on genomic coordinates of exons from the *Sus scrofa* gene transcripts (see Table 4.1 for accession numbers of mRNA sequences/gene transcripts) and concatenated. If a gene was found to have more than one transcript, the longest transcript was chosen for analysis. Coding sequences were aligned using ClustalW 1.81 [25].

Some of the coding sequences had lots of missing nucleotides. The following criteria were used to discard such sequences: (1) Sequences with more than 50% of missing amino acid residues (2) Aligned Sequences absent in any one of the species involved in this study. Consequently, the final data set was composed of 33 genes (Table 4.2).

Impact of natural selection

The impact of natural selection on genes within the TLR signaling pathway was determined by estimating nonsynonymous substitution rate (dn), synonymous substitution rate (ds) and their ratio (ω = dn/ ds) using the M0 model implemented in the program CODEML program from the PAML package version 4 [26,27]. ω values of <1, $= 1$ and > 1 are indication of purifying selection, neutral evolution and positive selection respectively targeting a gene. To determine whether some codon positions are under positive selection within gene orthologs, CODEML site model M1a, a nearly neutral evolution model where sites are assumed to be evolving under either purifying selection ($\omega < 1$) or neutral evolution ($\omega = 1$) was compared to model M2a that allows positive selection among sites. M7, which allows sites to evolve under either purifying selection or neutrally, was compared to model M8, which allows for positively selected sites. The F3x4 model of codon frequencies was used for the analyses. Models were run in duplicates with ω of 0.5 and 1.5 to increase the probability of convergence of model parameters. Multiple testing for positive selection on genes was corrected for by conducting a false discovery rate (FDR) test [28] at a q value of 0.05. The phylogenetic relationship among the Suidae species (Fig. 4.2) inferred from near complete genome sequences of each species [20] (L. Frantz, personal communication) was used for the analysis involving the CODEML program.

Network level analysis

To determine whether the evolution of molecules within each gene was affected by network structure, network parameters were computed for each gene and their correlation with each evolutionary parameter estimated by model M0 determined. Number of protein-protein interactions (connectivity) for proteins encoded by each gene was determined from the *Sus scrofa* interaction network in the STRING database (http://string.embl.de/). STRING is a database of known and predicted interactions which include direct (physical) and indirect (functional) associations from various sources including high throughput experiments and genomic contexts [29]. The magnitude for codon bias for individual genes

was measured by the Effective Number of Codons (ENC). The codon usage bias of each orthologous gene group was measured as the mean of ENC of each species gene. ENC values of each gene were obtained using DNASP software [30]. Bivariate correlation analysis between network parameters (pathway position within network, protein length, codon bias, connectivity (number of protein-protein interactions) and length of 3'UTR) and evolutionary parameters (ω, dn, ds) were conducted using Spearman's rank correlations. A false discovery rate (FDR) test [28] was performed to correct for multiple testing for correlations controlling for q value at 0.05. Bivariate correlations and corrections for multiple testing were performed using SAS software (SAS/STAT, version 9.1.3, SAS Institute Inc., Cary, NC). Figures illustrating bivariate correlations were obtained with the package R (http://www.rproject.org).

Multivariate analyses

The two multivariate analyses methods (partial correlation and path analysis) were used to determine whether the observed bivariate correlations were due to direct or indirect influences. Partial correlation analysis measures the strength of relationship between two variables, while holding one or more variables constant. Path analysis estimates direct and indirect relationships under a user defined causal model. For the path analysis, pathway position, number of protein-protein interactions, protein length, ENC and length of the 3UTR were considered as exogenous variables whereas ω and dn were considered as endogenous variables. Prior to performing the path analysis, data were log-transformed to improve normality. All statistics related to path analysis were calculated by the lavaan package in R.

Results

Orthologs of 33 *Sus scrofa* TLR signaling pathway genes were identified for 10 members of the family Suidae. The majority of genes within the TLR signal pathway as indicated in the KEGG database were within the genomes in each Suidae species involved in this study. A total of 330 sequences ranging

from 351 to 3153 nucleotides were used in analysis. The sequences from orthologs were aligned for a series of analysis including test for selection pressures acting on genes, bivariate correlations and multivariate analysis.

Analysis of protein sequence evolution

In order to estimate selective pressure acting on genes within the Suidae TLR signaling pathway, the M0 model, which provides a single estimate of ω across all codons and lineages, was utilized. The ω values ranged from 0.0001 (*MKK6, MEK1, MAPK1, MAPK9 and MAPK14*) to 1.0544 (*TLR10*) (Table 4.2). The mean ω was 0.1668. To test for positive selection, the models M1a (nearly neutral) vs M2a (positive selection); M7 vs M8 were used. The two tests were in agreement for all genes (Table 4.3) and demonstrated that a proportion of sites within *TLR1, TLR2, TLR6 and IRAK4* (Table 4.3) were under positive selection. However, *TLR1* was the only gene under positive selection after correcting for multiple testing at q = 0.05. Thus, results here indicated that genes within the TLR signaling pathway have evolved under strong functional constraint.

Relationship between evolutionary rates and pathway position

The relationship between the evolutionary parameter ω of genes and their position within the TLR signaling pathway was determined using a spearman's rank correlation test between the two variables. The ω values were negatively correlated with pathway position (Spearman's rank correlation coefficient ρ = -0.6250; P = 0.0005 after FDR correction; Table 4.4 and Fig. 4.3) indicating that downstream genes are under stronger purifying selection than upstream genes. ω is a ratio of dn and ds. To test for the evolutionary parameter accounting for the ω differences among genes, correlation between pathway position and dn, pathway position and ds was also tested. The dn was negatively correlated with the pathway position (Spearman's rank correlation coefficient ρ = -0.6110; P = 0.0007 after FDR correction; Table 4.4 and Fig. 4.3). There was no significant correlation between ds and pathway position (spearman's rank correlation coefficient ρ = -0.0990; P = 0.7117). These results indicated that the

decrease in ω (increase in the strength of purifying selection) from upstream to downstream genes is attributable to a decrease in dn (the rate of nonsynonymous substitution).

Relationship between evolutionary rates and other network parameters

Network parameters aside from pathway position can influence the evolutionary parameters (ω, dn, ds) and account for inferred polarity in ω and dn. In order to determine such relationships, Spearman's rank correlation test was performed between the number of protein-protein interactions (PPI) of proteins encoded by genes within the TLR signaling pathway, protein length, codon bias measured as effective number of codons (ENC), length of the 3 UTR region and the evolutionary parameters (Table 4.4 and Fig. 4.3). The ω values were negatively correlated with PPI (Spearman's rank correlation coefficient ρ = -0.6240; P = 0.0005 after FDR correction) and positively correlated with protein length (Spearman's rank correlation coefficient ρ = 0.5990; P = 0.0007 after FDR correction). Similarly, dn values were negatively correlated with PPI (Spearman's rank correlation coefficient ρ = -0.6740; P = 0.0005 after FDR correction) and positively correlated with protein length (Spearman's rank correlation coefficient ρ = 0.672; p = 0.0005 after FDR correction) indicating that apart from pathway position, the number of protein-protein interactions a given protein is involved in and its protein length influence it evolution. In addition, ds was also negatively correlated with ENC (Spearman's rank correlation coefficient ρ = -0.4650; p = 0.0163 after FDR correction) indicating stronger selection based on codon usage in genes with high codon bias than genes with low codon bias [31].

Since pathway position, PPI and protein length are intercorrelated and are each correlated with evolutionary parameters (ω and dn) (Table 4. 3), observed associations of these network parameters with the evolutionary parameters could be indirect (correlation between 2 parameters due to their both being correlated with a third parameter). Thus, to distinguish between direct and indirect effects, multivariate analysis (partial correlation and path analysis) were utilized. Partial correlation analysis revealed that when controlling for PPI, the correlation between ω and pathway position was still significant (ρ = -0.431, P = 0.014). When controlling for protein length, the correlation between ω and pathway position

remained significant (ρ = -0.386, P = 0.029). Thus, gene position within the TLR signaling pathway directly influenced the evolution of genes in the TLR signaling pathway. The correlation between ω and PPI while controlling for pathway position remained significant (ρ = -0.429, P = 0.014). It also remained significant (ρ = -0.452, P = 0.009) when controlling for protein length. This result indicated that PPI is an important factor affecting the evolution of genes within the TLR signaling pathway. The correlation between ω and protein length was not significant (ρ = 0.323, P = 0.072) when pathway position was held constant indicating that the correlation between ω and protein length was rather mediated by pathway position. Nonsynonymous (dn) substitution rates were not significantly (ρ = -0.306, P = 0.089) correlated with pathway position when controlling for protein length. In contrast, the correlation between dn and PPI remained significant (ρ = -0.506, P = 0.003) when controlling for protein length. If pathway position was held constant, the correlation between dn and PPI remained significant (ρ = -0.600, P = 0.003). Correlation between dn and protein length was still significant (ρ = 0.456, P = 0.009) after controlling for pathway position and remained significant (ρ = 0.504, P = 0.003) if controlling for PPI. Results for the relationship between dn values and network parameters indicated that the number of protein-protein interactions and protein length have direct effects on the rate of dn evolution.

The relationship between evolutionary parameters (ω and dn) and network parameters (pathway positon, PPI, protein length, ENC and L3UTR) were further analyzed using path analysis. As depicted in Fig. 4.4, path analysis indicated that dn values were affected by gene position within the TLR signaling pathway (standardized path coefficient, β = -0.366, P = 0.008). Pathway position had the largest direct effect on dn. The dn values were negatively associated with the number of protein-protein interactions (β = -0.314, P = 0.011) and positively associated with protein length (β = 0.281, P = 0.029). ω values were only associated with ENC (β = -0.230, P = 0.044). Thus pathway position, the number of protein interactions and protein length were factors that influenced Suidae TLR signaling pathway dn substitution rates.

Discussion

These results support our hypothesis that there is polarity of purifying selection along the TLR signaling pathway (from TLRs as upstream genes to genes downstream of TLRs) within the family Suidae with purifying selective pressure increasing along the pathway. Protein-protein interactions and protein length accounted for this polarity in the strength of purifying selection, where number of interactors of TLR signaling protein molecules increases and protein length decreasing along the TLR signaling pathway. The evolutionary parameter influenced by these network parameters was dn (nonsynonymous substitution rate), which is actually the metric of selective pressure [11]. Lesser amino acid substitutions were tolerated along the signaling pathway reflective of stronger purifying selection.

The polarity in the strength of purifying selection inferred in this study corroborated with other studies. Upstream genes revealed greater constrains than downstream genes in the carotenoid biosynthesis pathway [32,33] and in the dopamine catabolic pathway across mammals [34]. The relatively slow rate of evolution of upstream genes reflects these genes being required for a wider range of end products and therefore being more pleiotropic [35,36]. A negative correlation between strength of purifying selection and gene position within pathway, as inferred in this study, have also been reported for the Drosophila Toll and Imd signaling pathways [11] and TLR signaling across metazoan genomes [4]. This pattern of selective constraint might be attributed to purifying selection acting to maintain the function of downstream signal transduction elements and a concentration of adaptive changes in the upstream genes due to their interaction with external environment [13]. In this study, *TLR1* was the only gene inferred to be under positive selection after FDR correction. Therefore, adaptive changes were not prominent within upstream genes of the TLR signaling pathway resulting in the inferred pattern of selective constraints.

With the inferred significantly negative correlation between pathway position and the evolutionary parameters ω and dn, a negative correlation between pathway position and number of protein-protein interactions for a protein, as indicated in this study, imply a role for the number of protein-protein interactions in the polarity of selective constraints. This negative relationship is indicative of increasing selective constraints for proteins along the TLR signaling pathway, as a result of increase in

the number of interactors along the signaling pathway. Evidence for increasing connectivity of proteins along signaling pathway has also been demonstrated for the human signal transduction network [2]. Within metabolic pathways, connectivity plays a major role in constraining evolutionary rates with proteins interacting with more proteins being subject to stronger selective constraints [7,37,38]. The increasing number of interactors along the TLR signaling pathway indicates how essential the signaling process is and the involvement of protein molecules in other functions apart from TLR signaling. This places stronger selective constraint on protein molecules.

In this study, the finding of a significant negative correlation between protein length and pathway position had been detected in the insulin/TOR signaling of *Drosophila* [3] and indicated that protein length decreased and become more actively translated along the pathway [3]. Given that partial correlation and pathway analysis were both in agreement as to there been a positive relationship between protein length and dn, protein length appears to be a factor responsible for the polarity in dn along TLR signaling pathway. Thus, along the TLR signaling pathway, the decrease in protein length is associated with a decrease in the amino acid substitutions. A similar pattern has been reported for the woody perennial plant *Populus tremula*, where protein length is the main factor affecting selective constraints, with purifying selection weaker in genes with longer coding genes [39]. A possible explanation for the relationship between purifying selection and protein length is that selection at more than one site should cause an overall reduction in the effectiveness of selection (Hill-Robertson effect) [40,41]. In that case, for longer proteins, which may have many sites under selection simultaneously [39], there will be a reduced efficiency in natural selection.

Though path analysis indicated a negative association between ω and ENC, pathway position and ENC were not significantly correlated, therefore ENC cannot explain correlation between selective constraints and ENC. In this study, length of the 3UTR region of a gene did not have a relationship with any evolutionary parameter in disagreement with studies in the Toll/imd pathway in the Drosophila Toll and Imd signaling pathways [11] and human and mice miRNA target prediction data [42]. The relationship between the length of the 3UTR and evolutionary parameters is mediated by the number of

regulatory miRNAs, with genes with longer 3UTRs likely to be regulated by more miRNAs [11,42]. Genes regulated by more miRNAs are likely to be under stronger constraints [11,42]. Thus length of the 3UTR region has a rather indirect effect on protein evolution. It is therefore likely that in the case of the Suidae TLR signaling pathway genes, this indirect effect may be weak.

Genes within the signaling pathway had average ω value < 1, indicating that the TLR signaling pathway is selectively constrained across members of the Suidae. This is in keeping with the essential role of the pathway in innate immunity and host survival. Selectively constrained regions within the genome are likely to be functionally important [43]. The strongest selective constraints inferred for the MAP kinases (*MKK6, MEK1, MAPK1, MAPK9, MAPK14*) with dn substitution rates of zero for these genes suggest their involvement in critical roles during TLR signaling. The MAP kinases cascade components play important role in the production of proinflammatory mediators [44]. Recently, it has been demonstrated that MAP kinases play a role in agonist dependent regulation of cognate TLR mRNA levels [45]. MAP kinases are also required for the regulation of cellular development and differentiation processes [46,47]. The conservation of the MAP kinase genes in this study could therefore be attributed to their involvement in many processes. Results of increasing purifying along the TLR signaling pathway suggests that there is a greater need to protect the integrity of proteins as one moves down the signaling pathway [18]. Thus, genes downstream of the TLRs might be essential and nonredundant for survival of species involved in this study.

Conclusion

By analyzing the evolution of genes within the family Suidae TLR signaling pathway in the context of network structure, a polarity in the strength of selective constraints along the pathway was inferred. This polarity was such that nonsynonymous substitutions decreased along the TLR signaling pathway. The number of protein-protein interactions and protein length were significantly correlated with both gene position and rate of nonsynonymous substitution indicating a role of network parameters in the polarity of purifying selection along the family Suidae TLR signaling pathway.

Acknowledgements

We thank L. A. Rund of the department of animal sciences of the University of Illinois for reading the manuscript and making useful suggestions. We are also grateful to Laurent Frantz of the Animal Breeding and Genomics Centre, Wageningen University for providing the species tree of the Suidae used in this study. This work was funded by US Department of Agriculture (USDA) Agriculture Research Service (ARS) Grant 58-5438-2-307 to LBS.

Figures and Tables

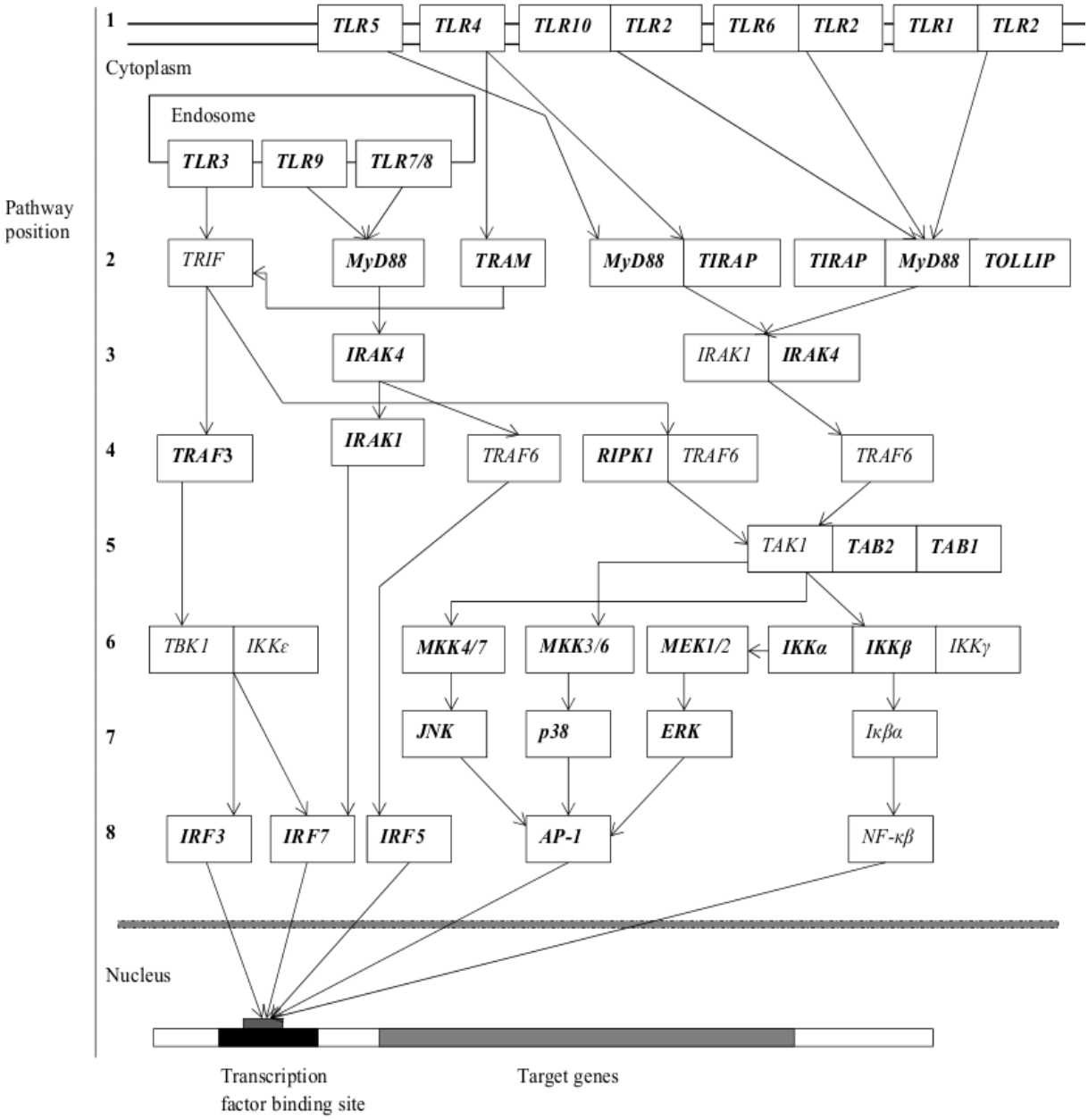

Fig. 4.1. The TLR signaling pathway genes. Redrawn following [4], with modification to include *TLR10*. The direction of signal transduction is indicated by the arrows. The numbers on the left side represent the position of the pathway genes. Genes used in this study are bolded.

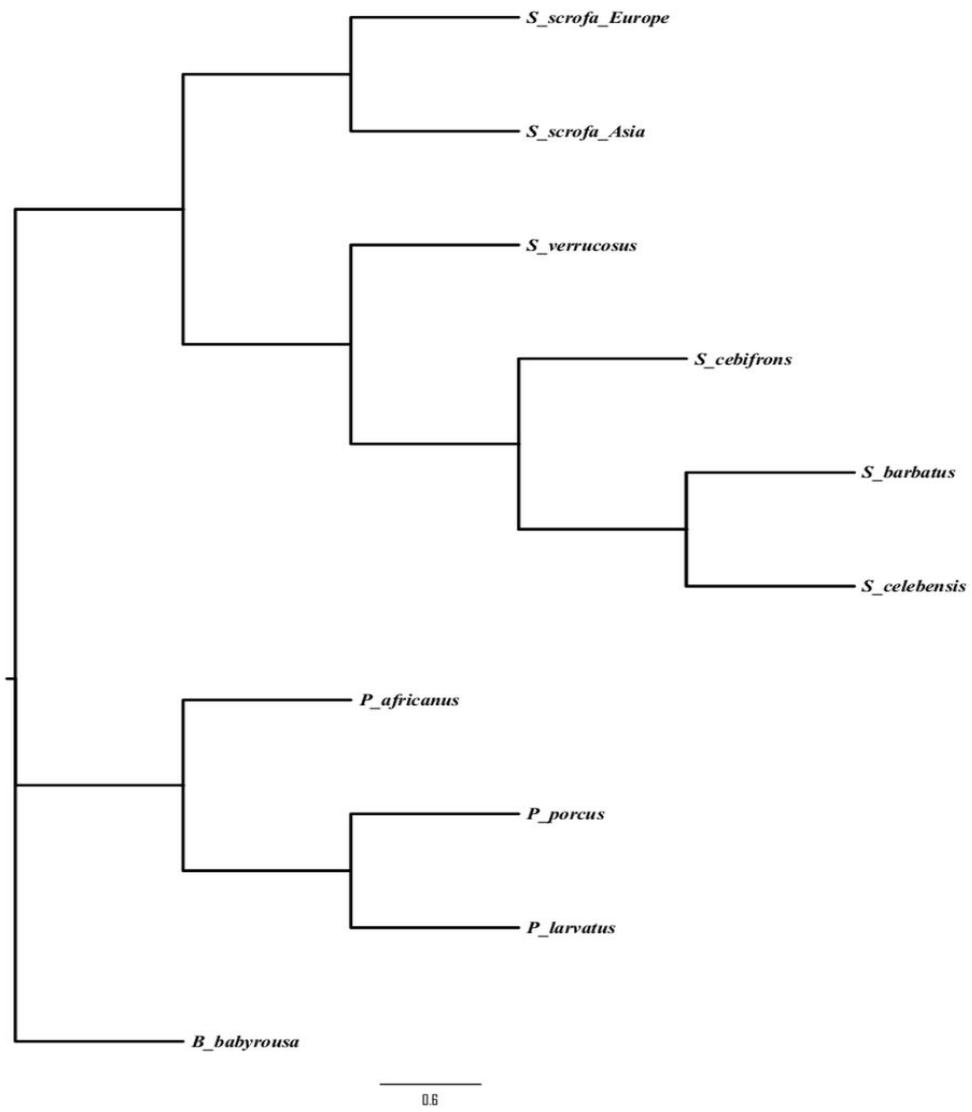

Fig. 4.2. Phylogenetic relationships among family Suidae species obtained from near complete genome data of each species. The posterior probability at each node is 1.

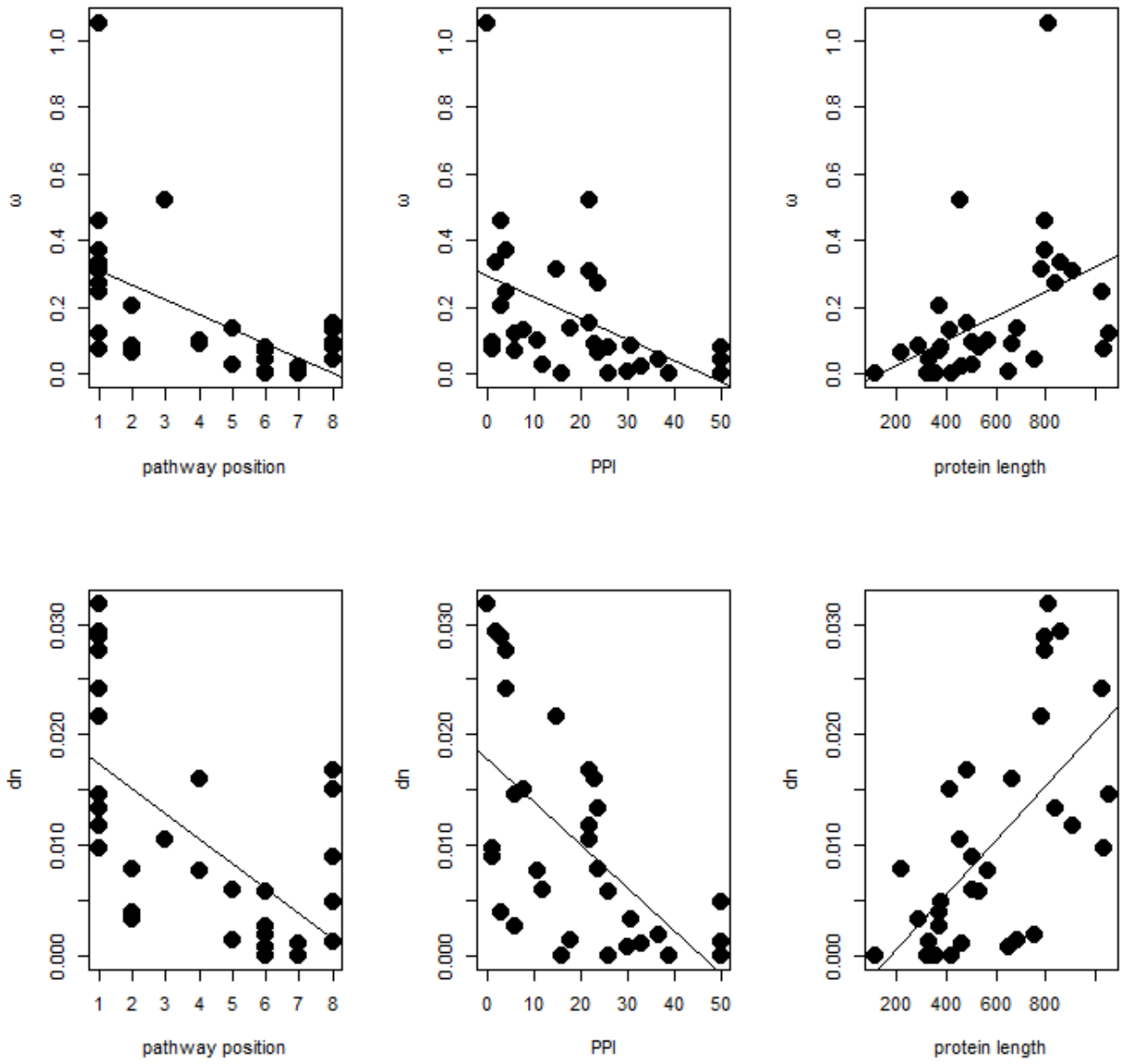

Fig. 4.3. ω and dn versus pathway position, number of protein-protein interactions (PP1) and protein length. All relationships are significant. Continuous lines represent regression lines.

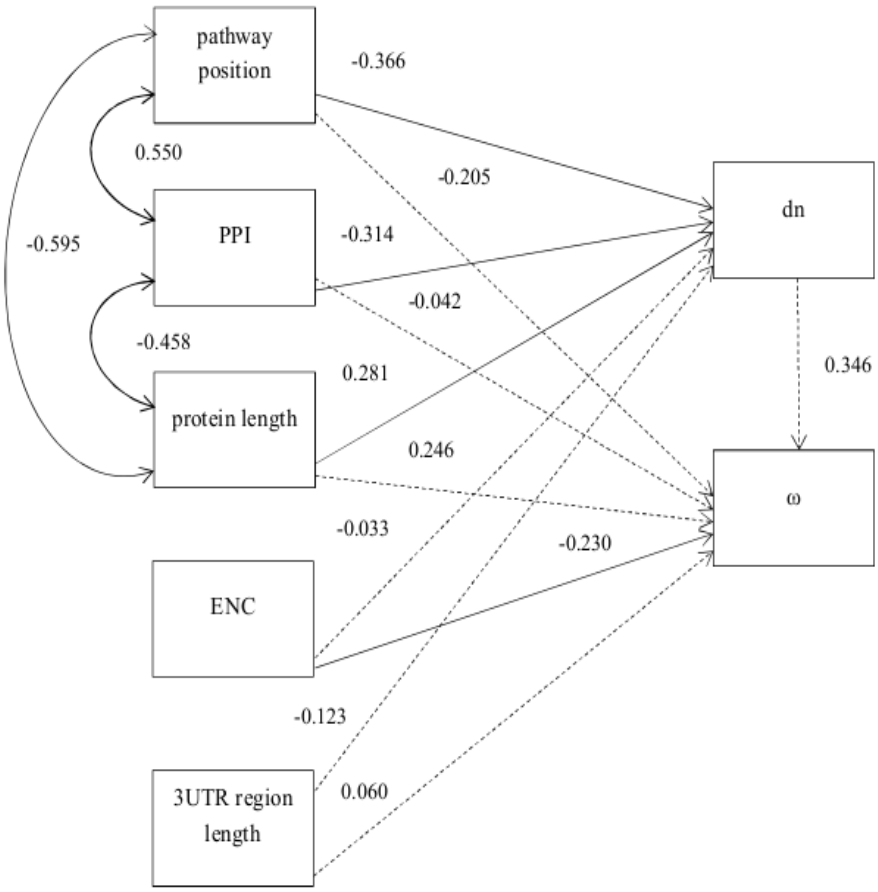

Fig.4.4. The causal model used to analyze the relationship among evolutionary and network parameters. The network parameter pathway position, number of protein-protein interactions (PPI), protein length, codon bias (measured as effective number of codons (ENC)) and length of the 3'UTR region are considered as exogenous variables. Continuous and dashed lines represent significant and nonsignificant relationships, respectively. Single-headed arrows indicate causal relationship between variables. Double-headed arrows indicate correlations between exogenous variables. Numbers on the arrows represent the standardized regression weights.

Table 4.1

Gene	Genomic coordinates	mRNA accession numbers/Transcript ID
TLR1	8: 31,627,788-31,635,628	NM_001031775
TLR2	8: 79,824,541-79,834,592	NM_213761
TLR3	15: 53,840,783-53,852,536	NM_001097444
TLR4	1: 289,775,822-289,785,847	NM_001113039
TLR5	10: 21,886,309-21,905,776	NM_001123202
TLR6	8: 31,641,826-31,660,823	NM_213760
TLR7	X: 10,450,659-10,476,655	NM_001097434
TLR8	X: 10,500,862-10,512,940	NM_214187
TLR9	13: 37,647,229-37,652,020	NM_213958
TLR10	8: 31,604,732-31,615,025	NM_001030534
MyD88	13: 25,181,051-25,185,351	NM_001099923
TIRAP	9: 59,051,622-59,056,203	ENSSSCT00000028018
TRAM	4: 70,399,155-70,429,614	ENSSSCT00000006787
IRAK4	5: 77,646,641-77,671,387	NM_001112693
TRAF3	7: 128,907,875-129,006,770	ENSSSCT00000002804
RIPK1	7: 1,921,742-1,945,193	ENSSSCT00000001101
TAB1	5: 6,122,718-6,146,864	NM_001244067
TAB2	1: 18,852,325-18,903,792	ENSSSCT00000004545
IKKα	14: 120,756,666-120,854,124	NM_001114279
IKKβ	17: 13,063,170-13,101,996	ENSSSCT00000007699
MKK6	12: 11,139,492-11,264,120	ENSSSCT00000018783
MKK4	2: 59,331,232-59,409,597	ENSSSCT00000019611
MKK7	2: 71,697,110-71,708,586	ENSSSCT00000014840
MEK1	1: 182,175,684-182,177,853	ENSSSCT00000023493
MAPK1	14: 53,590,167-53,614,842	ENSSSCT00000011042
MAPK9	2: 79,823,591-79,881,875	ENSSSCT00000035631
MAPK10	8: 141,997,575-142,137,195	ENSSSCT00000010107
MAPK14	7: 36,725,707-36,795,310	ENSSSCT00000001734
IRF3	6: 50,430,671-50,436,164	NM_213770
IRF5	18: 20,747,732-20,760,161	ENSSSCT00000018043
IRF7	2: 299,444-302,179	NM_001097428
FOS	7: 104,293,657-104,297,121	NM_001123113
JUN	6: 141,230,121-141,233,597	NM_213880

Table 4.1. Genes of the *Sus scrofa* TLR signaling pathway used in querying genomes of other Suidae species

Table 4.2

Gene	ω	*dn*	*ds*	Position	PPI	ENC	PLENGTH	L3UTR
TLR1	0.36932	0.0276	0.0747	1	4	53.28	796	181
TLR2	0.31229	0.0216	0.069	1	15	54.71	785	192
TLR3	0.30699	0.0117	0.038	1	22	54.28	905	218
TLR4	0.26995	0.0133	0.0493	1	24	53.54	841	806
TLR5	0.33407	0.0293	0.0878	1	2	52.27	856	1516
TLR6	0.46067	0.0288	0.0624	1	3	53.37	796	468
TLR7	0.12225	0.0145	0.1187	1	6	54.58	1050	655
TLR8	0.24534	0.0241	0.0984	1	4	56.33	1028	53
TLR9	0.0764	0.0097	0.1271	1	1	35.82	1030	167
TLR10	1.05445	0.0318	0.0301	1	0	52.49	811	272
MyD88	0.08294	0.0032	0.0384	2	31	45	293	1641
TIRAP	0.0655	0.0078	0.1197	2	24	41.24	221	0
TRAM	0.20405	0.0039	0.0192	2	3	55.13	374	111
IRAK4	0.52095	0.0104	0.0199	3	22	50.19	460	759
TRAF3	0.10269	0.0077	0.0747	4	11	43.56	568	4976
RIPK1	0.08849	0.016	0.181	4	23	39.77	664	1704
TAB1	0.02593	0.0059	0.2261	5	12	37.37	504	895
TAB2	0.1354	0.0013	0.0098	5	18	52.19	689	1894
IKKα	0.0417	0.0018	0.0422	6	37	52.77	755	548
IKKβ	0.00611	0.0007	0.1187	6	30	38.85	649	774
MKK6	0.0001	0	0.0386	6	16	53.84	334	277
MKK4	0.07146	0.0026	0.0369	6	6	52.7	376	2455
MKK7	0.07876	0.0058	0.074	6	26	43.74	536	0
MEK1	0.0001	0	0.0859	6	26	43	117	0
MAPK1	0.0001	0	0.0455	7	50	51.73	325	634
MAPK9	0.0001	0	0.0311	7	39	53.73	424	604
MAPK10	0.02238	0.001	0.0465	7	33	54.76	464	3555
MAPK14	0.0001	0	0.0299	7	50	53.59	360	2288
IRF3	0.13262	0.0151	0.114	8	8	40.63	419	68
IRF5	0.0975	0.0089	0.0908	8	1	39.74	503	1138
IRF7	0.15374	0.0167	0.1086	8	22	36.93	487	48
FOS	0.07942	0.0048	0.0608	8	50	43.3	380	830
JUN	0.04095	0.0012	0.0305	8	50	37.48	335	1290

Table 4.2. Summary statistics for genes

Table 4.3

Gene	LnM1a	LnM2a	LnM7	LnM8	2(LnM2a-LnM1a)	2(LnM8-LnM7)
TLR1	-3884.4	-3872.6	-3884.6	-3872.6	23.6*†	24*†
TLR2	-3758.0	-3753.6	-3758.2	-3753.6	8.8*	9.2*
TLR3	-4016.4	-4016.4	-4016.4	-4016.4	0	0
TLR4	-3832.2	-3832.0	-3832.3	-3832.0	0.4	0.6
TLR5	-4259.9	-4259.5	-4259.9	-4259.5	0.8	0.8
TLR6	-3856.1	-3852.0	-3856.4	-3852.0	8.2*	8.8*
TLR7	-5104.2	-5104.2	-5104.2	-5104.2	0	0
TLR8	-5085.7	-5084.2	-5085.8	-5084.2	3	3.2
TLR9	-4586.9	-4586.9	-4586.9	-4586.9	0	0
TLR10	-3828.6	-3828.2	-3828.6	-3828.2	0.8	0.8
MyD88	-1238.3	-1238.3	-1238.3	-1238.3	0	0
TIRAP	-1012.8	-1012.8	-1012.8	-1012.8	0	0
TRAM	-1574.8	-1574.8	-1574.8	-1574.8	0	0
IRAK4	-1971.3	-1966.4	-1971.3	-1966.4	9.8*	9.8*
TRAF3	-2492.8	-2490.4	-2493.1	-2490.4	4.8	5.4
RIPK1	-3067.3	-3067.3	-3067.2	-3067.2	0	0
TAB1	-2323.6	-2323.6	-2323.6	-2323.6	0	0
TAB2	-2818.0	-2818.0	-2818.0	-2818.0	0	0
IKKα	-3225.9	-3225.9	-3225.9	-3225.9	0	0
IKKβ	-2710.1	-2710.1	-2710.1	-2710.1	0	0
MKK6	-1427.5	-1427.5	-1427.5	-1427.5	0	0
MKK4	-1630.0	-1630.0	-1630.0	-1630.0	0	0
MKK7	-2321.4	-2320.2	-2321.7	-2320.2	2.4	3
MEK1	-490.6	-490.6	-490.6	-490.6	0	0
MAPK1	-1370.1	-1370.1	-1370.1	-1370.1	0	0
MAPK9	-1745.5	-1745.5	-1745.5	-1745.5	0	0
MAPK10	-2000.2	-1998.9	-2000.6	-1998.9	2.6	3.4
MAPK14	-1518.4	-1518.4	-1518.4	-1518.4	0	0
IRF3	-1915.3	-1913.8	-1915.4	-1913.8	3	3.2
IRF5	-2238.1	-2237.4	-2238.1	-2237.4	1.4	1.4
IRF7	-2205.9	-2205.9	-2205.9	-2205.9	0	0
FOS	-1649.3	-1649.3	-1649.3	-1649.3	0	0
JUN	-1305.2	-1305.2	-1305.2	-1305.2	0	0

Table 4.3. Results of codon based test for positive selection
*$p<0.05$, †statistical significance at FDR test of $q=0.05$

Table 4.4

		ω	***dn***	***ds***	**Position**	**PPI**	**ENC**	**PLENGTH**	**L3UTR**
ω	ρ		0.868	-0.054	-0.625	-0.624	0.219	0.599	-0.162
	*P*_raw		<.0001	0.7650	<.0001	0.0001	0.2216	0.0002	0.3685
	*P*_FDR		0.0005	0.8380	0.0005	0.0005	0.3266	0.0007	0.4690
dn	ρ			0.364	-0.611	-0.674	0.035	0.672	-0.259
	*P*_raw			0.0374	0.0002	<.0001	0.8446	<.0001	0.1452
	*P*_FDR			0.0806	0.0007	0.0005	0.8759	0.0005	0.2710
ds	ρ				-0.099	-0.245	-0.465	0.269	-0.250
	*P*_raw				0.5846	0.1692	0.0064	0.1303	0.1609
	*P*_FDR				0.7117	0.2787	0.0163	0.2606	0.2787
Position	ρ					0.552	-0.393	-0.654	0.193
	*P*_raw					0.0009	0.0237	<.0001	0.2812
	*P*_FDR					0.0028	0.0553	0.0005	0.3756
PPI	ρ						-0.083	-0.529	0.193
	*P*_raw						0.6460	0.0051	0.2817
	*P*_FDR						0.7537	0.0042	0.3756
ENC	ρ							0.226	-0.017
	*P*_raw							0.2064	0.9250
	*P*_FDR							0.3211	0.9250
PLENGTH	ρ								-0.051
	*p*_raw								0.7781
	*p*_FDR								0.8380

Table 4.4. Bivariate correlations among variables
ρ is correlation coefficient
*P*_raw is p value before FDR correction
*P*_FDR is p value after FDR correction

References

1. Ramsay H, Rieseberg LH, Ritland K. The correlation of evolutionary rate with pathway position in plant terpenoid biosynthesis. Mol Biol Evol. 2009;26: 1045–53. Available: http://www.ncbi.nlm.nih.gov/pubmed/19188263

2. Alvarez-Ponce D. The relationship between the hierarchical position of proteins in the human signal transduction network and their rate of evolution. BMC Evol Biol. 2012;12: 192. Available: http://www.biomedcentral.com/1471-2148/12/192

3. Alvarez-Ponce D, Aguadé M, Rozas J. Network-level molecular evolutionary analysis of the insulin/TOR signal transduction pathway across 12 Drosophila genomes. Genome Res. 2009;19: 234–42. Available: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=2652205&tool=pmcentrez&rendertype=abstract

4. Song X, Jin P, Qin S, Chen L, Ma F. The evolution and origin of animal Toll-like receptor signaling pathway revealed by network-level molecular evolutionary analyses. PLoS One. 2012;7: e51657. Available: http://journals.plos.org/plosone/article?id=10.1371/journal.pone.0051657

5. Gongora J, Cuddahee RE, Nascimento FF do, Palgrave CJ, Lowden S, Ho SYW, et al. Rethinking the evolution of extant sub-Saharan African suids (Suidae, Artiodactyla). Zool Scr. 2011;40: 327–335. Available: http://doi.wiley.com/10.1111/j.1463-6409.2011.00480.x

6. Lemos B, Bettencourt BR, Meiklejohn CD, Hartl DL. Evolution of proteins and gene expression levels are coupled in Drosophila and are independently associated with mRNA abundance, protein length, and number of protein-protein interactions. Mol Biol Evol. 2005;22: 1345–54. Available: http://www.ncbi.nlm.nih.gov/pubmed/15746013

7. Montanucci L, Laayouni H, Dall'Olio GM, Bertranpetit J. Molecular evolution and network-level analysis of the N-glycosylation metabolic pathway across primates. Mol Biol Evol. 2011;28: 813–23. Available: http://www.ncbi.nlm.nih.gov/pubmed/20924085

8. Wei W, Zhang T, Lin D, Yang Z-J, Guo F-B. Transcriptional abundance is not the single force driving the evolution of bacterial proteins. BMC Evol Biol. 2013;13: 162. Available: http://www.biomedcentral.com/1471-2148/13/162

9. Sharp PM. Determinants of DNA sequence divergence between Escherichia coli and Salmonella typhimurium: codon usage, map position, and concerted evolution. J Mol Evol. 1991;33: 23–33. Available: http://www.ncbi.nlm.nih.gov/pubmed/1909371

10. Wang M, Wang Q, Wang Z, Wang Q, Zhang X, Pan Y. The Molecular Evolutionary Patterns of the Insulin/FOXO Signaling Pathway. Evol Bioinform Online. 2013;9: 1–16. Available: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3547545&tool=pmcentrez&rendertype=abstract

11. Han M, Qin S, Song X, Li Y, Jin P, Chen L, et al. Evolutionary rate patterns of genes involved in the Drosophila Toll and Imd signaling pathway. BMC Evol Biol. 2013;13: 245. Available: http://www.biomedcentral.com/1471-2148/13/245

12. Wu X, Chi X, Wang P, Zheng D, Ding R, Li Y. The evolutionary rate variation among genes of HOG-signaling pathway in yeast genomes. Biol Direct. BioMed Central Ltd; 2010;5: 46. Available: http://www.biology-direct.com/content/5/1/46

13. Jovelin R, Phillips PC. Expression level drives the pattern of selective constraints along the insulin/Tor signal transduction pathway in Caenorhabditis. Genome Biol Evol. 2011;3: 715–22. Available: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3157841&tool=pmcentrez&rendertype=abstract

14. Alvarez-Ponce D, Guirao-Rico S, Orengo DJ, Segarra C, Rozas J, Aguadé M. Molecular population genetics of the insulin/TOR signal transduction pathway: a network-level analysis in Drosophila melanogaster. Mol Biol Evol. 2012;29: 123–32. Available: http://mbe.oxfordjournals.org/content/29/1/123

15. Riley RM, Jin W, Gibson G. Contrasting selection pressures on components of the Ras-mediated signal transduction pathway in Drosophila. Mol Ecol. 2003;12: 1315–23. Available: http://www.ncbi.nlm.nih.gov/pubmed/12694293

16. Fitzpatrick DA, O'Halloran DM. Investigating the Relationship between Topology and Evolution in a Dynamic Nematode Odor Genetic Network. Int J Evol Biol. 2012;2012: 548081. Available: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3465961&tool=pmcentrez&rendertype=abstract

17. Cormican P, Lloyd AT, Downing T, Connell SJ, Bradley D, O'Farrelly C. The avian Toll-Like receptor pathway--subtle differences amidst general conformity. Dev Comp Immunol. 2009;33: 967–73. Available: http://www.ncbi.nlm.nih.gov/pubmed/19539094

18. Fornarino S, Laval G, Barreiro LB, Manry J, Vasseur E, Quintana-Murci L. Evolution of the TIR domain-containing adaptors in humans: swinging between constraint and adaptation. Mol Biol Evol. 2011;28: 3087–97. Available: http://www.ncbi.nlm.nih.gov/pubmed/21659570

19. Song X, Jin P, Qin S, Chen L, Ma F. The evolution and origin of animal Toll-like receptor signaling pathway revealed by network-level molecular evolutionary analyses. Wicker-Thomas C, editor. PLoS One. Public Library of Science; 2012;7: e51657. Available: http://dx.plos.org/10.1371/journal.pone.0051657

20. Frantz LAF, Schraiber JG, Madsen O, Megens H-J, Bosse M, Paudel Y, et al. Genome sequencing reveals fine scale diversification and reticulation history during speciation in Sus. Genome Biol. 2013;14: R107. Available: http://genomebiology.com/2013/14/9/R107

21. Pickford M. Synopsis of the biochronology of African Neogene and Quaternary Suiformes. Trans R Soc South Africa. Taylor & Francis Group; 2006;61: 51–62. Available: http://www.tandfonline.com/doi/abs/10.1080/00359190609519953#.VN4m8vnF9zg

22. Groenen MAM, Megens H-J, Semiadi G, Schook LB. Insights from whole genome sequence alignment of the domestic (Sus scrofa) and Javan Warty (Sus verrucosus) pigs. Plant and Animal Genomes XVIII Conference. San Diego, CA, USA; 2010.

23. Guan Y, Ranoa DRE, Jiang S, Mutha SK, Li X, Baudry J, et al. Human TLRs 10 and 1 share common mechanisms of innate immune sensing but not signaling. J Immunol. 2010;184: 5094–103. Available: http://www.ncbi.nlm.nih.gov/pubmed/20348427

24. Hasan UA, Dollet S, Vlach J. Differential induction of gene promoter constructs by constitutively active human TLRs. Biochem Biophys Res Commun. 2004;321: 124–31. Available: http://www.ncbi.nlm.nih.gov/pubmed/15358224

25. Thompson JD, Gibson TJ, Plewniak F, Jeanmougin F, Higgins DG. The CLUSTAL_X windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. Nucleic Acids Res. 1997;25: 4876–82. Available: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=147148&tool=pmcentrez&rendertype=abstract

26. Yang Z. PAML: a program package for phylogenetic analysis by maximum likelihood. Comput Appl Biosci. 1997;13: 555–6. Available: http://www.ncbi.nlm.nih.gov/pubmed/9367129

27. Yang Z. PAML 4: phylogenetic analysis by maximum likelihood. Mol Biol Evol. 2007;24: 1586–91. Available: http://www.ncbi.nlm.nih.gov/pubmed/17483113

28. Benjamini Y. Controlling the false discovery rate : a practical and powerful approach to multiple testing. J R Stat Soc. 1995;57: 289–300. Available: http://ci.nii.ac.jp/naid/10021910059/en/

29. Szklarczyk D, Franceschini A, Kuhn M, Simonovic M, Roth A, Minguez P, et al. The STRING database in 2011: functional interaction networks of proteins, globally integrated and scored. Nucleic Acids Res. 2011;39: D561–8. Available: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3013807&tool=pmcentrez&rendertype=abstract

30. Librado P, Rozas J. DnaSP v5: a software for comprehensive analysis of DNA polymorphism data. Bioinformatics. 2009;25: 1451–2. Available: http://www.ncbi.nlm.nih.gov/pubmed/19346325

31. Comeron JM, Kreitman M. The Correlation Between Synonymous and Nonsynonymous Substitutions in Drosophila: Mutation, Selection or Relaxed Constraints? Genetics. 1998;150: 767–775. Available: http://www.genetics.org/content/150/2/767.short

32. Clotault J, Peltier D, Soufflet-Freslon V, Briard M, Geoffriau E. Differential selection on carotenoid biosynthesis genes as a function of gene position in the metabolic pathway: a study on the carrot and dicots. PLoS One. 2012;7: e38724. Available: http://journals.plos.org/plosone/article?id=10.1371/journal.pone.0038724

33. Livingstone K, Anderson S. Patterns of variation in the evolution of carotenoid biosynthetic pathway enzymes of higher plants. J Hered. 100: 754–61. Available: http://www.ncbi.nlm.nih.gov/pubmed/19520763

34. Ma X, Wang Z, Zhang X. Evolution of dopamine-related systems: biosynthesis, degradation and receptors. J Mol Evol. 2010;71: 374–84. Available: http://www.ncbi.nlm.nih.gov/pubmed/20890594

35. Eanes WF, Merritt TJS, Flowers JM, Kumagai S, Sezgin E, Zhu C-T. Flux control and excess capacity in the enzymes of glycolysis and their relationship to flight metabolism in Drosophila melanogaster. Proc Natl Acad Sci U S A. 2006;103: 19413–8. Available: http://www.pnas.org/content/103/51/19413.full

36. Flowers JM, Sezgin E, Kumagai S, Duvernell DD, Matzkin LM, Schmidt PS, et al. Adaptive evolution of metabolic pathways in Drosophila. Mol Biol Evol. 2007;24: 1347–54. Available: http://www.ncbi.nlm.nih.gov/pubmed/17379620

37. Vitkup D, Kharchenko P, Wagner A. Influence of metabolic network structure and function on enzyme evolution. Genome Biol. 2006;7: R39. Available: http://genomebiology.com/2006/7/5/R39

38. Fraser HB, Hirsh AE, Steinmetz LM, Scharfe C, Feldman MW. Evolutionary rate in the protein interaction network. Science. 2002;296: 750–2. Available: http://www.ncbi.nlm.nih.gov/pubmed/11976460

39. Ingvarsson PK. Gene expression and protein length influence codon usage and rates of sequence evolution in Populus tremula. Mol Biol Evol. 2007;24: 836–44. Available: http://www.ncbi.nlm.nih.gov/pubmed/17204548

40. Hill WG, Robertson A. The effect of linkage on limits to artificial selection. Genet Res. 1966;8: 269–94. Available: http://www.ncbi.nlm.nih.gov/pubmed/5980116

41. Birky CW, Walsh JB. Effects of linkage on rates of molecular evolution. Proc Natl Acad Sci. 1988;85: 6414–6418. Available: http://www.pnas.org/content/85/17/6414.abstract

42. Cheng C, Bhardwaj N, Gerstein M. The relationship between the evolution of microRNA targets and the length of their UTRs. BMC Genomics. 2009;10: 431. Available: http://www.biomedcentral.com/1471-2164/10/431

43. Sella G, Petrov DA, Przeworski M, Andolfatto P. Pervasive natural selection in the Drosophila genome? PLoS Genet. 2009;5: e1000495. Available: http://journals.plos.org/plosgenetics/article?id=10.1371/journal.pgen.1000495

44. Alessi DR, Cuenda A, Cohen P, Dudley DT, Saltiel AR. PD 098059 is a specific inhibitor of the activation of mitogen-activated protein kinase kinase in vitro and in vivo. J Biol Chem. 1995;270: 27489–94. Available: http://www.ncbi.nlm.nih.gov/pubmed/7499206

45. Peroval MY, Boyd AC, Young JR, Smith AL. A critical role for MAPK signalling pathways in the transcriptional regulation of toll like receptors. PLoS One. 2013;8: e51243. Available: http://journals.plos.org/plosone/article?id=10.1371/journal.pone.0051243

46. Schaeffer HJ, Weber MJ. Mitogen-activated protein kinases: specific messages from ubiquitous messengers. Mol Cell Biol. 1999;19: 2435–44. Available: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=84036&tool=pmcentrez&rendertype=abstract

47. Dickman MB, Yarden O. Serine/threonine protein kinases and phosphatases in filamentious fungi. Fungal Genet Biol. 1999;26: 99–117. Available: http://www.ncbi.nlm.nih.gov/pubmed/10328981