European Journal of Human Genetics (2014) 22, 551–557

www.nature.com/ejhg

ARTICLE

Haplotype structure and positive selection at *TLR1*

Christopher Heffelfinger*[,1,2], Andrew J Pakstis[1], William C Speed[1], Allison P Clark[1], Eva Haigh[1], Rixun Fang[3], Mahohar R Furtado[3,4], Kenneth K Kidd[1,6] and Michael P Snyder[5,6]

Toll-like receptor 1, when dimerized with Toll-like receptor 2, is a cell surface receptor that, upon recognition of bacterial lipoproteins, activates the innate immune system. Variants in *TLR1* associate with the risk of a variety of medical conditions and diseases, including sepsis, leprosy, tuberculosis, and others. The foremost of these is rs5743618 c.2079T>G(p.(Ile602Ser)), the derived allele of which is associated with reduced risk of sepsis, leprosy, and other diseases. Interestingly, 602Ser, which shows signatures of selection, inhibits TLR1 surface trafficking and subsequent activation of NFκB upon recognition of a ligand. This suggests that reduced TLR1 activity may be beneficial for human health. To better understand *TLR1* variation and its link to human health, we have typed all 7 high-frequency missense variants (>5% in at least one population) along with 17 other variants in and around *TLR1* in 2548 individuals from 56 populations from around the globe. We have also found additional signatures of selection on missense variants not associated with rs5743618, suggesting that there may be multiple functional alleles under positive selection in this gene.
European Journal of Human Genetics (2014) **22,** 551–557; doi:10.1038/ejhg.2013.194; published online 4 September 2013

Keywords: TLR1; haplotypes; SNPs; selection; evolution

INTRODUCTION

Toll-like receptors (TLRs) are a series of surface receptors that have a critical role in innate immunity.[1–3] Toll was originally identified in *Drosophila* as a developmental gene and later as being involved in immune response to fungal infection.[4] Later, TLRs, 10 of which are found in humans, were found to detect both Gram-negative and Gram-positive bacteria as well as viral nucleic acids, identifying them as critical component of the innate immune system.[5,6] Studies have found evidence of significant TLR family adaptation and selection in primates and other mammals,[7,8] as well as evidence of significant *TLR1* evolution along the vertebrate lineage.[9] Considerable population-stratified variation exists in human TLR genes, possibly contributing to disease response.[10–12]

TLR1, when in a heterodimer with TLR2, recognizes a variety of triacylated lipoproteins,[11] as well as other lipoproteins, and specific recognition targets include *M. leprae*, *M. tuberculosis*, *Borrelia* species, and a variety of fungal pathogens.[10,12–18] Activation of the TLR1–TLR2 heterodimer is not always beneficial to the organism, as multiple studies have found that derived alleles at some SNPs inhibit TLR1 activity and reduce the risk of sepsis, leprosy, prostate cancer, pelvic inflammatory disease, and tuberculosis.[19–27] The foremost of these is rs5743618, whose derived allele c.2079T>G (p.(Ile602Ser)) reduces surface trafficking of the TLR1–TLR2 heterodimer,[28] resulting in a reduced response to heterodimer antagonists and, in turn, reduced activation of the NFκB pathway.[29] This phenotype is especially evident in the homozygous derived state.[28]

There exists, however, considerable genetic diversity in *TLR1* beyond rs5743618, including multiple, high-frequency (>5% in at least one population) missense variants. Multiple SNPs, such as rs4833095, rs3923647, and rs5743613, have direct evidence suggesting an effect on TLR1 activity,[30] whereas others have been linked through association studies.[22,27,31,32] There is evidence that some of these alleles may be under selective pressure.[29]

We have typed all seven high-frequency (>5% in at least one population) missense variants in 2548 individuals from 56 populations from around the globe. These missense variants include rs5743618, which has not been typed before in a large panel of populations. In addition to typing these variants, we have examined the evolution of the core exonal haplotype and searched for signatures of selection via relative extended haplotype homozygosity (REHH),[33] integrated haplotype score (iHS),[34] and Wright's Fst.[35]

We suggest, based on our results regarding haplotype structure and selection in concert with the conclusions of various association studies, that there is an additional functional variation at the *TLR1* locus that has yet to be discovered or fully characterized.

METHODS

Genotyping

Twenty-four SNPs (Supplementary Table S1) in 2548 individuals from 56 populations were genotyped (Supplementary Table S2) using Applied Biosystems Taqman assays (Applied Biosystems, Foster City, CA, USA) using 50–100 ng of genomic DNA per well. Manufacturer's protocol was followed excepting that 3 μl rather than 5 μl of master mix was used. SNP typing results were analyzed via the ABI Prism Sequence Detection System (Applied Biosystems). Table 1 lists the missense SNPs with SIFT,[36] Grantham,[37] PhastCons,[38] and Fst[35] scores. Positions of each SNP are given as per RefSeq NM_003263.3.

Genomic DNA used in these assays was collected from lymphoblastoid cell lines established and/or grown in the laboratory of Kenneth and Judith Kidd.

[1]Department of Genetics, Yale University, New Haven, CT, USA; [2]Department of Molecular, Cellular, and Developmental Biology, Yale University, New Haven, CT, USA; [3]Life Technologies, Foster City, CA, USA; [4]President & Founder, Biology for Global Good, Sam Ramon, CA, USA; [5]Department of Genetics, Stanford University, Palo Alto, CA, USA

*Correspondence: Dr C Heffelfinger, Department of Genetics, School of Medicine, Department of Molecular, Cellular, and Developmental Biology, Yale University, SHM I-308, PO Box 208005, New Haven, CT 06520-8005, USA. Tel: +1 203 785 2654; Fax: +1 203 785 6568; E-mail: Christopher.Heffelfinger@yale.edu

[6]These authors contributed equally to this work.

Received 1 February 2013; revised 2 July 2013; accepted 24 July 2013; published online 4 September 2013

Table 1 Typed missense SNPs with SIFT, Grantham, phastCons, and Fst scores

dbSNP ID	Reference sequence (NM_003263.3)	Putative aa substitution	SIFT[a]	Grantham[b]	Conservation (phastCons)[c]	Fst (Wright) 56 pops	Known protein motif
rs5743621	2472C>T	Pro733Leu	0.00	98	0.993	0.032	Toll/interleukin-receptor homology domain
rs5743618[d]	2079T>G	Ile602Ser	1.00	142	0.429	0.473	Transmembrane helices
rs76796448	1328C>A	His352Asn	0.01	68	0.011	0.098	
rs3923647	1188A>T	His305Leu	0.00	99	0.879	0.067	
rs4833095	1017G>A	Ser248Asn	0.52	46	0.013	0.207	Low-complexity segment
rs5743612	626C>T	His118Tyr	0.15	83	0.915	0.168	Leucine-rich repeat
rs5743611	513G>C	Arg80Thr	0.06	71	0.98	0.15	Leucine-rich repeat

[a]A score <0.05 is considered to be a deleterious change.
[b]A score >40 is considered to be deleterious.
[c]Probability that nucleotide belongs to a conserved element.
[d]Present in a six amino acid sequence in the cytoplasmic region essential for surface trafficking of TLR1. Reduced surface trafficking when serine variant is present.

Cell lines were established from human samples collected from normal, apparently healthy adults, with informed consent under protocols approved by government and institutional human subject agencies of all countries involved.

Haplotype structure

Haplotypes were constructed using PHASE,[39] HAPLO,[40] and MaCH 1.06.[41] Data sets used for figures and REHH testing were produced using HAPLO. The extended data set used for iHS testing was produced using fastPHASE. MaCH and PHASE were also used to validate HAPLO results.

Linkage disequilibrium

Linkage disequilibrium was determined by calculating r^2 (Δ^2 in Devlin and Risch[42]; Supplementary Table S3) on all populations grouped by region (Supplementary Table S2). Regions of high LD for adjacent SNPs in individual populations are displayed by population via HAPLOT[43] (Supplementary Figure S1).

Selection testing

For the purposes of selection testing, we used REHH,[33] iHS,[34] and Wright's Fst.[35] REHH and iHS are metrics based on relative probability of identity by descent measured outward from alleles at a central SNP or haplotype. REHH is more sensitive to selection as it measures the maximum relative value, whereas iHS is more stringent. High Fst values may indicate unusual variance of allele frequencies for a given SNP across a set of populations.

For REHH, populations were grouped by region (Supplementary Table S2) so that sample size would be sufficient for selection testing. REHH was then calculated on all polymorphisms with >10% frequency in a group. Populations with haplotype evidence of admixture and regions with insufficient individuals were excluded from this analysis.

An extended data set of 196 SNPs (Supplementary Table S4) previously typed on 506 individuals from 15 populations[44] (Supplementary Table S5) was used to test rs5743618 and rs4833095 for selection in the Middle East, Europe, and East Asia. Insufficient individuals were present in the Pacific Islands group to test rs5743612 for selection.

Scores from REHH and iHS were tested against scores from a matched data set consisting of 1000 simulated populations created under the Wright–Fisher neutral model using Hudson's mksamples (ms).[45] Parameters were constant population size, 0.0001 mutation rate, and 1.12 cM/Mb recombination rate based on the Chr 4 average.[46] Actual REHH values were then plotted *versus* simulated REHH values. Actual REHH values exceeding the simulated 95% confidence interval were deemed potentially significant.

Fst was calculated for each missense variant using the Wright definition[35] and compared with values calculated from 369 neutral biallelic markers in a similar cohort.[47]

Evolution of the core exonal haplotype

For the core haplotype, we considered the seven high-frequency (>5% in at least one population) missense variants plus the silent variant rs5743614, c.1792G>A in the exon of *TLR1*. For rs5743614, even though the 'G' allele is listed as ancestral in dbSNP Build 138, the 'A' allele is supported as being ancestral by the majority of current primate genome assemblies.[48] Of the two haplotypes consisting of the ancestral alleles of the seven nonsynonymous SNPs (CTACAGCG and CTGCAGCG) surrounding rs5743614 (the third position), the G allele haplotype is found on only 125 chromosomes, whereas A allele haplotype was found on 2302 chromosomes. It remains possible that the rs5743614 is ancestrally variable. Haplotypes were plotted outward from the ancestral haplotype in a stepwise fashion.

RESULTS

Frequency of missense alleles

All missense alleles with a previously characterized frequency (>5%) as per 1000 Genomes (www.1000genomes.org), HapMap (www.hapmap.org), and the Human Genome Diversity Project (http://hagsc.org/hgdp/files.html) were analyzed in 2607 individuals (Supplementary Table S2) using ABI Taqman assays. Global frequencies are given in all 56 populations in Figure 1 and Supplementary Table S6. Haplotype structure is described in Figure 2. Allele frequencies and sample sizes for all *TLR1* gene region SNPs in the current study have been contributed to the ALlele FRequency Database (http://alfred.med.yale.edu) along with the haplotype frequencies underlying Figure 2.[44]

Evolution of the core *TLR1* haplotype

All core *TLR1* haplotypes present on >20 chromosomes globally were successfully plotted in a stepwise manner from the ancestral haplotype (Figure 3). Of interest for selection testing is the 'B' path per Figure 3, where haplotypes containing the derived alleles of rs4833095, rs5743618, and rs5743611 sequentially evolved from the ancestral. Of these, rs4833095 is the only SNP that can be tested for selection independently, as it was the first to evolve and a haplotype containing its derived allele appears in multiple regions without the derived alleles of the other two. Although the derived allele of rs5743618 is only found on haplotypes with rs4833095's derived allele, its high frequency does make it possible to look for an additive effect in selection testing. The derived allele of rs5743611, c.513G>C(p.(Arg80Thr)) only appears on haplotypes with the derived alleles of rs4833095 and rs5743618, and thus could produce no truly unique signature of selection. The derived allele of SNP 57435612, c.626C>T(p.His118Thr), appears only on the 'C' haplotype, where it is most commonly found in India and the Pacific Islands, and thus produces selection scores independent from other tested alleles.

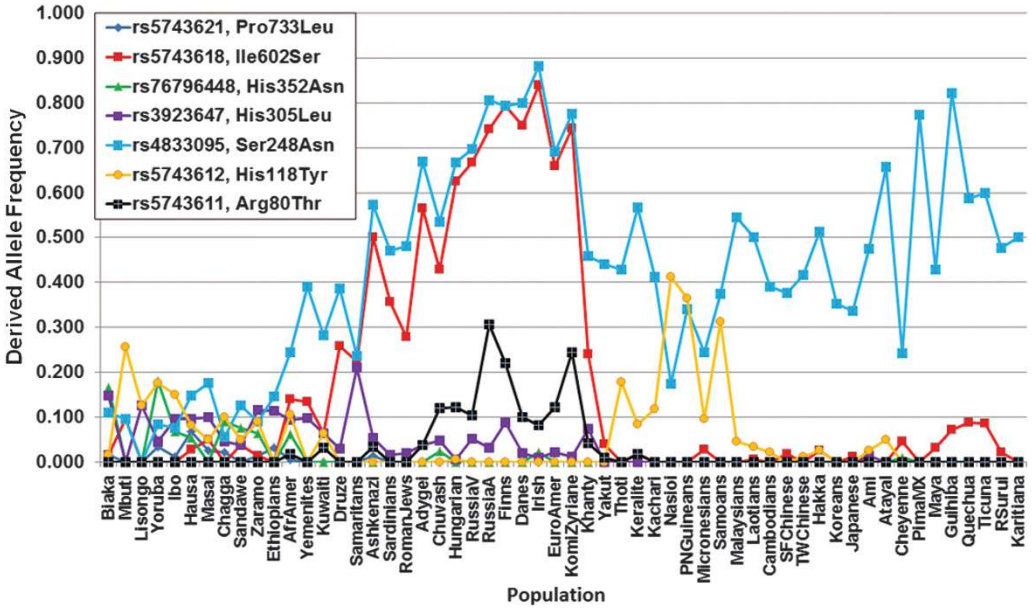

Figure 1 The derived allele frequencies for seven *TLR1* high-frequency missense SNPs in all 56 populations.

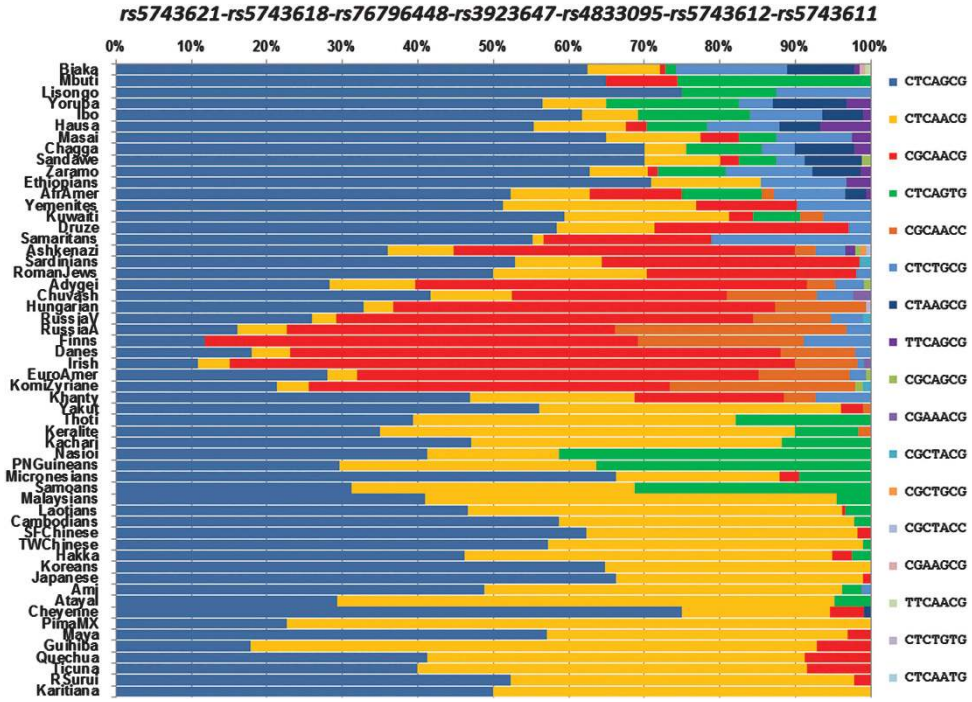

Figure 2 Haplotype structure for the seven common missense variants at *TLR1*. The most common global haplotype is CTCAGCG, which is also the ancestral haplotype. Haplotypes containing the derived allele of rs4833095, the SNP with the highest global heterozygosity, are indicated with warm colors. Haplotypes with the ancestral allele of rs4833095 are indicated using cool colors. Of interest to selection tests is the green haplotype, CTCAGTG. The green haplotype is the only haplotype that contains the derived allele of rs5743612, but it lacks the derived allele of rs4833095. The derived alleles of both rs5743612 and rs4833095 were found to be under selection in the Pacific Islanders, but their separate haplotypes suggest their high REHH values are two distinct signals.

Selection at rs5743618

The derived allele at the missense SNP rs5743618, c.2079T>G (p.(Ile602Ser)), occurs at high frequencies in Europe and low-to-moderate frequencies in some parts of the Middle East and is at low frequencies in the rest of the world. REHH indicates a weak signature of positive selection in both Europe (Supplementary Figure S2A) and the Middle East (Supplementary Figure S2B; 2.015 and 4.329, respectively). Both of these scores are above the 95% confidence interval, but still within the range of simulated neutral values. In addition, its Fst was 0.47, the highest of any missense SNP

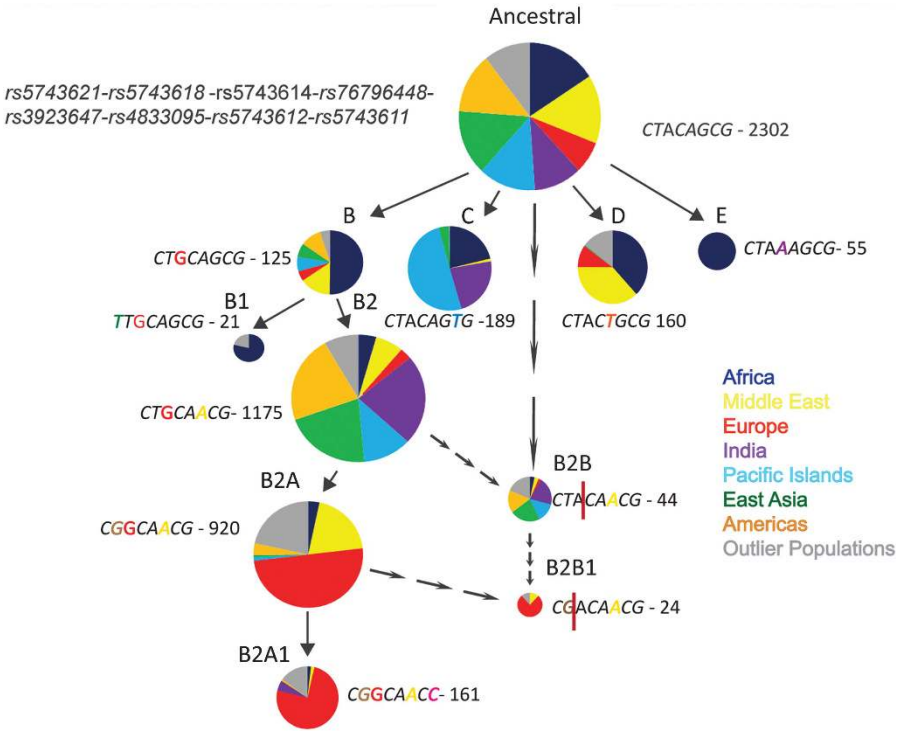

Figure 3 Evolution of the core *TLR1* haplotype. Solid lines indicate new haplotypes formed from mutation events, whereas spaced lines indicate recombination events. The most common haplotypes were used as the sources for recombination events, although there are other possible haplotypes of origin for them. Colored letters indicate derived alleles. Italics indicate missense polymorphisms and alleles. Letters above a haplotype indicate how they are referred to in text. Numbers for each haplotype indicate the number of chromosomes containing it in the data set. Outlier populations are populations whose haplotype structure differed from other populations in their region, or whose numbers were insufficient for independent study as a region. These include the Ashkenazi, Komi Zyrian, Khanty, and Yakut.

in this study and well above the mean neutral Fst of 0.136 (SD = 0.068). SNP rs5743618 was not significant by iHS score (Supplementary Figures S5A, B).

Selection at rs4833095

SNP rs4833095, c.1017G > A (p.(Ser248Asn)), has the highest global heterozygosity of any *TLR1* missense polymorphism. The derived allele is common in all regions of the world, with derived allele frequencies ranging from ~10% in most African populations to over 85% in some European populations. The possibility that this allele is under selection was tested using REHH. Two population groups, East Asia (Supplementary Figure S3A) and India (Supplementary Figure S3B), showed strong evidence of selection with REHH values (4.235 and 5.716, respectively) falling clearly outside the 95% confidence interval of the simulated data set. Two other populations groups, the Pacific Islanders (Supplementary Figure S3C) and the Middle Eastern (Supplementary Figure S3D) group, had lower scores (4.044 and 3.141, respectively) and derived allele frequencies that fell outside the 95% confidence interval but were within the range of noise produced by the simulated data set. In other population groups, the derived allele at rs4833095 either failed to show significant evidence of selection or the REHH values barely crossed the 95% confidence interval. In no population did rs4833095 produce a significant iHS score (Supplementary Figure S5A,B,C). Its overall Fst, 0.206, was not significant.

Selection at rs5743612

The derived allele of rs5743612, c.626C > T(p.(His118Tyr)), is rare or nonexistent in most global populations. The exception to this is some Pacific Islander populations in whom its frequency is ~30%. The derived allele at rs5743612 has a maximum REHH score of 5.999 in Pacific Islanders (Supplementary Figure S4A). The REHH score was also found to be climbing at the extremes of the tested SNP range, raising the possibility that the actual REHH for the derived allele of rs5743612 is even higher (Supplementary Figure S4B). The only other common missense allele (for rs4833095) had an REHH score of 4.044. The frequencies for the derived alleles of these two SNPs are similar as well (~25%); however, the LD between them is low (0.2). An analysis of haplotype structure reveals that the derived alleles of rs5743612 and rs4833095 appear on different haplotypes (Figure 2). The size of the Pacific Islander population in the extended data set was under 50 individuals, and thus it could not be effectively tested by iHS. Its global Fst of 0.168 was not significant, as it falls in under one standard deviation (0.068) of the neutral mean (0.136).

Previously detected signatures of selection

In addition to selection being detected on rs5743618,[29] c.2079T > G (p.(Ile602Ser)), positive selection has previously been detected at TLR1 on one intergenic and two intronic SNPs, rs5743595, rs5743565, and rs5743557, via a composite methodology.[49] To investigate if their signal of selection could be attributed to a *TLR1* missense allele we detected selection on, we determined the global haplotype structure of these six alleles (Supplementary Figure S6). In no population was a common haplotype (>5%) typed that contains both the three Grossman-derived alleles and any of the derived alleles we detected selection on.

DISCUSSION

Distribution and positive selection on missense variants at TLR1

The TLR1–TLR2 heterodimer recognizes a variety of triacylated lipoproteins and, when bound, activates the innate immune system; however, polymorphisms inhibiting its activity may actually be beneficial to the host for several reasons. There is evidence that leprosy hijacks the TLR1–TLR2 dimer to cause infection.[14,21,25,26] Reduced TLR1 activity has also been shown to result in a reduced risk of sepsis.[19,20] *TLR1*-derived alleles have also been linked to a variety of other diseases including placental malaria, IgA nephropathy, and invasive aspergillosis.[18,31,50]

Positive selection and distribution at rs5743618

This study represents the first global typing in a large panel of well-defined populations of rs5743618, c.2079T>G (p.(Ile602Ser)), a missense polymorphism in *TLR1* that inhibits the surface trafficking of the TLR1–TLR2 dimer.[28] It has previously been identified as being common among European populations as well as in one Turkish cohort, but rare or virtually absent elsewhere.[25,26,29] We have better characterized its distribution in European and Middle Eastern populations, as well as in Native American and African-American individuals where its presence is likely due to European admixture.

Of the SNPs tested for selection, only rs5743618 had been previously identified as a target of selection.[29] The biological evidence strongly supports the amino-acid change resulting from the derived variant being functional.[28] Reporter assays testing NFκB expression in HepG2 cells using constitutively expressed *TLR1* constructs with the ancestral and the derived states of c.2079T>G(p.(Ile602Ser)) have shown that 602Ser results in reduced NFκB activation.[29] Further studies using fluorescent constructs have shown this is likely a result of reduced surface trafficking.[28]

Positive selection and distribution of rs4833095

Of the *TLR1* amino-acid missense variants, the one with the highest global heterozygosity is rs4833095, c.1017G>A(p.(Ser428Asn)). In African populations, the derived allele has frequencies of <20% in most populations, but elsewhere in the world it tends to be around 50% in frequency, with frequencies over 80% in some European and Native American populations. It has the second highest SIFT score (0.53) and the lowest Grantham score[48] of any substitution in *TLR1*, although its Grantham score[48] is nonetheless above what is considered modest or negligible.[37] The presence of rs4833095 in a low-complexity domain of TLR1 may support it having a phenotypic effect, but the conservation score of the actual site is very low at 0.013. In addition, previous studies involving *TLR1* constructs with both states of the amino-acid substitution, serine and asparagine, have produced mixed results regarding functionality.[29,30]

We tested the derived allele at rs4833095 for positive selection using REHH in all global populations. In India and East Asia, and to a lesser degree in the Pacific Islands and the Middle East, it achieved significance by REHH but not with iHS. In the Middle East, evidence of selection can be dismissed as a result of it being in LD with rs5743618; however, in all other regions it is not in LD with any known missense variant.

The derived allele has mixed evidence indicating it may reduce TLR1 activity.[29,30] In addition, it has been linked to disease risk in populations where the derived allele of rs5743618 is >5% frequency, such as IgA nephropathy in Koreans[50] and placental malaria in Ghanians.[31] This, at the very least, suggests a phenotypic effect on a variant linked to rs4833095 independent of rs5746318, if not rs4833095 itself.

Positive selection and distribution of rs5743612

The derived allele of rs5743612, c.626C>T(p.(His118Tyr)), is present at frequencies above 40% in some Pacific Islander populations. Outside of the Pacific Islanders, it is rare with frequencies of 25.7% in the Mbuti and 17.5% in the Thoti, and lower frequencies in other African and Indian populations. It was absent in Europe and at low frequency (<5%) in East Asians. Its phastCons score (0.915) indicates it is in a conserved region of a leucine-rich repeat. The substitution is from a basic amino acid (histidine) to a neutral polar amino acid (tyrosine).

The only world region where the derived allele of rs5743612 was frequent enough to be tested for selection was the Pacific Islanders, where it was found to be significant with a REHH score of 5.999. This score was increasing at the fringe of our tested chromosomal region, indicating that the actual score may be even higher. Although the Pacific Islanders also showed significant REHH scores on rs4833095, a detailed analysis of haplotype structure reveals that the derived alleles of each appear on different haplotypes (Figure 2) and would not confer an effect on each other. Thus, it is possible that both rs4833095 and rs5743612 contribute toward a beneficial phenotype, or neither contributes toward a beneficial phenotype but both are associated with a variant that does via linkage disequilibrium.

Other high-frequency missense variants

Four other high-frequency missense variants have been typed in this study. SNPs rs5743621, c.2472C>T(p,(Pro733Leu)), and rs76796448, c.1328C>A(p.(His352Asn)), are variable in African populations but fixed in non-Africans. SNP rs3923647, c.1188A>T(p.(His305Leu)), is variable at low-to-moderate frequencies in Africa, the Middle East, Europe, India, and North Asia and is fixed elsewhere. Another SNP, rs5743611, c.513G>C(p.(Arg80Thr)), is variable only in Europe and the Middle East. Of these, all save rs76796448 are located in highly conserved regions (phastCons >0.8) and have SIFT and Grantham scores indicating potentially deleterious amino-acid substitutions (<0.05 and >50, respectively; Table 1). All, save rs76796448 are also located in known protein motifs. Most interestingly, the amino-acid substitution resulting from the derived allele of rs3923647 results in a modest loss of TLR1 activity as measured by luciferase assay.[30] Of these, only rs5743511 was at a high enough frequency to test for selection but could not produce a separate signal from rs5743618 or rs4833095.

Reconciliation with previous studies

The sites we found to be under selection do not agree with the results of Grossman *et al.*[49] Their composite method identified selection on the derived alleles of three SNPs, rs5743595, rs5743565, and rs5743557. Two of these SNPs are intronic and one is upstream of the TSS. We found that the haplotype formed by these three alleles was not associated with any of our proposed derived alleles under selection. SNP rs5743618 is not typed in the HapMap and thus was not tested for selection in the Grossman study. Although the HapMap data set[51] does contain rs5743612, it does not include any populations from the Pacific Islands, where we found signatures of selection. The derived allele of rs4833095 is present in many of their populations at reasonable frequencies and selection was not detected upon it. Our larger population sizes and more diverse data set may explain this.

One concern is how multiple haplotypes under selection may affect the results of the identity-by-descent selection tests (REHH and iHS).

Both rely on the assumption that measured sites will have a neutral allele. If multiple haplotypes are under selection at a locus, the absolute scores of both alleles will be increased. This reduces relative scores and testing sensitivity. This may, in part, explain why rs5743618 obtained only modest scores in spite of strong biological evidence for selection.

CONCLUSIONS

The main goal of this study was to add a global perspective to our knowledge of variation at *TLR1*, a gene previously associated with the risk of leprosy, sepsis, and other diseases.[20,21,50] Including many populations from the major geographical regions around the world has helped highlight the diversity of *TLR1* haplotypes that have evolved and the different frequency patterns they display across the globe. Possible future directions for study include a focused search for regulatory variation as well as functional study of rs5743612. Expanding the human populations genotyped will also assist in proper extrapolation of disease association studies at this locus.

The primary site of interest was rs5743618, a functional amino-acid change believed to directly affect sepsis risk.[20,28] This substitution has never before been studied in a major panel of diverse, defined populations, in part perhaps because typing is complicated by the region around it being a duplication of a region in *TLR6*.

Alongside rs5743618, rs4833095 and rs5743612 were also found to have signatures of selection. While neither rs4833095 nor rs5743612 can be definitively ruled as being a target of selection, they nonetheless indicate the presence of selection in populations outside of Europe and the Middle East—the only regions where rs5743618 is present at high frequencies. This, in combination with association studies at *TLR1* suggests that there is functional variation at *TLR1* beyond rs5743618.

That three distinct haplotypes would be found to be under selection in a single gene would normally give pause, as the traditional model of molecular evolution is that most nonsynonymous changes are neutral or deleterious rather than beneficial.[52] The simplifying explanation is that all of these polymorphisms are in strong LD with an uncharacterized regulatory variant. However, reduced TLR1 function decreases the risk of sepsis,[19,20] leprosy,[21,22,25,26,28] tuberculosis,[13,27] and other diseases.[18,31,50] Thus, variants that disrupt the function of TLR1 may be beneficial to the survival of the organism. Combined with the view that over time and place human populations have experienced rather different disease vector environments, which have interacted with the divergent genetic profiles present in those human populations, our data support the conclusion that multiple 'beneficial' amino-acid substitutions have arisen at *TLR1*.

CONFLICT OF INTEREST

The authors declare no conflict of interest.

ACKNOWLEDGEMENTS

This research was performed at Yale University School of Medicine, Department of Genetics, New Haven, CT. This work was supported in part by grants 2010-DN-BX-K225 to KKK awarded by the National Institute of Justice, Office of Justice Programs, US Department of Justice. Points of view in this document are those of the authors and do not necessarily represent the official position or policies of the US Department of Justice. Much of the support for the original collection of the population samples was from GM057672 awarded to KKK by the US National Institutes of Health. CH is supported by 5T32 GM007223. We would like to acknowledge all of our collaborators who helped collect the samples used in this research as well as the National Laboratory for the Genetics of Israeli Populations at Tel Aviv University and the Coriell Cell Repositories. Finally we would like to thank the thousands of individuals who donated samples without whom this research would not be possible.

1 Takeda K, Kaisho T, Akira S: Toll-like receptors. *Annu Rev Immunol* 2003; **21**: 335–376.
2 Medzhitov R: Toll-like receptors and innate immunity. *Nat Rev Immunol* 2001; **1**: 135–145.
3 Iwasaki A, Medzhitov R: Toll-like receptor control of the adaptive immune responses. *Nat Immunol* 2004; **5**: 987–995.
4 Lemaitre B, Nicolas E, Michaut L, Reichhart JM, Hoffmann JA: The dorsoventral regulatory gene cassette spatzle/Toll/cactus controls the potent antifungal response in *Drosophila* adults. *Cell* 1996; **86**: 973–983.
5 Medzhitov R, Preston-Hurlburt P, Janeway CA Jr: A human homologue of the *Drosophila* Toll protein signals activation of adaptive immunity. *Nature* 1997; **388**: 394–397.
6 Yoshimura A, Lien E, Ingalls RR, Tuomanen E, Dziarski R, Golenbock D: Cutting edge: recognition of Gram-positive bacterial cell wall components by the innate immune system occurs via Toll-like receptor 2. *J Immunol* 1999; **163**: 1–5.
7 Wlasiuk G, Nachman MW: Adaptation and constraint at Toll-like receptors in primates. *Mol Biol Evol* 2010; **27**: 2172–2186.
8 Enard D, Depaulis F, Roest Crollius H: Human and non-human primate genomes share hotspots of positive selection. *PLoS Genet* 2010; **6**: e1000840.
9 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.
10 Hawn TR, Misch EA, Dunstan SJ *et al*: A common human TLR1 polymorphism regulates the innate immune response to lipopeptides. *Eur J Immunol* 2007; **37**: 2280–2289.
11 Ranoa DR, Kelley SL, Tapping RI: Human lipopolysaccharide-binding protein (LBP) and CD14 independently deliver triacylated lipoproteins to Toll-like receptor 1 (TLR1) and TLR2 and enhance formation of the ternary signaling complex. *J Biol Chem* 2013; **288**: 9729–9741.
12 Takeuchi O, Sato S, Horiuchi T *et al*: Cutting edge: role of Toll-like receptor 1 in mediating immune response to microbial lipoproteins. *J Immunol* 2002; **169**: 10–14.
13 Uciechowski P, Imhoff H, Lange C *et al*: Susceptibility to tuberculosis is associated with TLR1 polymorphisms resulting in a lack of TLR1 cell surface expression. *J Leukoc Biol* 2011; **90**: 377–388.
14 Krutzik SR, Ochoa MT, Sieling PA *et al*: Activation and regulation of Toll-like receptors 2 and 1 in human leprosy. *Nat Med* 2003; **9**: 525–532.
15 Oosting M, Ter Hofstede H, Sturm P *et al*: TLR1/TLR2 heterodimers play an important role in the recognition of *Borrelia* spirochetes. *PLoS One* 2011; **6**: e25998.
16 Netea MG, Ferwerda G, van der Graaf CA, Van der Meer JW, Kullberg BJ: Recognition of fungal pathogens by toll-like receptors. *Curr Pharm Des* 2006; **12**: 4195–4201.
17 Bellocchio S, Montagnoli C, Bozza S *et al*: The contribution of the Toll-like/IL-1 receptor superfamily to innate and adaptive immunity to fungal pathogens in vivo. *J Immunol* 2004; **172**: 3059–3069.
18 Kesh S, Mensah NY, Peterlongo P *et al*: TLR1 and TLR6 polymorphisms are associated with susceptibility to invasive aspergillosis after allogeneic stem cell transplantation. *Ann N Y Acad Sci* 2005; **1062**: 95–103.
19 Wurfel MM, Gordon AC, Holden TD *et al*: Toll-like receptor 1 polymorphisms affect innate immune responses and outcomes in sepsis. *Am J Respir Crit Care Med* 2008; **178**: 710–720.
20 Pino-Yanes M, Corrales A, Casula M *et al*: Common variants of TLR1 associate with organ dysfunction and sustained pro-inflammatory responses during sepsis. *PLoS One* 2010; **5**: e13759.
21 Misch EA, Macdonald M, Ranjit C *et al*: Human TLR1 deficiency is associated with impaired mycobacterial signaling and protection from leprosy reversal reaction. *PLoS Negl Trop Dis* 2008; **2**: e231.
22 Schuring RP, Hamann L, Faber WR *et al*: Polymorphism N248S in the human Toll-like receptor 1 gene is related to leprosy and leprosy reactions. *J Infect Dis* 2009; **199**: 1816–1819.
23 Stevens VL, Hsing AW, Talbot JT *et al*: Genetic variation in the toll-like receptor gene cluster (TLR10-TLR1-TLR6) and prostate cancer risk. *Int J Cancer* 2008; **123**: 2644–2650.
24 Taylor BD, Darville T, Ferrell RE, Kammerer CM, Ness RB, Haggerty CL: Variants in toll-like receptor 1 and 4 genes are associated with *Chlamydia trachomatis* among women with pelvic inflammatory disease. *J Infect Dis* 2012; **205**: 603–609.
25 Wong SH, Gochhait S, Malhotra D *et al*: Leprosy and the adaptation of human toll-like receptor 1. *PLoS Pathog* 2010; **6**: e1000979.
26 Hart BE, Tapping RI: Genetic diversity of Toll-like receptors and immunity to *M. leprae* infection. *J Trop Med* 2012; **2012**: 415057.
27 Ma X, Liu Y, Gowen BB, Graviss EA, Clark AG, Musser JM: Full-exon resequencing reveals toll-like receptor variants contribute to human susceptibility to tuberculosis disease. *PLoS One* 2007; **2**: e1318.
28 Johnson CM, Lyle EA, Omueti KO *et al*: Cutting edge: a common polymorphism impairs cell surface trafficking and functional responses of TLR1 but protects against leprosy. *J Immunol* 2007; **178**: 7520–7524.
29 Barreiro LB, Ben-Ali M, Quach H *et al*: Evolutionary dynamics of human Toll-like receptors and their different contributions to host defense. *PLoS Genet* 2009; **5**: e1000562.

30 Omueti KO, Mazur DJ, Thompson KS, Lyle EA, Tapping RI: The polymorphism P315L of human toll-like receptor 1 impairs innate immune sensing of microbial cell wall components. *J Immunol* 2007; **178**: 6387–6394.
31 Hamann L, Bedu-Addo G, Eggelte TA, Schumann RR, Mockenhaupt FP: The toll-like receptor 1 variant S248N influences placental malaria. *Infect Genet Evol* 2010; **10**: 785–789.
32 Mikacenic C, Reiner AP, Holden TD, Nickerson DA, Wurfel MM: Variation in the TLR10/TLR1/TLR6 locus is the major genetic determinant of interindividual difference in TLR1/2-mediated responses. *Genes Immun* 2013; **14**: 52–57.
33 Sabeti PC, Reich DE, Higgins JM *et al*: Detecting recent positive selection in the human genome from haplotype structure. *Nature* 2002; **419**: 832–837.
34 Voight BF, Kudaravalli S, Wen X, Pritchard JK: A map of recent positive selection in the human genome. *PLoS Biol* 2006; **4**: e72.
35 Wright S: Genetical structure of populations. *Nature* 1950; **166**: 247–249.
36 Ng PC, Henikoff S: SIFT: predicting amino acid changes that affect protein function. *Nucleic Acids Res* 2003; **31**: 3812–3814.
37 Grantham R: Amino acid difference formula to help explain protein evolution. *Science* 1974; **185**: 862–864.
38 Siepel A, Bejerano G, Pedersen JS *et al*: Evolutionarily conserved elements in vertebrate, insect, worm, and yeast genomes. *Genome Res* 2005; **15**: 1034–1050.
39 Stephens M, Smith NJ, Donnelly P: A new statistical method for haplotype reconstruction from population data. *Am J Hum Genet* 2001; **68**: 978–989.
40 Hawley ME, Kidd KK: HAPLO: a program using the EM algorithm to estimate the frequencies of multi-site haplotypes. *J Hered* 1995; **86**: 409–411.
41 Li Y, Willer CJ, Ding J, Scheet P, Abecasis GR: MaCH: using sequence and genotype data to estimate haplotypes and unobserved genotypes. *Genet Epidemiol* 2010; **34**: 816–834.
42 Devlin B, Risch N: A comparison of linkage disequilibrium measures for fine-scale mapping. *Genomics* 1995; **29**: 311–322.
43 Gu S, Pakstis AJ, Kidd KK: HAPLOT: a graphical comparison of haplotype blocks, tagSNP sets and SNP variation for multiple populations. *Bioinformatics* 2005; **21**: 3938–3939.
44 Cheung KH, Osier MV, Kidd JR, Pakstis AJ, Miller PL, Kidd KK: ALFRED: an allele frequency database for diverse populations and DNA polymorphisms. *Nucleic Acids Res* 2000; 28: 361–363.
45 Hudson RR: Generating samples under a Wright-Fisher neutral model of genetic variation. *Bioinformatics* 2002; **18**: 337–338.
46 Jensen-Seaman MI, Furey TS, Payseur BA *et al*: Comparative recombination rates in the rat, mouse, and human genomes. *Genome Res* 2004; **14**: 528–538.
47 Kidd KK, Pakstis AJ, Speed WC, Kidd JR: Understanding human DNA sequence variation. *J Hered* 2004; **95**: 406–420.
48 Kent WJ, Sugnet CW, Furey TS *et al*: The human genome browser at UCSC. *Genome Res* 2002; **12**: 996–1006.
49 Grossman SR, Shlyakhter I, Karlsson EK *et al*: A composite of multiple signals distinguishes causal variants in regions of positive selection. *Science* 2010; **327**: 883–886.
50 Lee JS, Park HK, Suh JS *et al*: Toll-like receptor 1 gene polymorphisms in childhood IgA nephropathy: a case-control study in the Korean population. *Int J Immunogenet* 2011; **38**: 133–138.
51 The International HapMap Project Consortiuum. The International HapMap Project. *Nature* 2003; **426**: 789–796.
52 Kimura M: Preponderance of synonymous changes as evidence for the neutral theory of molecular evolution. *Nature* 1977; **267**: 275–276.

Supplementary Information accompanies this paper on European Journal of Human Genetics website (http://www.nature.com/ejhg)