The selective footprints of viral pressures at the human RIG-I-like receptor family

Estelle Vasseur1,2,3, Etienne Patin1,2, Guillaume Laval1,2, Sandra Pajon1,2, Simona Fornarino1,2, Brigitte Crouau-Roy3 and Lluis Quintana-Murci1,2,*

1Unit of Human Evolutionary Genetics and 2Centre National de la Recherche Scientifique, URA3012, Institut Pasteur, 25–28 Rue du Dr Roux, F-75015 Paris, France and 3CNRS, Université de Toulouse, EDB (Laboratoire Evolution et Diversité Biologique), UMR 5174, 118 route de Narbonne, F-31062 Toulouse, France

Received June 28, 2011; Revised and Accepted August 21, 2011

The RIG-I-like receptors (RLRs)—RIG-I, IFIH1 (or MDA5) and LGP2—are thought to be key actors in the innate immune system, as they play a major role in sensing RNA viruses in the cytosol of host cells. Despite the increasingly recognized importance of the RLR family in antiviral immunity, no population genetic studies have yet attempted to compare the evolutionary history of its different members in humans. Here, we characterized the levels of naturally occurring genetic variation in the RLRs in a panel of individuals of different ethnic origins, to assess to what extent natural selection has acted on this family of microbial sensors. Our results show that amino acid-altering variation at RIG-I, particularly in the helicase domain, has been under stronger evolutionary constraint than that at IFIH1 and LGP2, reflecting an important role for RIG-I in sensing numerous RNA viruses and/or functional constraints related to the binding of viral substrates. Such evolutionary constraints have been much more relaxed at IFIH1 and LGP2, which appear to have evolved adaptively in specific human populations. Notably, we identified several mutations showing signatures of positive selection, including two non-synonymous polymorphisms in IFIH1 (R460H and R843H) and one in LGP2 (Q425R), suggesting a selective advantage related to the sensing of RNA viruses by IFIH and to the regulatory functions of LGP2. In light of the fact that some of these mutations have been associated with altered risks of developing autoimmune disorders, our study provides an additional example of the evolutionary conflict between infection and autoimmunity.

INTRODUCTION

Viral recognition by host cells and the subsequent induction of interferons (IFNs) are mediated by pattern-recognition receptors, which sense pathogen-associated molecular patterns (PAMPs) from diverse groups of microbes (1–3). Two systems of RNA virus detection and IFN induction have been characterized so far; the Toll-like receptor (TLR) (4–6) and the RIG-I-like receptor (RLR) systems (7,8). Although the endosomally located TLRs, specialized in the sensing of viral nucleic acids, have been studied from different angles—immunological, genetic and evolutionary (9,10), much less attention has been paid to the RLRs. Identified in 2004, RLRs are RNA helicases that function as cytosolic sensors of viral RNA infection to initiate and modulate antiviral immunity (7,8,11–13). To date, three RLR members have been identified: RIG-I (retinoic acid-inducible gene I, or DDX58), IFIH1 (IFN-induced helicase C domain-containing protein 1, or MDA5) and LGP2 (laboratory of genetics and physiology 2, or DHX58) (11). They share a number of structural similarities including their organization into different protein domains: (i) a C-terminal domain (CTD) mostly involved in RNA recognition, with a particular conformation in the repressor domain (RD) for RIG-I and LGP2; (ii) a central DExD/H box RNA helicase domain with the capacity to hydrolyze ATP and to bind and possibly unwind RNA; and (iii) two caspase recruitment domains (CARDs), mostly associated with cell death and inflammatory signalling pathways for RIG-I and IFIH1 (12,14–16). RLRs and TLRs use distinct and specific adaptors to initiate their activity.
respective signalling cascades, which ultimately converge to the expression of type-I IFNs and proinflammatory cytokines (8,17–19).

The three RLRs, although activating the same signalling cascade, differ from each other in both their specific functions and the ligands which they sense. RIG-I and IFIH1 are both involved in the direct recognition of viral PAMPs, but they differ in their viral substrates and mechanisms of recognition [see (11) for an extensive review]. RIG-I detects a large range of RNA viruses including Newcastle disease virus, vesicular stomatitis virus, Sendai Virus, hepatitis C virus, influenza A virus, rabies virus, measles virus and respiratory syncytial virus (20–25). In turn, the viral substrates of IFIH1 are essentially restricted to picornaviruses, such as the encephalomyocarditis virus, rhinovirus and enterovirus (in particular Coxsackie B virus) (26–28). Some flaviviruses, such as Dengue type-2, or reoviruses have been shown to be sensed by both RIG-I and IFIH1, illustrating the occasionally overlapping functions of these two receptors (22). The different viral substrates of RIG-I and IFIH1 are mainly accounted for by the distinct structures of viral RNA the two sensors recognize; RIG-I is specialized in the sensing of ssRNAs bearing a 5′-triphosphate moiety as well as short dsRNAs (∼1 kb) (29–31), whereas IFIH1 binds long dsRNA (1–5 kb) (11,30). In both cases, these specificities allow them to distinguish viral ligands from self RNA (32). In contrast to RIG-I and IFIH1, LGP2 is not able to initiate antiviral signalling owing to its complete lack of a CARD domain, and has been shown instead to negatively regulate RIG-I signalling and positively regulate IFIH1 signalling (33).

In the past few years, genetic association studies and functional analyses have allowed a better understanding of the role of RLRs in human disease, suggesting a critical relationship between RLR polymorphisms, viral infections and autoimmunity (11). Genetic screens have led to the identification of a number of functional polymorphisms in both RIG-I and IFIH1, which are in some cases associated with altered cell responses to viral infection (34–36). The direct involvement of RLR polymorphisms in susceptibility to disease is well supported in the case of autoimmune disorders. Specifically, IFIH1 has been replicably identified as a strong susceptibility gene for type-I diabetes (37–41). The non-synonymous mutation identified as the most strongly associated (rs1990760, A946T) has also been associated with other autoimmune diseases, including Graves’ disease (42) systemic lupus erythematosus (43) and psoriasis (44). The A946T variant, which confers susceptibility to these diseases, also increases susceptibility to immunoglobulin A deficiency, the most common heritable immunodeficiency (45), which is diagnosed more often than expected in patients with autoimmune disorders. Altogether, these findings point to an important role of the RLR system in human autoimmunity. Interestingly, the observation that type-I diabetes susceptibility alleles in IFIH1 are found at high population frequencies (39) may attest to an advantage conferred by these variants against past infections (46,47). Indeed, the maintenance of risk alleles for autoimmune or inflammatory disorders, such as Crohn’s disease, celiac disease or multiple sclerosis, as the by-product of past adaptation to pathogen exposure is supported by a growing body of data (reviewed in 46,48). Perhaps the best documented example is that of celiac disease, where evolutionary and functional analyses have shown that several celiac disease-risk alleles have been positively selected and that individuals carrying these alleles benefit from increased protection against certain infectious agents (46,49,50).

Despite the increasingly recognized importance of the RLR family in host defence, no population genetic study has so far attempted to compare the evolutionary history of its different members. In this context, the evolutionary genetics approach has been shown to provide an indispensable complement to clinical and epidemiological genetics studies in helping to delineate the essential and redundant functions of host defence genes in the natural setting (10,46,51). Here, we characterized the levels of naturally occurring sequence-based variation of the three members of the RLR family in a panel of healthy individuals, representative of various populations worldwide. We used these data to search for selection signatures, by means of allele frequency spectrum- and linkage disequilibrium (LD)-based tests, together with population differentiation indexes. Our analyses reveal that natural selection has targeted to different extents the members of the RLR family and identify various genetic variants as putative targets of positive selection, providing a number of good candidates that may affect the susceptibility to infectious and autoimmune diseases.

RESULTS

To gain insight into the levels of naturally occurring genetic variation at the three members of the RLR family, we resequenced them in a panel of 186 healthy individuals from sub-Saharan Africa, Europe and East-Asia (Supplementary Material, Table S1). A total of 21.3 kb was sequenced per individual (7.5, 8.0 and 5.8 kb for RIG-I, IFIH1 and LGP2, respectively), 37% of which corresponded to exonic regions, the rest accounting for intronic and putative cis-regulatory regions (Supplementary Material, Table S2 and Fig. S1). Overall, we identified 206 single nucleotide substitutions, 8 insertions and 3 deletions (Supplementary Material, Table S3). Specifically, we detected 69 mutations at RIG-I, 78 at IFIH1 and 59 at LGP2, of which 14, 16 and 11 corresponded to non-synonymous variants, respectively (Table 1 and Fig. 1).

Varying levels of nucleotide and functional diversity among RLRs

Globally, RIG-I presented lower levels of nucleotide diversity per site (π) (4.5 × 10⁻⁴) than IFIH1 and LGP2, which displayed similar values (6.7 × 10⁻⁴ and 7.4 × 10⁻⁴, respectively) (Table 2). To compare these levels of nucleotide diversity with background genetic expectations, we used the SeattleSNPs database, which reports the population sequence diversity of 327 genes associated with inflammatory pathways. As the SeattleSNPs database does not include Asian populations, we excluded our Asian sample from this analysis to have comparable panels. As a consequence, the overall levels of diversity at each gene were slightly increased, because the Asian population is known to present the lowest genetic diversity, as a result of greater genetic drift (52). RIG-I (5.1 × 10⁻⁴)
A full description of all SNPs identified in this study is available in Supplementary Material, Table S3.

Table 1. Non-synonymous changes identified in RLRs in this study

<table>
<thead>
<tr>
<th>Gene</th>
<th>SNP position</th>
<th>AA change</th>
<th>Domain</th>
<th>Polyphen 2</th>
<th>dbSNP</th>
<th>Allele frequency</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Africa</td>
</tr>
<tr>
<td>RIG-I</td>
<td>19 C&gt;T</td>
<td>R7C</td>
<td>CARD1</td>
<td>P damaging</td>
<td>rs10813831</td>
<td>18.55</td>
</tr>
<tr>
<td>RIG-I</td>
<td>23 G&gt;A</td>
<td>S8N</td>
<td>CARD1</td>
<td>Benign</td>
<td>–</td>
<td>0.81</td>
</tr>
<tr>
<td>RIG-I</td>
<td>88 G&gt;A</td>
<td>A30T*</td>
<td>CARD1</td>
<td>Benign</td>
<td>–</td>
<td>1.61</td>
</tr>
<tr>
<td>RIG-I</td>
<td>89 C&gt;T</td>
<td>A30V*</td>
<td>CARD1</td>
<td>Benign</td>
<td>–</td>
<td>1.61</td>
</tr>
<tr>
<td>RIG-I</td>
<td>25272 A&gt;G</td>
<td>M51V</td>
<td>CARD1</td>
<td>Benign</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td>RIG-I</td>
<td>25333 G&gt;A</td>
<td>R71H</td>
<td>CARD1</td>
<td>P damaging</td>
<td>rs72710678</td>
<td>0.81</td>
</tr>
<tr>
<td>RIG-I</td>
<td>32398 A&gt;C</td>
<td>I139L</td>
<td>CARD2</td>
<td>P damaging</td>
<td>rs7886043</td>
<td>4.84</td>
</tr>
<tr>
<td>RIG-I</td>
<td>33636 C&gt;T</td>
<td>S144F</td>
<td>CARD2</td>
<td>Benign</td>
<td>rs55789327</td>
<td>6.45</td>
</tr>
<tr>
<td>RIG-I</td>
<td>36758 A&gt;G</td>
<td>N245S</td>
<td>Helicase</td>
<td>Benign</td>
<td>–</td>
<td>4.03</td>
</tr>
<tr>
<td>RIG-I</td>
<td>44826 G&gt;A</td>
<td>R546Q</td>
<td>Helicase</td>
<td>P damaging</td>
<td>rs61752945</td>
<td>2.42</td>
</tr>
<tr>
<td>RIG-I</td>
<td>45914 T&gt;A</td>
<td>D580E</td>
<td>Helicase</td>
<td>Ps damaging</td>
<td>rs17217280</td>
<td>11.29</td>
</tr>
<tr>
<td>RIG-I</td>
<td>45957 G&gt;A</td>
<td>E773K</td>
<td>RD</td>
<td>Benign</td>
<td>–</td>
<td>4.03</td>
</tr>
<tr>
<td>RIG-I</td>
<td>68933 T&gt;C</td>
<td>I889T</td>
<td>RD</td>
<td>P damaging</td>
<td>–</td>
<td>0.81</td>
</tr>
<tr>
<td>RIG-I</td>
<td>69014 T&gt;C</td>
<td>I916T</td>
<td>RD</td>
<td>Benign</td>
<td>rs77788008</td>
<td>4.03</td>
</tr>
<tr>
<td>IFIHI</td>
<td>413 G&gt;T</td>
<td>C138F</td>
<td>CARD2</td>
<td>P damaging</td>
<td>–</td>
<td>0.81</td>
</tr>
<tr>
<td>IFIHI</td>
<td>30124 A&gt;G</td>
<td>K349H</td>
<td>Helicase</td>
<td>Benign</td>
<td>rs72650664</td>
<td>0.81</td>
</tr>
<tr>
<td>IFIHI</td>
<td>35733 T&gt;A</td>
<td>V366E</td>
<td>Helicase</td>
<td>P damaging</td>
<td>–</td>
<td>0.81</td>
</tr>
<tr>
<td>IFIHI</td>
<td>35817 C&gt;A</td>
<td>T394N</td>
<td>Helicase</td>
<td>Benign</td>
<td>–</td>
<td>0.81</td>
</tr>
<tr>
<td>IFIHI</td>
<td>35593 T&gt;C</td>
<td>V432A</td>
<td>Helicase</td>
<td>Ps damaging</td>
<td>rs10930046</td>
<td>52.42</td>
</tr>
<tr>
<td>IFIHI</td>
<td>36835 G&gt;A</td>
<td>R460H</td>
<td>Helicase</td>
<td>P damaging</td>
<td>–</td>
<td>3.23</td>
</tr>
<tr>
<td>IFIHI</td>
<td>36937 A&gt;G</td>
<td>G494V</td>
<td>Helicase</td>
<td>P damaging</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td>IFIHI</td>
<td>45914 T&gt;A</td>
<td>D580E</td>
<td>Helicase</td>
<td>Ps damaging</td>
<td>rs17217280</td>
<td>11.29</td>
</tr>
<tr>
<td>IFIHI</td>
<td>45961 G&gt;A</td>
<td>A946T</td>
<td>CTD</td>
<td>P damaging</td>
<td>–</td>
<td>1.61</td>
</tr>
<tr>
<td>IFIHI</td>
<td>50181 A&gt;G</td>
<td>I923V</td>
<td>CTD</td>
<td>Ps damaging</td>
<td>rs35667974</td>
<td>0.81</td>
</tr>
<tr>
<td>IFIHI</td>
<td>50199 G&gt;A</td>
<td>V929I</td>
<td>CTD</td>
<td>P damaging</td>
<td>–</td>
<td>1.61</td>
</tr>
<tr>
<td>IFIHI</td>
<td>50767 G&gt;A</td>
<td>A946T</td>
<td>CTD</td>
<td>Benign</td>
<td>rs1990760</td>
<td>11.29</td>
</tr>
<tr>
<td>IFIHI</td>
<td>50930 C&gt;A</td>
<td>A967D</td>
<td>CTD</td>
<td>Benign</td>
<td>–</td>
<td>0.81</td>
</tr>
<tr>
<td>LGP2</td>
<td>125 T&gt;C</td>
<td>L42P</td>
<td>Helicase</td>
<td>P damaging</td>
<td>–</td>
<td>0.81</td>
</tr>
<tr>
<td>LGP2</td>
<td>453 A&gt;G</td>
<td>T76A</td>
<td>Helicase</td>
<td>Benign</td>
<td>rs34891485</td>
<td>2.42</td>
</tr>
<tr>
<td>LGP2</td>
<td>511 G&gt;A</td>
<td>R95Q</td>
<td>Helicase</td>
<td>Ps damaging</td>
<td>rs35118457</td>
<td>2.42</td>
</tr>
<tr>
<td>LGP2</td>
<td>994 G&gt;A</td>
<td>V129M</td>
<td>Helicase</td>
<td>P damaging</td>
<td>–</td>
<td>0.81</td>
</tr>
<tr>
<td>LGP2</td>
<td>5906 C&gt;T</td>
<td>R334C</td>
<td>Helicase</td>
<td>Ps damaging</td>
<td>rs76998797</td>
<td>1.61</td>
</tr>
<tr>
<td>LGP2</td>
<td>6748 A&gt;G</td>
<td>Q425R</td>
<td>Helicase</td>
<td>Benign</td>
<td>rs2074158</td>
<td>84.68</td>
</tr>
<tr>
<td>LGP2</td>
<td>6856 A&gt;G</td>
<td>N461S</td>
<td>Helicase</td>
<td>P damaging</td>
<td>rs34016093</td>
<td>2.42</td>
</tr>
<tr>
<td>LGP2</td>
<td>7048 T&gt;C</td>
<td>I495T</td>
<td>RD</td>
<td>Benign</td>
<td>–</td>
<td>1.61</td>
</tr>
<tr>
<td>LGP2</td>
<td>8099 G&gt;A</td>
<td>R523Q</td>
<td>RD</td>
<td>Benign</td>
<td>rs7241640</td>
<td>21.77</td>
</tr>
<tr>
<td>LGP2</td>
<td>8168 A&gt;G</td>
<td>Q546R</td>
<td>RD</td>
<td>Benign</td>
<td>–</td>
<td>2.42</td>
</tr>
<tr>
<td>LGP2</td>
<td>10056 G&gt;A</td>
<td>R654H</td>
<td>RD</td>
<td>Benign</td>
<td>–</td>
<td>0.81</td>
</tr>
</tbody>
</table>

A full description of all SNPs identified in this study is available in Supplementary Material, Table S3.

*aSNP numbering is related to the ATG position, the first nucleotide corresponding to the ancestral allele.

*bAA change stands for ‘amino acid change’, the first amino acid corresponding to the ancestral state.

*c‘Ps damaging’ stands for ‘possibly damaging’, and ‘P damaging’ for ‘probably damaging’.

*dThe frequencies (in %) of each SNP in the different populations refer to the derived allele frequency.

*Nucleotide substitutions at positions 88 and 89 are always observed together; consequently, the actual amino acid substitution is A30I, which is predicted as benign by Polyphen v2.

was found to be in the 17th percentile of genes presenting the lowest nucleotide diversity, whereas IFIHI (7.9×10^-4) and LGP2 (2.2×10^-5) felt in the 54th and 58th, respectively. At the continental level, we observed fluctuations in the levels of nucleotide diversity across populations, the most extreme case being that of IFIHI, which displays very high levels of nucleotide diversity in Africa while very low in European and Asian populations (Supplementary Material, Fig. S2). We next focused on the comparison of non-synonymous/synonymous ratios between species (divergence, dS/dN) and within humans (polymorphism, πS/πS). Strong differences among genes emerged from our data: RIG-I exhibited a dramatic decay in its dS/dN ratio with respect to the other RLRs (Fig. 2), consistent with a strong deficit of fixed non-synonymous mutations (i.e. no non-synonymous mutations fixed between the chimpanzee and human lineages were detected). Furthermore, when comparing the πS/πS ratios between RLRs, we observed both a smaller proportion and a lower frequency of segregating non-synonymous polymorphisms for RIG-I (Fig. 2).

Finally, we predicted the functional effects at the protein level of the 41 non-synonymous mutations at the RLRs found to segregate in the population (Table 1), using the Polyphen v2 HumDiv algorithm (53). This method, which considers protein structure and/or sequence conservation for each gene, has been shown to be the best available predictor of
the fitness effects of non-synonymous variants (53,54). IFIH1 exhibited the highest proportion of ‘probably damaging’ mutations and these mutations were found at high population frequencies. In contrast, RIG-I and LGP2 displayed a relatively low proportion of probably damaging mutations, which were observed at low population frequencies (Fig. 3). In light of this, RIG-I and LGP2 seem to be much more constrained with respect to non-synonymous mutations that are likely to have an impact on protein function.

Low amino acid-altering variation in the RIG-I helicase domain

We next aimed to understand whether constraints of amino acid sequence variation are evenly distributed among the different protein domains. To this end, we compared the levels of non-synonymous and synonymous polymorphisms, considering both the number of nucleotide substitutions \((p_N/p_S)\) and the average number of differences between pairs of individuals \((\pi_N/\pi_S)\), the latter of which also takes into account population frequency data. With respect to the CARD domain, involved in cell death and inflammation signalling, IFIH1 exhibits a lower occurrence of non-synonymous mutations than RIG-I, as attested to by the \(p_N/p_S\) ratio (Fig. 4A). Moreover, non-synonymous mutations in the CARD domain for IFIH1 were generally at low frequency; the \(\pi_N/\pi_S\) of IFIH1 was in the lowest 25% of the resampled values while that of RIG-I was among the 14% highest, although it did not reach statistical significance. With respect to the Helicase domain, RIG-I exhibits a lower \(p_N/p_S\) ratio than IFIH1 and LGP2 (Fig. 4B), as well as a significant deficit in high-frequency non-synonymous mutations (resampling \(P < 0.0001\)). Finally, with respect to the CTD/RD domain, primarily responsible for RNA recognition, RIG-I displayed lower \(p_N/p_S\) and \(\pi_N/\pi_S\) ratios than the other two RLRs, although this did not reach statistical significance (Fig. 4C). Overall, the comparisons of \(p_N/p_S\) and \(\pi_N/\pi_S\) ratios per domain illustrated lower occurrences and frequencies of non-synonymous mutations in the helicase and CTD domains at RIG-I than at IFIH1 and LGP2, a deficit that was significant in the case of the helicase domain.

Significant deviations from neutrality at IFIH1

We tested for deviations of the allele frequency spectrum from neutral expectations over the genomic regions encompassing the three RLRs. We used summary statistics based on within-population allele frequency distribution, such as Tajima’s \(D\), Fu & Li’s \(D^*\) & \(F^*\) and normalized Fay & Wu’s \(H\) (reviewed in 55). As these tests are known to be sensitive to demographic effects, we considered two previously validated demographic models based on a set of unlinked noncoding regions sequenced in a panel of populations from Africa, Europe and Asia (56,57). The main difference between the two models is that the one of Voight et al. (57) does not consider inter-continental population migration, while the one of Laval et al. (56) does. Our analyses unmasked an excess of low-frequency variants for the three genes, indicative of either negative or positive selection. This was illustrated by the significantly negative values of the various neutrality tests, which were virtually restricted to the European population (Table 2). To be conservative, we focused mainly on those signatures of selection that turned out to be significant after considering both demographic models. The strongest signals were clearly detected at IFIH1, as illustrated by the significantly negative Tajima’s \(D\), Fu & Li’s \(D^*\) and \(F^*\) in Europe. With respect to the Asian population, we also detected a signature, although weaker, of positive selection at IFIH1, as attested to by the significantly negative value of the Fay and Wu’s \(H\) statistics, which detects an excess of high-frequency-derived alleles. This signature was supported by the Tajima’s \(D\) values, which were significantly negative (or close to significance) for the two models in Asia (Table 2). Overall, the strongest

Figure 1. Distribution of the non-synonymous variants identified in this study across the RLRs. The location of each non-synonymous variant within the different protein domains is shown.
deviations from neutrality were observed in IFIHI1, especially in Europe.

**Strong population differentiation at IFIHI1 and LGP2**

To gain further insight into the signatures of selection detected by the previous sequence-based neutrality tests, we estimated the degree of population differentiation by means of FST. Local positive selection is known to increase the levels of population differentiation with respect to neutrally evolving loci (55,58–61). When calculating FST averaged across each gene, both worldwide and population-pairwise, IFIHI1 and LGP2 were strongly differentiated between populations (FST = 0.25 and 0.40, respectively) with respect to RIG-I (FST = 0.05). The very high FST displayed by LGP2 (FST = 0.40) with respect to the mean across the genome (FST = 0.15) suggests that this gene has been targeted by population-specific positive selection. The analysis of pairwise FST revealed that the strong differentiation observed at IFIHI1 and LGP2 is mainly accounted for by differences between Africans and non-Africans (IFIHI1 FST = 0.30, 0.23 and 0.12 and LGP2 FST = 0.46, 0.51 and 0.05 for Africa/Europe, Africa/Asia and Europe/Asia, respectively), whereas all pairwise FST gave similarly low values for RIG-I (FST = 0.05, 0.06 and 0.06).

At the level of individual single nucleotide polymorphisms (SNPs) (i.e. 206 SNPs), strongly differentiated SNPs appeared to be restricted to IFIHI1 and LGP2, when estimating FST as a function of expected heterozygosity (Fig. 5). IFIHI1 presented 11 highly differentiated SNPs in the Africa/Europe comparison, with respect to the rest of the genome: SNPs 29773 (P = 0.03), 36033 (P = 0.009), 36835 (P = 0.007), 36947 (P = 0.009), 38047 (P = 0.009), 38091 (P = 0.013), 40330 (P = 0.009), 45935 (P = 0.009), 46093 (P = 0.008), 49917 (P = 0.008) and 51114 (P = 0.009), including one non-synonymous SNP in the helicase domain (SNP 36835, R460H). Three of these highly differentiated SNPs were also significant outliers in the Africa/Asia comparison (SNPs 40330, 45935 and 51114, P = 0.013), while the non-synonymous SNP 36835 showed a very high FST but did not reach statistical significance. The
derived state of SNP 36835 was found at a frequency of ~52% in Africa, at fixation in Europe and at ~93% frequency in Asia (Supplementary Material, Table S3). Concerning LD levels among highly differentiated SNPs at \textit{IFIH1}, we found a group of seven SNPs (SNPs 36033, 36947, 38047, 38091, 40330, 45935 and 51114) and a group of four (29773, 36835, 46093 and 49917) in high LD ($r^2 > 0.8$) in Africa, and a group of seven SNPs (SNPs 36033, 36835, 36947, 38047, 38091, 46093 and 49917) in complete LD in Asia. No SNPs with a MAF > 0.05 were found in high LD ($r^2 > 0.8$) in Europe (Supplementary Material, Fig. S3).

With respect to \textit{LGP2}, we detected five SNPs (SNPs −751, −215, 3931, 5784 and 6748) that were outliers for the Africa/Europe ($P = 0.009, 0.01, 0.01, 0.016$ and 0.042, respectively) and the Africa/Asia ($P = 0.001, 0.023, 0.023, 0.054$ and 0.043, respectively) comparisons (Fig. 5). Three of these five SNPs are in LD in all populations, namely SNP −215 with SNPs 3931 and 5784. SNP −751 is in strong LD with these SNPs in Africa and Europe only (Supplementary Material, Fig. S3). Interestingly, SNPs −751 and −215 are located in the potential \textit{cis}-regulatory region of the gene. SNP 6748, which is in low LD with the ‘potential regulatory’ block of LD (Supplementary Material, Fig. S3), is a non-synonymous mutation (Q425R) located, as for \textit{IFIH1}, in the helicase domain.

**The footprints of recent positive selection targeting \textit{IFIH1} and \textit{LGP2}**

Finally, we searched for signatures of population-specific positive selection by means of several LD-based tests. Specifically, we used statistics based on specific signatures of recent positive selection including low levels of derived intra-allelic nucleotide diversity (DIND) (9) and extended haplotype homozygosity [integrated haplotype scores (iHS), cross population extended haplotype homozygosity (XP-EHH) and extended haplotype homozygosity (EHH)] (61–63). The strongest signatures of positive selection were again observed at \textit{IFIH1} and \textit{LGP2}, identifying potential targets within both of them. For \textit{IFIH1}, SNP 36835, which is found to be highly differentiated in Europe and to a lesser extent in Asia based on $F_{ST}$, was not assessed for EHH in Europeans because it reaches fixation. In Asia, the levels of EHH around the derived allele of SNP 36835 were similar to those observed in Africa (Supplementary Material, Fig. S4). In Africa, SNP 36835 (together with SNPs 46093 and 49917 which are in complete LD, see Supplementary Material, Fig. S3) appeared to be an outlier in the DIND test when using the recombination rates of UCSC ($P = 0.006$ for Laval’s model and $P = 0.039$ for Voight’s model) but not when using HapMap ($P = 0.164$ for Laval’s model and $P = 0.181$ for Voight’s model). In this view, LD-based tests did not support the notion that the signature of positive selection at SNP 36835 revealed by $F_{ST}$ was of recent nature. In contrast, the DIND test identified another non-synonymous mutation (SNP 45994, R843H) as a clearer outlier in Africa ($P = 0.000$ Laval/UCSC; $P = 0.002$ Voight/UCSC; $P = 0.023$ Laval/HapMap; $P = 0.027$ Voight/HapMap) and, to a lesser extent, in Asia ($P = 0.0128$ Laval/UCSC; $P = 0.048$ Voight/UCSC; $P = 0.080$ Laval/HapMap; $P = 0.116$ Voight/HapMap) (Supplementary Material, Fig. S5). Likewise, EHH analysis for SNP 45994 revealed a much higher homozygosity for the derived haplotype in Africa and Asia (Supplementary Material, Fig. S4). Note that SNPs 36835 and 45994 display intermediate LD levels in Africa and low in Asia (Supplementary Material, Fig. S3) (LD was not assessed in Europe because of the derived allele fixation of SNP 36835). Finally, the DIND test provided strongly significant values for two additional non-coding SNPs in complete LD in Africa (Supplementary Material, Fig. S3), SNPs 36531/41712 ($P = 4 \times 10^{-4}$ Laval/UCSC; $P = 0.004$ Voight/UCSC; $P = 0.018$ Laval/HapMap; $P = 0.022$ Voight/HapMap), but they present a frequency lower than 0.20, at which the DIND test has been shown to have a higher false discovery rate (9).

With respect to \textit{LGP2}, the non-synonymous SNP 6748, which was found to be highly differentiated between Africans and non-Africans based on $F_{ST}$, exhibited an iHS value in Europe of 2.3, suggesting the action of positive selection on the ancestral allele. In Asia, the iHS value was of recent nature. In contrast, the DIND test identified another non-synonymous mutation (SNP 45994, R843H) as a clearer outlier in Africa ($P = 0.000$ Laval/UCSC; $P = 0.002$ Voight/UCSC; $P = 0.023$ Laval/HapMap; $P = 0.027$ Voight/HapMap) and, to a lesser extent, in Asia ($P = 0.0128$ Laval/UCSC; $P = 0.048$ Voight/UCSC; $P = 0.080$ Laval/HapMap; $P = 0.116$ Voight/HapMap) (Supplementary Material, Fig. S5). Likewise, EHH analysis for SNP 45994 revealed a much higher homozygosity for the derived haplotype in Africa and Asia (Supplementary Material, Fig. S4). Note that SNPs 36835 and 45994 display intermediate LD levels in Africa and low in Asia (Supplementary Material, Fig. S3) (LD was not assessed in Europe because of the derived allele fixation of SNP 36835). Finally, the DIND test provided strongly significant values for two additional non-coding SNPs in complete LD in Africa (Supplementary Material, Fig. S3), SNPs 36531/41712 ($P = 4 \times 10^{-4}$ Laval/UCSC; $P = 0.004$ Voight/UCSC; $P = 0.018$ Laval/HapMap; $P = 0.022$ Voight/HapMap), but they present a frequency lower than 0.20, at which the DIND test has been shown to have a higher false discovery rate (9).

![Figure 3](https://example.com/fig3.png) Distribution of functional diversity among the different RLRs. The functional impact of each non-synonymous mutation observed at \textit{RIG-I}, \textit{IFIH1} and \textit{LGP2} was predicted by means of the Polyphen algorithm v2 HumDiv. (A) Proportion of ‘Probably damaging’ mutations among non-synonymous mutations and (B) proportion of chromosomes carrying at least one ‘probably damaging’ mutation.
Figure 4. Occurrence and frequency of non-synonymous mutations across the different RLR protein domains. Non-synonymous/synonymous polymorphisms in humans ($p_N/p_S$), and non-synonymous/synonymous diversity ($\pi_N/\pi_S$) assessed by the mean numbers of pairwise differences between human sequences ($\pi$), for RIG-I, IFIH1 and LGP2 in (A) the CARD domain, (B) the helicase domain and (C) the CTD/RD domain. Grey and black bars represent $p_N/p_S$ and $\pi_N/\pi_S$, respectively.
Indeed, as well as by its low tolerance of amino acid-altering variation. Levels of nucleotide diversity and population differentiation vary substantially among the three RLR members. Selection and with different intensities. Selection has driven the evolution of RLRs, in different directions and with different intensities. Results presented herein showed indeed that natural selection has driven the evolution of RLRs, in different directions and with different intensities.

DISCUSSION

Detecting to what extent natural selection has targeted host microbial sensors, such as the RLRs, can provide insight into their respective biological roles in host defence as well as into the genetic basis of their phenotypic variation. Because RLRs constitute key players in the initial sensing of RNA viruses, genetic variation at these genes may impact downstream immune responses and therefore contribute to the susceptibility or pathogenesis of various infectious diseases. This makes these genes putative substrates of natural selection. Results presented herein showed indeed that natural selection has driven the evolution of RLRs, in different directions and with different intensities.

First, we showed that the degree of gene and protein diversity vary substantially among the three RLR members. RIG-I is clearly the most constrained RLR, as attested by its low levels of nucleotide diversity and population differentiation as well as by its low tolerance of amino acid-altering variation. Indeed, RIG-I presents fewer, and lower frequencies of, non-synonymous mutations than IFIH1 and LGP2. Such a constraint may reflect the pressure imposed by the specific viruses recognized by RIG-I with respect to IFIH1; RIG-I senses a much larger variety of viruses, including several families such as paramyxoviridae, rhabdoviridae, filoviridae, arenaviridae and coronaviridae, whereas IFIH1 is mainly involved in the recognition of picornaviruses [see (11) and references therein]. In addition, the structural nature of the viral substrates of RIG-I may also account for the stronger constraint observed. Indeed, RIG-I and IFIH1 detect short and long dsRNA, respectively, but RIG-I further senses 5′PPPssRNA (20,30,31). Short dsRNA and 5′PPPssRNA binding could require more strict binding sites, thus allowing for less protein flexibility, compared with long dsRNA. In this context, it is interesting to note that the C-terminal and helicase domains of RIG-I, notably involved in RNA recognition, are more constrained with respect to amino acid-altering variation than in the other RLRs. Our results suggest that, in the context of sensing RNA viruses, the function fulfilled by RIG-I may allow for less protein variation than that of IFIH1. However, viral presence may not be the only force driving the evolution of these microbial sensors, as suggested by experimental data in mice. Indeed, both RIG-I+/− and IFIH1−/− mice are highly susceptible, to varying extents, to specific RNA virus infections. However, while IFIH1−/− mice grow healthily without any developmental abnormalities before 24 weeks of age, most RIG-I−/− mice have been shown to be embryonic lethal due to developmental defects (18,20). These observations suggest a generally more important role for RIG-I compared with IFIH1, which may go well beyond immunity to infection. In light of this, our evolutionary findings in humans could also reflect different levels of involvement for these two genes in other functions, such as reproduction and development. This would not be surprising, considering that mutations in various NOD-like receptors—another major family of microbial sensors—have been shown to lead to reproductive wastage or abnormalities in early development both in mice and humans (64–67).

Secondly, the patterns of diversity displayed by IFIH1 and LGP2 indicate that these two genes, in contrast to RIG-I, have

Figure 5. Levels of population differentiation displayed by the three RLR members. FST as a function of heterozygosity per SNP is shown between (A) Africans versus Europeans, (B) Africans versus East-Asians and (C) Europeans versus East-Asians, for RIG-I, IFIH1 and LGP2. The 95th and 99th percentiles of the HGDP–CEPH genotyping data set using the same individuals are represented with dashed lines, while the density blue area corresponds to the 99.9th percentile. Black and red points represent silent and non-synonymous SNPs, respectively. For each outlier SNP, the gene name is indicated, followed by position to ATG. Groups of outlying SNPs separated by a comma correspond to SNPs in complete LD. Outlying non-synonymous SNP positions are underlined.
evolved under more relaxed selective constraints. Most importantly, different lines of evidence based on the allele frequency spectrum, the levels of population differentiation and LD-based tests clearly show that IFIH1 and LGP2 appear to evolve adaptively. The action of positive selection at IFIH1 has been previously proposed, with the non-synonymous SNP 36835 (R460H) suggested as a plausible target (47). Our study extends previous observations and clarifies the nature of the selective event targeting IFIH1. Our \( F_{ST} \) analysis clearly indicates that the R460H variant is the most differentiated non-synonymous SNP in Eurasian populations, pointing to this variant as the genuine target of selection. Further support for the action of positive selection comes from the median-joining network, where the branch carrying the derived 460H allele includes virtually all Europeans and Asians (Supplementary Material, Fig. S7). In addition, the absence of signals of recent positive selection at this SNP based on LD-based tests (iHS test) (61), which is not surprising given the very high frequencies of the derived allele (100 and 92.7% in Europe and Asia, respectively), suggests that the selective event targeting R460H is of a sufficiently ancient nature to have allowed recombination to reduce haplotype lengths. In support of this, the results of the XP-EHH test over the entire gene were close to the threshold of 2 in Europe and East-Asia, but were not significant (data not shown). Interestingly, the ancestral allele at this position (460R) has been associated with higher resistance against psoriasis, a chronic autoimmune disease caused by an abnormal speeding up of growth cycle for skin cells (68). It is thus tempting to hypothesize that the derived 460H variant has increased in frequency among non-Africans owing to an increased advantage in host defence, despite its deleterious, rather minor effects associated with higher risk of psoriasis. This provides an additional example that supports the notion whereby the present-day increased incidence of autoimmune, inflammatory or allergic diseases may result from past adaptation to infectious agents (46,48).

In addition, our study has identified another non-synonymous variant at IFIH1, the SNP 45994 (R843H), which exhibit signatures of positive selection in Africa and, to a lesser extent, in Asia, based on LD tests. The R843H variant has been suggested to behave neutrally or almost neutrally and to have hitch-hiked with the genuine selective variant, the SNP 36835-R460H variant (47). Our data challenge this view, because the strongest signatures on SNPs 45994 were found in Africa, while those detected at SNP 45994 reflect a selective event of different and more ancient nature to have allowed recombination to reduce haplotype lengths. In support of this, the results of the XP-EHH test over the entire gene were close to the threshold of 2 in Europe and East-Asia, but were not significant (data not shown). Interestingly, the ancestral allele at this position (460R) has been associated with higher resistance against psoriasis, a chronic autoimmune disease caused by an abnormal speeding up of growth cycle for skin cells (68). It is thus tempting to hypothesize that the derived 460H variant has increased in frequency among non-Africans owing to an increased advantage in host defence, despite its deleterious, rather minor effects associated with higher risk of psoriasis. This provides an additional example that supports the notion whereby the present-day increased incidence of autoimmune, inflammatory or allergic diseases may result from past adaptation to infectious agents (46,48).

Our study extends previous observations and clarifies the nature of the selective event targeting IFIH1. Our \( F_{ST} \) analysis clearly indicates that the R460H variant is the most differentiated non-synonymous SNP in Eurasian populations, pointing to this variant as the genuine target of selection. Further support for the action of positive selection comes from the median-joining network, where the branch carrying the derived 460H allele includes virtually all Europeans and Asians (Supplementary Material, Fig. S7). In addition, the absence of signals of recent positive selection at this SNP based on LD-based tests (iHS test) (61), which is not surprising given the very high frequencies of the derived allele (100 and 92.7% in Europe and Asia, respectively), suggests that the selective event targeting R460H is of a sufficiently ancient nature to have allowed recombination to reduce haplotype lengths. In support of this, the results of the XP-EHH test over the entire gene were close to the threshold of 2 in Europe and East-Asia, but were not significant (data not shown). Interestingly, the ancestral allele at this position (460R) has been associated with higher resistance against psoriasis, a chronic autoimmune disease caused by an abnormal speeding up of growth cycle for skin cells (68). It is thus tempting to hypothesize that the derived 460H variant has increased in frequency among non-Africans owing to an increased advantage in host defence, despite its deleterious, rather minor effects associated with higher risk of psoriasis. This provides an additional example that supports the notion whereby the present-day increased incidence of autoimmune, inflammatory or allergic diseases may result from past adaptation to infectious agents (46,48).

In addition, our study has identified another non-synonymous variant at IFIH1, the SNP 45994 (R843H), which exhibit signatures of positive selection in Africa and, to a lesser extent, in Asia, based on LD tests. The R843H variant has been suggested to behave neutrally or almost neutrally and to have hitch-hiked with the genuine selective variant, the SNP 36835-R460H variant (47). Our data challenge this view, because the strongest signatures on SNPs 45994 were found in Africa, while those detected at SNP 45994 reflect a selective event of different and more ancient nature to have allowed recombination to reduce haplotype lengths. In support of this, the results of the XP-EHH test over the entire gene were close to the threshold of 2 in Europe and East-Asia, but were not significant (data not shown). Interestingly, the ancestral allele at this position (460R) has been associated with higher resistance against psoriasis, a chronic autoimmune disease caused by an abnormal speeding up of growth cycle for skin cells (68). It is thus tempting to hypothesize that the derived 460H variant has increased in frequency among non-Africans owing to an increased advantage in host defence, despite its deleterious, rather minor effects associated with higher risk of psoriasis. This provides an additional example that supports the notion whereby the present-day increased incidence of autoimmune, inflammatory or allergic diseases may result from past adaptation to infectious agents (46,48).

In addition, our study has identified another non-synonymous variant at IFIH1, the SNP 45994 (R843H), which exhibit signatures of positive selection in Africa and, to a lesser extent, in Asia, based on LD tests. The R843H variant has been suggested to behave neutrally or almost neutrally and to have hitch-hiked with the genuine selective variant, the SNP 36835-R460H variant (47). Our data challenge this view, because the strongest signatures on SNPs 45994 were found in Africa, while those detected at SNP 45994 reflect a selective event of different and more ancient nature to have allowed recombination to reduce haplotype lengths. In support of this, the results of the XP-EHH test over the entire gene were close to the threshold of 2 in Europe and East-Asia, but were not significant (data not shown). Interestingly, the ancestral allele at this position (460R) has been associated with higher resistance against psoriasis, a chronic autoimmune disease caused by an abnormal speeding up of growth cycle for skin cells (68). It is thus tempting to hypothesize that the derived 460H variant has increased in frequency among non-Africans owing to an increased advantage in host defence, despite its deleterious, rather minor effects associated with higher risk of psoriasis. This provides an additional example that supports the notion whereby the present-day increased incidence of autoimmune, inflammatory or allergic diseases may result from past adaptation to infectious agents (46,48).

In addition, our study has identified another non-synonymous variant at IFIH1, the SNP 45994 (R843H), which exhibit signatures of positive selection in Africa and, to a lesser extent, in Asia, based on LD tests. The R843H variant has been suggested to behave neutrally or almost neutrally and to have hitch-hiked with the genuine selective variant, the SNP 36835-R460H variant (47). Our data challenge this view, because the strongest signatures on SNPs 45994 were found in Africa, while those detected at SNP 45994 reflect a selective event of different and more ancient nature to have allowed recombination to reduce haplotype lengths. In support of this, the results of the XP-EHH test over the entire gene were close to the threshold of 2 in Europe and East-Asia, but were not significant (data not shown). Interestingly, the ancestral allele at this position (460R) has been associated with higher resistance against psoriasis, a chronic autoimmune disease caused by an abnormal speeding up of growth cycle for skin cells (68). It is thus tempting to hypothesize that the derived 460H variant has increased in frequency among non-Africans owing to an increased advantage in host defence, despite its deleterious, rather minor effects associated with higher risk of psoriasis. This provides an additional example that supports the notion whereby the present-day increased incidence of autoimmune, inflammatory or allergic diseases may result from past adaptation to infectious agents (46,48).

In addition, our study has identified another non-synonymous variant at IFIH1, the SNP 45994 (R843H), which exhibit signatures of positive selection in Africa and, to a lesser extent, in Asia, based on LD tests. The R843H variant has been suggested to behave neutrally or almost neutrally and to have hitch-hiked with the genuine selective variant, the SNP 36835-R460H variant (47). Our data challenge this view, because the strongest signatures on SNPs 45994 were found in Africa, while those detected at SNP 45994 reflect a selective event of different and more ancient nature to have allowed recombination to reduce haplotype lengths. In support of this, the results of the XP-EHH test over the entire gene were close to the threshold of 2 in Europe and East-Asia, but were not significant (data not shown). Interestingly, the ancestral allele at this position (460R) has been associated with higher resistance against psoriasis, a chronic autoimmune disease caused by an abnormal speeding up of growth cycle for skin cells (68). It is thus tempting to hypothesize that the derived 460H variant has increased in frequency among non-Africans owing to an increased advantage in host defence, despite its deleterious, rather minor effects associated with higher risk of psoriasis. This provides an additional example that supports the notion whereby the present-day increased incidence of autoimmune, inflammatory or allergic diseases may result from past adaptation to infectious agents (46,48).
LGP2, highly differentiating Africans and non-Africans, and support the notion that some genes appear to be more subject to recurrent events of positive selection (70).

More generally, in the context of pathogen sensing of RNA viruses, RLRs are not the only family that induce antiviral immunity but are part of a concerted antiviral program mediated by a variety of microbial sensors including the TLRs (11,13). Specifically, among the 10 functional members of the TLR family in humans, the endosomally located TLR3, TLR7 and TLR8 have been characterized as principal sensors of RNA viruses, while the other TLRs are responsible for detecting bacteria, fungi and DNA viruses (4,6,71). Earlier work from our laboratory has shown that endosomally located TLRs evolve under strong purifying selection, indicating their essential role in host survival, while the remaining TLRs display higher levels of immunological redundancy (9). When comparing the two groups of microbial sensors specialized in the detection of viral RNA (i.e. RLRs versus TLRs 3/7/8), we found that RLRs and TLRs display similar levels of nucleotide diversity (\(\pi = 6.2 \times 10^{-4}\) and \(6.4 \times 10^{-4}\), respectively) and \(d_N/d_S\) ratios (0.18 and 0.23, respectively). However, we observed both a higher proportion and a higher frequency of segregating non-synonymous polymorphisms at RLRs than at TLRs, as attested by the 10-fold higher \(\pi_d/\pi_s\) ratio of RLRs with respect to TLRs (1.26 and 0.12, respectively). It should be noted that the very high \(\pi_d/\pi_s\) ratio of RLRs is mainly accounted by the values of LGP2; however, the combined \(\pi_d/\pi_s\) ratio of RIG-I and IFIH1 is 0.44, a value which is still three times higher than that observed at the TLRs. Furthermore, among all non-synonymous mutations identified, RLRs displayed a higher proportion of mutations predicted as having an impact on protein function [‘probably damaging’ mutations as predicted by Polyphen v2 (53)] with respect to the TLRs (39% for RLRs versus 15% for TLRs 3/7/8). Overall, our data suggest that the TLRs specialized in the detection of viral RNA have been under stronger evolutionary constraints than their RLR counterparts. Nonetheless, RLRs have been under stronger selective constraints than the other group of TLRs, the cell-surface TLRs. This group not only displays generally high levels of genetic and functional diversity but, in contrast to RLRs and TLR3/7/8, can accumulate stop mutations at high population frequencies (9). As TLRs 3/7/8 and RLRs share partially overlapping functions (8,17–19), the generally lower constraints observed at RLRs taken as a group could reflect some level of redundancy in their innate immunity functions.

In summary, our study has shown that natural selection has targeted differently the three members of the RLR family of microbial sensors. Evolutionary constraint in amino acid-altering variation at RIG-I, especially in its RNA recognition domains, suggests an important role for this gene in fighting against numerous and severe RNA viruses, structurally strict mechanisms of recognition or, more generally, an essential role in host survival. In turn, such evolutionary constraints have been much more relaxed at IFIH1 and LGP2, which appear to have evolved adaptively. We identified three non-synonymous mutations, two in IFIH1 (R460H and R843H) and one in LGP2 (Q425R), as well as several SNPs located in the regulatory region of LGP2 showing signatures of positive selection. Our study provides evolutionary evidence that these mutations have conferred a selective advantage to specific human populations, which may have impacted the sensing of RNA viruses by IFIH1 and the regulatory effect of LGP2 on RIG-I and/or IFIH1. Delineation of the functional impact of these mutations on immune responses to viral infections is now needed to understand the extent to which these variants have conferred a phenotypic advantage on the human host. Given the role of the RLRs in immunity to infection, particularly to RNA viruses, the variants identified as potential targets of positive selection should be further studied in the context of disease, as they may play a role in the susceptibility to, or pathogenesis of, infectious and autoimmune diseases.

**MATERIAL AND METHODS**

**DNA samples**

We resequenced RIG-I, IFIH1 and LGP2 in a total of 372 chromosomes from the human genome diversity panel (HGDP)—centre d’étude du polymorphisme humain (CEPH) panel (72). Specifically, this subpanel includes 62 sub-Saharan Africans, 62 Europeans and 62 East-Asians. Sub-Saharan African populations were composed of 19 Bantu from Kenya, 21 Mandenka from Senegal and 22 Yoruba from Nigeria; European populations include 20 French, 14 Italians, 6 Orcadians and 22 Russians; and East-Asian populations were composed of 10 Japanese, 4 Cambodians, 15 Han Chinese and 33 individuals from Chinese minorities. For a complete description of this HGDP–CEPH subpanel, see Supplementary Material, Table S1. This study was approved by the Institut Pasteur Institutional Review Board (no. RBM 2008.06).

**Resequencing**

For each gene, we resequenced the whole of the exonic region and at least an equivalent amount of non-exonic sequence, including intronic, 5’ and 3’ regions (Supplementary Material, Table S2). We used NM_014314.3, NM_022168.2 and NM_024119.2 as reference sequences for RIG-I, IFIH1 and LGP2, respectively. Sequence files and chromatograms were inspected using the GENALYS software (73). All sequences were analysed by two different operators to avoid SNP discovery errors, and ambiguous polymorphisms were systematically reapplied and resequenced. Concerning nucleotide numbering, SNP positions correspond to locations in genomic DNA, taking the A of the ATG translation initiation codon in the reference sequence as the +1 position. For various neutrality tests that need the definition of the ancestral/derived state at each SNP, we used the orthologous regions of each of the three RLRs from various species. To determine ancestral states at each SNP by parsimony, we used the UCSC database to retrieve the orthologous sequences of chimpanzee, gorilla, orangutan and rhesus macaque.

**Statistical analyses**

Haplotype reconstruction was performed by means of the Bayesian statistical method implemented in Phase (v.2.1.1) (74). We applied the algorithm five times, using different
randomly generated seeds, and consistent results were obtained across runs. LD levels between SNPs were determined for the three RLRs in Africa, Europe and Asia separately, using Haplovew 4.0 (75). We assessed the $d_s/d_S$ (divergent non-synonymous sites/divergent synonymous sites) ratio for the three genes using DnaSP package v. 5.1 (76); divergent sites refer to positions that are different between the human and chimpanzee lineages, whereas polymorphic sites refer to alleles that are segregating within humans. The different summary statistics, such as the number of segregating sites (S), haplotype diversity (Hd), the average number of pairwise differences ($\pi$) and the sequence-based neutrality tests, such as Tajima’s D, Fu and Li’s $D^*$, $F^*$ and normalized Fay and Wu’s $H$ tests were also performed using DnaSP. P-values for the various neutrality tests were estimated from $10^4$ coalescent simulations, performed using SIMCOAL 2.0 (77), under a finite-site neutral model and considering the recombination rate of the concerned region as reported in UCSC (78) and HapMap Phase II (79). Each of the $10^4$ coalescent simulations was conditional on the sample size and the number of segregating sites observed in each gene. To correct for the mimicking effects of demography on the patterns of diversity, we considered two previously validated demographic models based on resequencing data from noncoding autosomal regions in a set of populations similar to ours (i.e. African, European and Asian) (56,57). The main difference between these two demographic models is that the model of Laval et al. (56) considers inter-continental population migration.

To detect events of positive selection, we used a number of haplotype-based tests as well as levels of population differentiation. Specifically, we used the DIND test based on the ratio $\pi_A/\pi_D$, where $\pi_A$ and $\pi_D$ are the levels of nucleotide diversity associated with the haplotypes carrying the ancestral and the derived allele, respectively (9). The rationale behind this test is that a derived allele under positive selection that is at high population frequency should present lower levels of nucleotide diversity at linked sites than expected. We also used tests based on the levels of haplotype homozygosity, such as the extended haplotype homozygosity (EHH) web calculator (http://ihg2.helmholtz-muenchen.de/cgi-bin/mueller/webehh.pl) (80) and the XP-EHH test (63). When available, we also used the integrated iHS obtained from HapMap Phase II (61). To assess the levels of population differentiation for the entire SNP panel, we used the $F_{ST}$ statistics derived from the analysis of variance (81). To identify SNPs presenting extreme levels of population differentiation, we compared the observed $F_{ST}$ values at each SNP within RLRs against the $F_{ST}$ distribution of 659,000 SNPs genotyped in the same panel of individuals (82). $F_{ST}$ comparisons were conditioned to SNPs presenting similar expected heterozygosity. Median-joining networks were built using NETWORK 4.6 (83).

**FUNCTIONAL IMPACT OF NON-SYNONYMOUS MUTATIONS AND PROTEIN DOMAIN ANALYSIS**

The fitness status of all amino acid-altering mutations (i.e. benign, possibly damaging and probably damaging) was predicted using the Polyphen algorithm v2 HumDiv (53). Nucleotide diversity indexes such as $p_N/p_S$ and $\pi_N/\pi_S$ were assessed for each protein domain (CARD1, CARD2, helicase and CTD/ RD) and compared among RLRs. The significance of differences in $\pi_N/\pi_S$ ratios observed between domains was assessed by resampling. For each domain, we resampled 10,000 times a region of the same length as the domain of interest at a given RLR from the entire coding region of the two other RLRs put end to end and alternatively ordered first or second. The $\pi_N/\pi_S$ value observed in a specific domain at one RLR was then compared with that obtained in the resampling along the coding regions of the other two RLRs. We then determined where each specific domain falls in the resampling in terms of $\pi_N/\pi_S$ and deduced the corresponding $P$-value. The domain division was done according to Baum and Garcia-Sastre (13). Details for the CARD1/CARD2 division were retrieved from the Uniprot database (84).

**SUPPLEMENTARY MATERIAL**

Supplementary Material is available at HMG online.

**ACKNOWLEDGEMENTS**

We would like to thank Katherine Siddle for critical reading of the manuscript.

**Conflict of Interest statement.** None declared.

**FUNDING**

This work has been supported by the Institut Pasteur, the ANR (ANR-08-MIEN-009-01), the Fondation pour la Recherche Médicale, the CNRS, Merck-Serono and a EPFL-Debiopharm Life Sciences Award to L.Q.-M.

**REFERENCES**


