

# Adaptive Evolution in TRIF Leads to Discordance between Human and Mouse Innate Immune Signaling

Edel M. Hyland<sup>\*,1,2</sup>, Andrew E. Webb<sup>1</sup>, Kathy F. Kennedy<sup>3</sup>, Z. Nevin Gerek Ince<sup>4</sup>, Christine E. Loscher<sup>3</sup> and Mary J. O'Connell<sup>\*,1,5</sup>

<sup>1</sup>Bioinformatics and Molecular Evolution Group, School of Biotechnology, Dublin City University, Glasnevin, Dublin 9, Ireland

<sup>2</sup>School of Biological Sciences, Queen's University Belfast, Belfast, United Kingdom

<sup>3</sup>Immunomodulation Group, School of Biotechnology, Dublin City University, Glasnevin, Dublin 9, Ireland

<sup>4</sup>Institute for Genomics and Evolutionary Medicine, Temple University, Philadelphia, Pennsylvania, USA

<sup>5</sup>Computational and Molecular Evolutionary Biology Group, School of Life Sciences, Faculty of Medicine and Health Sciences, University of Nottingham, Nottingham, United Kingdom

\*Corresponding authors: E-mails: e.hyland@qub.ac.uk; mary.o'connell@nottingham.ac.uk.

Accepted: 24 November 2021

## Abstract

The TIR domain-containing adapter inducing IFN- $\beta$  (TRIF) protein is an innate immune system protein that mediates the MyD88-independent toll-like receptor response pathway in mice and humans. Previously, we identified positive selection at seven distinct residues in mouse TRIF (mTRIF), as compared with human and other mammalian orthologs, thus predicting protein functional shift in mTRIF. We reconstructed TRIF for the most recent common ancestor of mouse and human, and mutated this at the seven sites to their extant mouse/human states. We overexpressed these TRIF mutants in immortalized human and mouse cell lines and monitored TRIF-dependent cytokine production and gene expression induction. We show that optimal TRIF function in human and mouse is dependent on the identity of the positively selected sites. These data provide us with molecular data relating observed differences in response between mouse and human MyD88-independent signaling in the innate immune system with protein functional change.

**Key words:** innate immune signaling, positive selection, ancestral gene resurrection.

## Significance

There are many similarities between human and mouse immune systems. However, there are a number of cases, for example, TRIF-dependent responses, where the human and mouse respond differently to infection. Previously, we proposed that species-specific positive selection in immune system genes may be a predictor of species-specific immune response (Webb et al. 2015). Here we perform immunological assays and demonstrate that sites under positive selection in the mouse *TRIF* (*mTRIF*) gene drive the known mouse-specific response. These findings show that the observed phenotypic/physiological differences between human and mouse in the MyD88-independent pathway response are driven by sites under positive selection in *mTRIF*.

## Introduction

There are numerous cases whereby the mouse model fails to fully mimic human physiology (Sango et al. 1995; Rangarajan and Weinberg 2003; Hatzipetros et al. 2014), including differences in the immune response (Sellers 2017). Understanding

the molecular underpinnings of such variation between human and mouse will improve experimental design and relevance of mouse models. Given that mouse and human lineages diverged from a common ancestor ~88 Ma (Hedges et al. 2006), their life histories and evolutionary

© The Author(s) 2021. Published by Oxford University Press on behalf of the Society for Molecular Biology and Evolution.

This is an Open Access article distributed under the terms of the Creative Commons Attribution License (<https://creativecommons.org/licenses/by/4.0/>), which permits unrestricted reuse, distribution, and reproduction in any medium, provided the original work is properly cited.

trajectories in the intervening  $\sim 180$  Myr have given rise to lineage-specific mutations. By identifying genes (and sites) under species-specific positive selection using comparative phylogenomics we can approximate functional divergence between human and mouse. Indeed we, and others have shown that sites under positive selection are involved in protein functional shift for a variety of proteins (Goriely et al. 2005; Levasseur et al. 2006; Loughran et al. 2012). Therefore, comparative phylogenomics can be used to systematically identify functionally discordant pathways between mouse and human, to assess the utility of mice more accurately as models for human biology and disease.

Variation in immune response between human and mouse is not unsurprising given that idiosyncratic diets, climates, and exposure to distinct pathogens shapes immune responses (Mestas and Hughes 2004; Kammerer et al. 2007; Barreiro et al. 2010; Vasseur and Quintana-Murci 2013). Indeed, genes involved in the innate immune response have been identified as some of the most rapidly evolving across a broad range of taxa (Obbard et al. 2009; McTaggart et al. 2012; Lazzaro and Clarke 2013; Behrman et al. 2018; Adrian et al. 2019), amplifying differences between species. Previously, we identified positive selection in 35 innate immune system genes in mouse (Webb et al. 2015) and our aim in this study is to determine the functional consequence(s) of positive selection in one of these genes. We chose to focus on the mouse TRIF (mTRIF) protein as previous studies demonstrated divergence in TRIF-dependent signaling between mouse and human (see below). Furthermore, the number of predicted residues in *mTRIF* that contribute to the signature of positive selection are amenable to experimentation, and the assays to monitor TRIF function are well established.

TRIF is an intracellular adapter protein that acts in the MyD88-independent toll-like receptor (TLR) pathway, signaling specifically through TLR3 and TLR4 (reviewed Ullah et al. 2016) (fig. 1A). TLR3 and TLR4 receptors are activated by viral (dsRNA) and bacterial (lipopolysaccharide [LPS]) signals respectively, promoting the antiviral and proinflammatory response. Specifically, TRIF is recruited to activated TLR3 and TLR4, and through its interaction with the signaling proteins RIP1 and TRAF6 leads to the downstream activation of IRF3 transcription factor. This induces the expression of type-1 interferons and pro-inflammatory cytokines such as RANTES, and numerous interleukins namely, IL6, IL8, IL12, and IL-1 $\beta$  (fig. 1A).

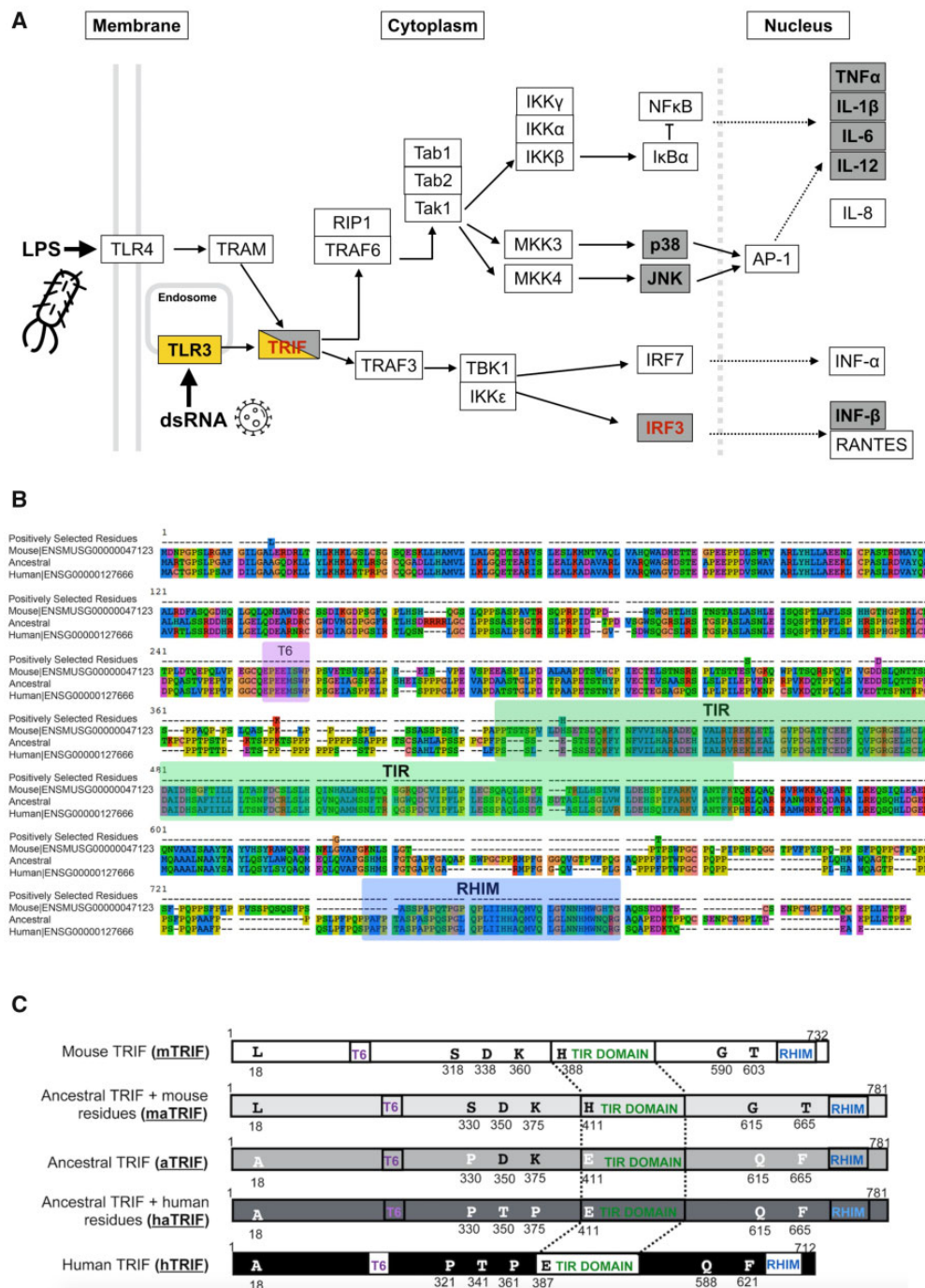
Documented inconsistencies exist in TLR3 and TLR4 signaling between mouse and human. First, the structure of the *TLR3* and *TLR4* genes themselves differ between mouse and human (Rehli 2002, Iwami et al. 2000), and well-documented differences exist in the regulation of TLR3 expression between both species (Heinz et al. 2003). Second, the loss of TLR3 function leads to discordant phenotypes in mice and humans. For example, TLR3-deficient mice show disruption in susceptibility to many viruses ranging from increased susceptibility to increased resistance phenotypes

depending on the virus, indicating a broad role for TLR3 signaling in protection against viral infections in mice (reviewed Zhang et al. 2013). In humans however, individuals born with deficiencies in TLR3 signaling show increased susceptibility to only a single virus, HSV-1 as TLR3-independent signaling contributes more substantially to antiviral immunity in humans (Zhang et al. 2013). Furthermore, differences exist in the regulation of TRIF-dependent cytokine production between mouse and human involving PKC- $\alpha$  (Langlet et al. 2010), and TLR3-specific activation leads to discordant responses between rodents and primates (Mitchell et al. 2014). Taken together, this evidence suggests that TRIF-dependent pathways and signaling in mouse and human have functionally diverged.

## Results

### The MyD88-Independent Signaling Pathway Shows Evidence of Divergence between Human and Mouse

In a previous study, we reported that the sequences of 37 innate immune response genes, including *mTRIF*, showed differences in selective pressure between human and mouse orthologs (Webb et al. 2015). Here, we extended this selective pressure analysis using CodeML (Yang 2007) to include 11 additional genes involved in TRIF-dependent signaling and increased our sequence data set from 21 to 43 mammalian genomes (supplementary tables S1 and S2, Supplementary Material online). Again, we detected multiple differences in selective pressures between mouse and human for 9 TRIF-related proteins (fig. 1A and table 1). We reconfirmed lineage-specific positive selection in *TRIF* in the mouse lineage (Webb et al. 2015), although there was not a complete overlap in the identity of the amino acids under selection between both studies (supplementary fig. S3A, Supplementary Material online). This presumably is due to the additional genomes present in our analysis resulting in changes in the multiple sequence alignment. For the *TRIF* alignment, we determined the confidence score to be 0.883, and  $<4\%$  of the columns in this alignment were deemed “unreliable.” We also detected mouse lineage-specific positive selection in *IRF3*, the major downstream transcription factor in the MyD88-independent signaling cascade. Interestingly, none of the genes analyzed are predicted to be under positive selection in the human lineage. These results further suggest that orthologous genes in the TRIF-dependent immune response pathways have functionally diverged between human and mouse. Furthermore, we identified site-specific positive selection downstream in the same pathway in the following genes: p38, and the cytokines TNF $\alpha$ , IL-1B, IL-6, IL-12A, and Interferon  $\beta$ 1. Overall, these findings indicate a strong selective pressure for change of function in the MyD88-independent signaling cascade.



**Fig. 1.**—Signatures of adaptive evolution identified in *TRIF* and TRIF-dependent signaling pathways. (A) Schematic adapted from the Kyoto Encyclopedia of Genes and Genomes (KEGG) illustrating TRIF-dependent pathways in toll-like signaling in innate immunity. Selective pressure variation analysis was performed for both lineage-specific (model A) and site-specific (model 8) selection. Genes boxed in gray represent those that are predicted to have signatures of positive selection; gene names in bolded black contain residues that are predicted to be under site-specific selection, and genes in red are genes under selection in the mouse lineage only. Highlighted in yellow are *TLR3* and *TRIF* predicted to be under selection in mouse by our earlier analysis (Webb et al. 2015). (B) A portion of the PRANK amino acid alignment of TRIF homologues showing the human, mouse, and ancestral TRIF protein sequences. The positions of the positively selected sites are indicated, as well as the position (relative to human—O’Neill and Bowie 2007) of the three known functional TRIF domains: Traf6 binding motif (T6), Toll/interleukin-1 receptor (TIR) domain, and the receptor-interacting protein (RIP) homotypic interaction motif (RHIM). (C) Schematic illustrating the different protein sequences of TRIF used in this study. Residues highlighted in the mouse protein (mTRIF) are those predicted to be under positive selection. Their position relative to the three functional domains is indicated. The position and identity of the homologous residues in the hTRIF, the reconstructed ancestral TRIF (aTRIF), the humanized ancestral TRIF protein (haTRIF), and the murinized ancestral TRIF (maTRIF) are also shown.

**Table 1.**

Summary of CodeML-Positive Selection Analysis on TRIF-Related Immune Response Genes

Gene Name	Type of Positive Selection Predicted <sup>a</sup>	Sites <sup>b</sup>	P Value of Positively Selected Sites <sup>c</sup>
TRIF	Lineage specific, mouse	18, <u>218</u> , <u>270</u> , <u>318</u> , <u>338</u> , <u>360</u> , <u>375</u> , 387, 388, 512, 603	11 > 0.50, <u>3</u> > 0.90, 0 > 0.99
	Site specific	180, 332, 362, <u>369</u> , <u>406</u> , 424, 425, 465, <u>479</u> , 496	10 > 0.50, <u>4</u> > 0.90, 0 > 0.99
TLR3	Lineage specific, mouse	<u>266</u> , <u>297</u> , <u>603</u>	3 > 0.50, 0 > 0.90, 0 > 0.99
IRF3	Lineage specific, mouse	<u>96</u> , <u>140</u> , <u>152</u> , <u>229</u>	4 > 0.50, <u>2</u> > 0.90, 0 > 0.99
p38-β/MAPK11	Site specific	<b>16</b>	1 > 0.50, <u>1</u> > 0.90, <b>1</b> > 0.99
TNFα	Site specific	16, <u>71</u> , <u>74</u> , <u>75</u> , <u>163</u> , 164, 173, 184, <b>186</b> , 194, 249	11 > 0.50, <u>5</u> > 0.90, <b>2</b> > 0.99
IL-1B	Site specific	109, 143, 245, 246, 247, 252, 281, 284, 285, 289, 291, 308, <u>332</u> , 360, 400, 445, 456	17 > 0.50, <u>2</u> > 0.90, 0 > 0.99
IL-6	Site specific	<u>102</u> , 128, 149, 156, 164, 166, 173, 176, 187, 231, 250, <u>254</u> , 255, 258, 259, 262, 273, 286, 287, 288, <u>289</u> , <u>292</u> , 298, 299, 320, 324, 328, 333, 335	29 > 0.50, <u>4</u> > 0.90, 0 > 0.99
IL-12A	Site specific	21, 110, <u>152</u> , <u>178</u> , 190, 277, 281, 327	8 > 0.50, <u>2</u> > 0.90, 0 > 0.99
INF β1	Site specific	25, 26, 28, 36, 37, <u>43</u> , 48, 50, <u>52</u> , <u>78</u> , 82, <u>90</u> , <u>111</u> , 114, 119, 122, 135, <u>137</u> , <u>138</u> , 140, 145, 153, 156, 157, 173, <b>188</b> , 191	27 > 0.50, <u>8</u> > 0.90, <b>1</b> > 0.99

<sup>a</sup>For site-specific selection, results are based on model M8 alone (reporting BEB sites).

<sup>b</sup>For genes under site-specific selection, site numbers refer to position in the multiple sequence alignment used for CodeML analysis. For lineage-specific selection, site number (in red) refers to the amino acid position in lineage/species where selection has been detected.

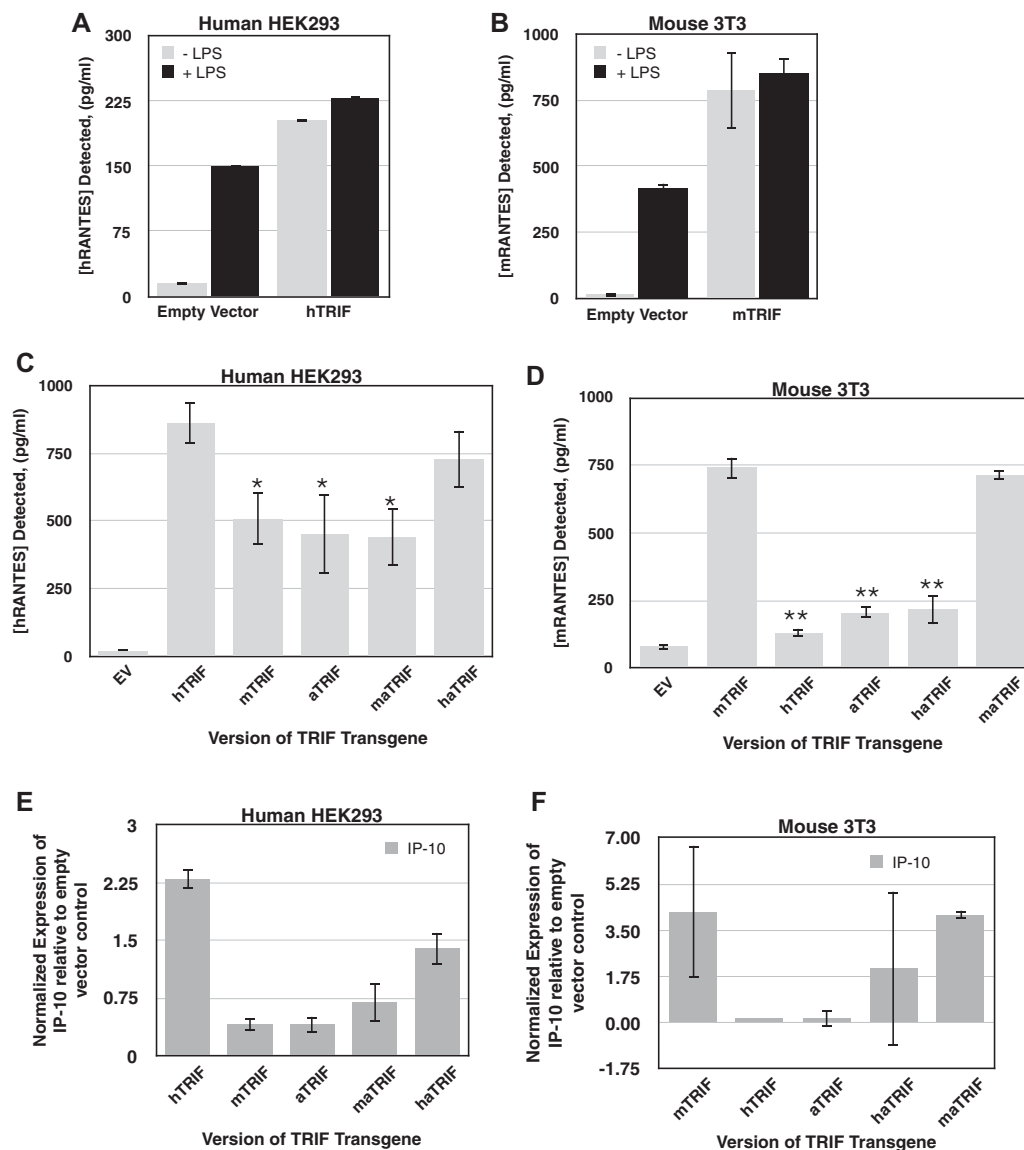
<sup>c</sup>The associated posterior probability is indicated through formatting: underlined >0.90, in bold >0.99, and no formatting implies 0.50 < PP < 0.90.

### A Resurrected Mouse–Human Ancestral TRIF Protein Shows Species-Specific Functionality

Mammalian *TRIF* contains three annotated domains: highly conserved TIR and TRAF6 binding domains, and a less well-conserved RHIM (receptor-interacting-protein homotypic interaction motif) (fig. 1B). Of the 10 sites predicted to be under selection in *mTRIF*, we chose the seven with the highest confidence to focus on (see supplementary fig. S3B, Supplementary Material online for details). Of those seven amino acid residues under positive selection, only one is located within an annotated (TIR) domain (fig. 1B). To understand the evolutionary relevance of the positively selected mouse amino acids, we employed ancestral gene reconstruction (AGR) (for review, see Thornton 2004). Using FastML (Ashkenazy et al. 2012), we reconstructed the most recent common ancestor of mouse–human TRIF (*aTRIF*) (as in Bickelmann et al. 2015, Voordeckers et al. 2012, Randall et al. 2016, Castro-Fernandez et al. 2017, and Furukawa et al. 2020), from an alignment of 10 rodent and mammalian genomes. This alignment obtained a Guidance2 confidence score of 0.947, with unreliable columns making up <1% of the total alignment. This predicted *aTRIF* sequence was 781 amino acids in length, with a median posterior probability score of 0.934 for individual codons (supplementary fig. S4, Supplementary Material online), giving us sufficient

confidence in its composition. Interestingly, within the known annotated domains, the predicted *aTRIF* protein is more similar in sequence to the extant human TRIF (*mTRIF*) than *mTRIF* (supplementary table S4, Supplementary Material online). Using gene synthesis, we resurrected this *aTRIF* gene sequence and cloned it into a mammalian overexpression plasmid under the control of the constitutive CMV promoter.

Initially we sought to verify the functionality of the *aTRIF* protein and transfected both human embryonic kidney (HEK) 293 TLR4+ cells, and murine fibroblast NIH 3T3 cells with our synthetic gene construct. Additionally, we used two control plasmids that expressed either *hTRIF* or *mTRIF* genes under the same regulation. Upon transfection of the cell lines with each TRIF construct, we used ELISA to monitor the production of the cytokine RANTES (downstream of MyD88-independent TLR4 signaling) as a proxy for TRIF function. The overexpression of *hTRIF* and *mTRIF* in their native context, HEK 293 cells and 3T3 cells respectively, induces the production of RANTES, bypassing the requirement for LPS, the canonical stimulus for TLR4 signaling (fig. 2A and B). However, not surprisingly, the expression of *hTRIF* and *mTRIF* constructs in their nonnative context, that is, 3T3 cells and HEK 293 cells respectively, significantly decreased the levels of RANTES produced compared with their native hosts, indicating reduced TLR4 signaling by these cross-species TRIF proteins. For the ancestral *aTRIF*



**FIG. 2.**—Amino acid residues under positive selection in *TRIF*, modulate species-specific function. Functional analysis of *TRIF* gene constructs. TRIF-induced RANTES production for human HEK 293 cells (A + C) and mouse 3T3 cells (B + D). Cell lines were transiently transfected with plasmids constitutively expressing each of the versions of the *TRIF* transgenes as indicated, or an empty vector (EV) control. In (A) and (B), cells were stimulated with LPS 12 h post-infection. At 24 h post transfection, supernatants were recovered and assayed for the presence of the RANTES cytokine by ELISA. Each graph represents data from three biological replicates assayed in triplicate from a single day's experiment. Similar trends were observed on different days. Error bars denote SD. \* $P < 0.05$  compared with hRANTES value in HEK cells. \*\* $P < 0.01$  compared with mRANTES value in 3T3 cells. Quantitative RT-PCR analysis as outlined in figure indicating the expression of IP-10 in human HEK 293 TLR4 cells (E) and in murine 3T3 cells (F) expressing the indicated TRIF constructs. Baseline of 0 is set to the expression of IP-10 in cells transfected with the empty vector control.

construct, RANTES production was stimulated to levels above those detected for the empty vector control for both cell types indicating that the aTRIF protein is functional in this assay. However, the signal obtained was greater in HEK293 cells suggesting that aTRIF functions better within the human cellular context (fig. 2C and D). As a second functional test of each TRIF construct, we assessed the expression of the chemokine IP-10 upon transfection of both HEK 293 and 3T3

cells using RT-qPCR. IP-10 gene expression occurs in response to LPS stimulation of TLR4 and is TRIF dependent (Fitzgerald et al. 2003b). Our expression data corroborate the ELISA results showing that in HEK293 cells, TLR4 signaling is supported by hTRIF and to a much lesser extent mTRIF and aTRIF (fig. 2E). In murine NIH 3T3 cells robust IP-10 expression is detected for mTRIF, but negligible IP-10 RNA is present when hTRIF or aTRIF is present in the cells (fig. 2F).

### Amino Acid Sites Predicted to Be under Positive Selection Confer Optimal TRIF Function in Native Cellular Context

The resurrected *aTRIF* sequence provided us with the appropriate background to test the effects of species-specific mutations on functional diversification in *TRIF* (Harms and Thornton 2010). An alignment of the *aTRIF* sequence with *mTRIF* and *hTRIF* (fig. 1B) identifies amino acids in *aTRIF* that are homologous to the seven sites under positive selection in mouse. Most of these sites (five out of seven) reflect the specific residue found in extant *hTRIF*, which is expected given our prediction that positive selection has acted specifically on the mouse, and not human lineage. The fact that two of the mouse sites under selection, Asp350 and Lys360, are present in the predicted ancestral sequence, will be discussed below. We hypothesize that the seven sites under positive selection in *mTRIF* confer an altered function to mTRIF-signaling that will be absent in the aTRIF protein. Furthermore, by combining AGR with rational mutagenesis (Loughran et al. 2012; Baldwin et al. 2014), we can approximate the evolution of functional divergence in TRIF protein along the human and mouse lineages. We generated two “pseudo-evolved” versions of aTRIF: 1) mouse ancestral TRIF (*maTRIF*) and 2) human ancestral TRIF (*haTRIF*) where the homologous sites in *aTRIF* were mutated to reflect those found in either present day *mTRIF* or *hTRIF*, respectively (fig. 1C). Again, we used gene synthesis to resurrect these different versions of aTRIF and determined the effects of the mutations in the protein background in which the substitutions most probably occurred.

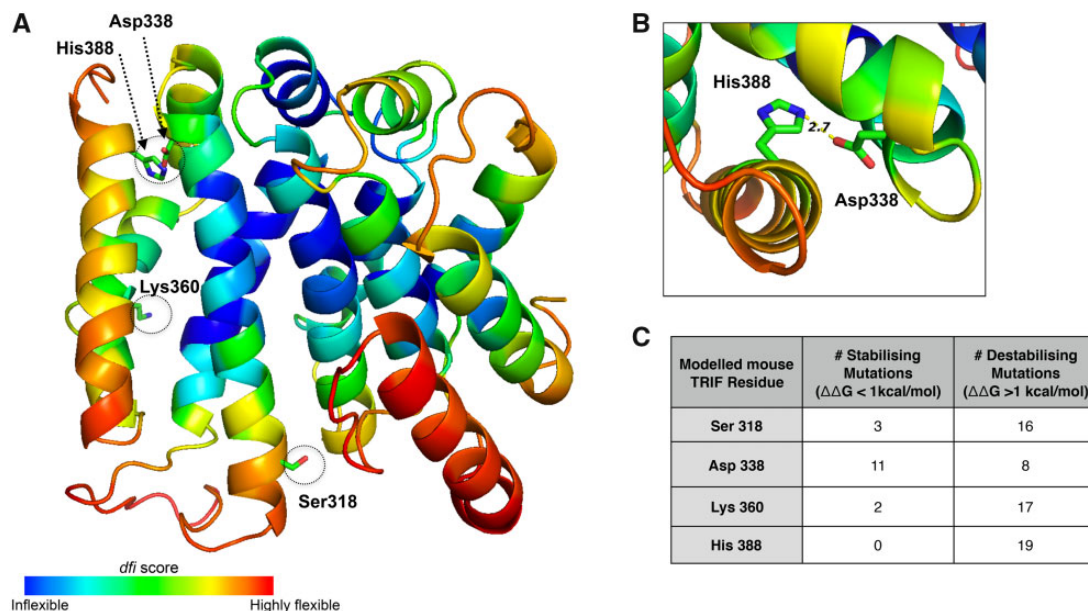
Analysis of the different versions of the ancestral TRIF protein indicated that in HEK 293 cells, *maTRIF* led to impaired signaling, whereas the *haTRIF* construct restored the ability of aTRIF to produce RANTES at a level almost comparable with *hTRIF* (fig. 2C). In contrast in murine 3T3 cells, the *haTRIF* protein significantly compromised the production of RANTES, whereas the *maTRIF* protein restored signaling through the pathway to wild-type mTRIF levels (fig. 2D). Again, quantitative RT-PCR analysis of IP-10 mRNA supported our ELISA data and showed that in HEK293 human cells, *haTRIF* significantly ( $P$  value > 0.05) increased IP-10 mRNA abundance over aTRIF, whereas the mouse-specific *maTRIF* did not (fig. 2E). Conversely, the *maTRIF* restored IP-10 mRNA expression levels to that of mTRIF levels in 3T3 cells, indicating that the mouse-specific substitutions were sufficient to improve the functionality of the ancestral *TRIF* gene to normal levels in this species. However, the *haTRIF* construct did not robustly alter IP-10 expression levels in the mouse cellular context (fig. 2F).

To determine whether expression differences between our TRIF constructs could account for the observed phenotypic differences, we determined the levels of TRIF mRNA using RT-qPCR with construct-specific primers (supplementary fig. S5, Supplementary Material online). Our data indicate strong expression (>15 times relative to empty vector control) of

each *TRIF* gene from the CMV promoter for all constructs in HEK 293 cells, except *maTRIF*. In 3T3 cells, although RNA abundance of each construct was lower than in HEK 293 cells, reproducible levels were detected for all, but the *maTRIF* construct. That said, there is clearly some variability of expression between the constructs in different cellular context. For 3T3 cells however, the pattern of expression variability is not consistent with the phenotypic results. For example, the *maTRIF* construct is the least expressed gene but shows optimal TRIF signaling in both the Rantes ELISA and IP-10 expression analysis. Similarly, in HEK cells the ancestral TRIF (*aTRIF*) sequence is the most highly expressed construct yet performs poorly in the functional analysis. Together, these observations are consistent with a lack of obvious correlation between TRIF mRNA abundance and TRIF function in our experimental setup, providing us with confidence in our conclusions about TRIF genotype/phenotype relationship.

### In Silico Analysis of TRIF Protein Structure Predicts That Amino Acids under Positive Selection in Mouse Are Structurally Important

Our experimental data lend support to our computational predictions about the altered function of mTRIF. We show that the predicted seven amino acid residues under positive selection in mTRIF; Leu18, Ser318, Asp338, Lys360, His388, Gly590, and Thr603, are required for optimal TRIF-dependent signaling in mouse cells. Clearly a detailed dissection of the contribution of each of the seven residues to mTRIF function is needed. However, the prediction that two of these mouse-specific residues (Asp338 and Lys360) existed in ancestral TRIF, suggests that they are either false positives of our CodeML analysis and are not under positive selection, or they reflect the limitations of AGR (Thornton 2004). In support of this, the *haTRIF* protein that lacks Asp338 and Lys360 functions equally poorly in mouse cells as the aTRIF protein which retains them, indicating that by themselves, Asp338 and Lys360 do not contribute to TRIF function in mouse in the context of the ancestral protein. Furthermore, both these residues lie in positions of the multiple sequence alignment with low confidence scores, 0.290 for Asp338, and 0.208 for Lys360. However, it is also possible that they are evidence of epistasis between the sites under positive selection (O’Maille et al. 2008; Phillips 2008), whereby Asp338 and Lys360 are required in the context of the mouse sequence to promote the altered function conferred by the remaining five sites, or indeed the identity of the residues has switched back and forth over evolution to account for amino acid substitutions at other positions. In support of epistasis, we used MODELLER (Fiser and Sali 2003) to predict the structure of a portion of mTRIF protein (Supplementary Material online) and identified a possible electrostatic interaction between Asp338 and a second site under positive selection, His388 (fig. 3).



**FIG. 3.**—In silico analysis highlights the structural importance of sites under positive selection in mTRIF. (A) The predicted structure of a portion of the mTRIF protein (residues Pro110 to Gln394) generated using MODELLER (Fiser and Sali 2003). Residues are highlighted based on their dynamic flexibility index (dfi) score, as per the indicated scale, and the position of four mouse residues under positive selection are indicated. (B) A close up of His 388 and Asp 338, two mTRIF residues under positive selection, in the predicted TRIF protein structure. The shortest distance between their two side chains is indicated (2.7 Å) and suggests a potential interaction between these two residues. Both (A) and (B) were generated using pymol (Delano 2002). (C) A summary of the in silico mutagenesis analysis done on mTRIF protein structure using dezyme software (<https://soft.dezyme.com>). The number of mutations that led to either a decrease (stabilizing) or increase (destabilizing) in the folding free energy ( $\Delta\Delta G$ ) of mTRIF is indicated.

We used our predicted mTRIF protein structure to assess a structure/function role for other mouse sites under positive selection and calculated the dynamic flexibility index (dfi) at each site. A low dfi score signifies high structural conservation and an intolerance to substitutions at that position. A positive correlation exists between dfi score and evolutionary rate (Nevin Gerek et al. 2013). Both Ser318 and His388 have relatively intermediate dfi scores (0.71 and 0.65, respectively) indicating moderate flexibility at these positions (fig. 3A). However, in silico mutagenesis, using dezyme PoPMuSiC analysis (Dehouck et al. 2009) (<https://soft.dezyme.com>) predicts that the majority of substitutions at these sites, including those found in hTRIF, are destabilizing (fig. 3C), highlighting their structural importance in mTRIF.

In human cells, we found that the impairment of signaling by the ancestral protein aTRIF is partially restored by the introduction of two specific residues, corresponding to Thr341 and Pro361 in the hTRIF protein, indicating that in the context of the ancestral protein the identity of these residues is important for TRIF signaling in human cells. Molecular dynamic simulations using a predicted hTRIF protein structure (supplementary fig. S6, Supplementary Material online) indicates that for Thr341, the majority of substitutions (14) are destabilizing, whereas replacing Pro361 with all amino acids except glycine, leads to a more stable hTRIF protein. These predicted contrasting effects of mutagenesis highlights the complexity in

understanding the molecular basis of positive selection and underscores the need for independent experimental analysis of each of these residues.

## Discussion

This study sought evidence that residues under positive selection in mTRIF confer species-specific functionality to TRIF signaling pathway. Our data support this prediction and demonstrate the effectiveness of combining AGR with rational mutagenesis to experimentally test this. That said, we fully acknowledge that there are documented limitations with AGR (Thornton 2004; Ekman et al. 2008) not least that the ancestral TRIF sequence presented here is an approximation. To this end, we report high confidence (mean posterior probability per codon = 0.934) in our predicted ancestral sequence, as well as in the identity of the homologous sites under selection. However, we acknowledge that a more robust strategy might be to resurrect multiple versions of aTRIF containing alternative states at the most ambiguously reconstructed positions. If these different versions of aTRIF behaved the same, it would eliminate doubt regarding ancestral state inference and provide a level of robustness to our experimental conclusions.

Furthermore, we acknowledge that placing an ancient protein inside a modern cell and assaying function holds many caveats, not least the possible loss of interaction(s) with vital

proteins, and therefore the decreased activity of aTRIF in human and mouse cells is not unexpected. To date, TRIF protein has three known interacting partners: TIR domains of TLRs, RIP, and Traf6. The TRIF sequences mediating these interactions are highly conserved; in fact, previous reports show that the TRIF TIR domain is under purifying selection (Fornarino et al. 2011). As such, the ancestral TRIF protein maintains a high level (79–100%) of sequence similarity with the extant mouse and human proteins at these interaction sites (supplementary table S4, Supplementary Material online). Therefore, it is possible that the known interactions of TRIF are sufficiently supported by the ancient protein in the modern cell. Regardless, our results indicate that despite these limitations to AGR, the activity of aTRIF was recovered in mouse cells by reintroducing the seven amino acids under positive selection found in present day mTRIF.

Given that six of the seven sites under positive selection in *mTRIF* are located in un-annotated regions, their specific (and individual) mechanistic contribution(s) to TRIF function will require further analyses. It is possible that these residues influence TRIFs known functions, for example by modifying its interaction with pathway components. However, given that positive selection is thought to result in protein functional shift (Loughran et al. 2012; Baldwin et al. 2014) it is feasible that all/some of these sites in TRIF mediate a novel interaction/function in mouse. To this end, we acknowledge that our study falls short of uncovering what this might be, and that substantially more in-depth mechanistic analyses need to be undertaken. That said, our experiments have identified a small set of residues in the ancestral TRIF protein that by changing them to be mouse specific or human specific is sufficient to recapitulate TRIF signaling activity within mouse and human cells respectively.

## Conclusion

To conclude, there are many similarities between human and mouse immune systems. However, there are cases where the human and mouse respond differently to infection, for example, TRIF-dependent responses. Previously, we proposed that species-specific positive selection in immune system genes may be a predictor of species-specific immune response (Webb et al. 2015). Here we performed immunological assays and demonstrate that sites under positive selection in the *mTRIF* gene drive the known mouse-specific response. These findings show that the observed phenotypic/physiological differences between human and mouse in the MyD88-independent pathway response are driven by sites under positive selection in *mTRIF*. Alongside the evolutionary implications, this work contributes to efforts in humanizing mouse responses when modeling disease and paves the way for future work on uncovering the species-specific function(s) of mTRIF.

## Materials and Methods

### Selective Pressure Analysis

Selective pressure analyses were performed using codeML from the PAML software package (v4.4) (Yang 2007). Briefly, CodeML examines nested codon-based models of evolution in a maximum-likelihood framework to determine the ratio of nonsynonymous substitutions per nonsynonymous site (Dn) to synonymous substitutions per synonymous site (Ds) ( $\omega = Dn/Ds$ ). We employed both site-specific and branch-specific models to test for positive selection at sites throughout our data set as well as in mouse and human lineages, respectively (Yang 2007). If positive selection is inferred, the posterior probability of the positively selected site is estimated using two calculations: Naïve empirical Bayes (NEB) or Bayes Empirical Bayes (BEB) (Yang 2007). If both BEB and NEB are predicted, we utilized the BEB results as they have been reported to be more robust (Yang 2007). We analyzed 12 genes in the TRIF-dependent immune response pathway (supplementary table S1, Supplementary Material online). Single gene orthologs were identified from the transcripts of 43 high quality genomes (supplementary table S2, Supplementary Material online) using the ENSEMBL Biomart database (<https://www.ensembl.org/biomart>). This analysis was therefore more in depth than our previous one which utilized up to 21 species (Webb et al. 2015). Alignments for CodeML analysis contained between 28 and 37 species depending on what was available on ENSEMBL (supplementary table S1, and Supplementary Material online). Alignment files can be found in the Supplementary Material online, and the TRIF alignment is shown in supplementary figure S1, Supplementary Material online. Alignment confidence was assessed using Guidance2 (Penn et al. 2010) and confidence scores obtained. Any positively selected sites inferred using CodeML under a given model were mapped back to the species-specific Uniprot protein sequence using software described in Webb et al. (2017).

### Ancestral TRIF Gene Resurrection

The most probable primate-rodent ancestral state for all sites in TRIF was determined using the FastML web-server (Ashkenazy et al. 2012), employing maximum-likelihood methods and the Yang codon model of protein evolution (Yang and Nielsen 2002). Briefly, homologous TRIF sequences from the primates, Human (ENSG00000127666), Marmoset (ENSCJAG00000017459), Gorilla (ENSGGOG00000006675), Macaque (ENSMUG00000014163), Gibbon (ENSNLEG00000013388), Bushbaby (ENSOGAG00000024483), Chimpanzee (ENSPTRG00000010322), and Orangutan (ENSPYGG00000009418) as well as the rodents Mouse (ENSMUSG00000047123), and Guinea Pig (ENSCPOG00000006486) were aligned using PRANK (Loytynoja and Goldman 2008). Initial attempts at AGR included larger sequence data sets, but



specific orthologs were removed whose TRIF sequence had evolved under different mechanisms. The confidence of our ten-species alignment for AGR was assessed using Guidance2 (Penn et al. 2010). A phylogenetic tree was generated for these species based on the topology as determined by Morgan et al. (2013). The species, Gibbon, was inserted manually onto this topology based on its positioning by the “Tree of life” online resource (Letunic and Bork 2011). The most likely marginally reconstructed sequence from the primate-rodent ancestral node was selected (fig. 1B).

The following versions of TRIF were synthesized by Invitrogen GeneART (<https://www.thermofisher.com/uk/en/home/life-science/cloning/gene-synthesis/geneart-gene-synthesis.html#>): 1) mTRIF, 2) ancestral TRIF (aTRIF), 3) ancestral TRIF whereby the amino acid residues predicted to be under selection were mutated to those residues present in either the extant mouse (maTRIF), or 4) the extant human (haTRIF). The mutagenesis strategy is detailed in figure 1C. The hTRIF mammalian expression construct, based on the pCMV2-FLAG vector, is described previously (Fitzgerald et al. 2003a). The synthetic *TRIF* genes were cloned into pCMV2-FLAG utilizing the same strategy as was used to engineer the hTRIF expression construct. Briefly, each gene was amplified from the original GeneArt vector using the primers oEMH1013 (5'-CTCACTATAGGGCGAATTGG-3') and oEMH1014 (5'-GAAGGCCCATGAGGCCAGTTAATT-3'), subsequently digested with BamHI and Sall and cloned into the pCMV2-FLAG vector generating the plasmids pEMH1001 (mTRIF insert), pEMH1002 (aTRIF insert), pEMH1003 (maTRIF insert), and pEMH1004 (haTRIF insert). Two independent clones were sequence verified for each construct and used for subsequent phenotypic analysis.

### In Silico Protein Modeling and Mutational Analysis

To predict the structure of human and mTRIF protein, we first identified template proteins with high sequence similarity to the TRIF protein sequence, and whose structures are available. Then we constructed a theoretical structural model of the TRIF proteins from its amino acid sequence through homology modeling. To generate reliable models, we used MODELLER (Fiser and Sali 2003) software and the ModBase database. The method relies on an input sequence alignment between the amino acid sequence to be modeled and a template protein whose structure has been solved. For both aTRIF and mTRIF, the sequence identity of the mouse sequence is 11%, and the template pdb used was 4k6j. We calculated dynamic flexibility index (dfi) scores of each residue in these homology modeled structures using our in-house program. To probe the effect of mutating the sites under selection in both the mouse and human predicted structures, we uploaded the corresponding pdb files into dezyme.com website and ran the systematic PoPMuSiC analysis which predicts changes in protein stability

by all possible mutations at a given site using statistical potentials and neural networks (Dehouck et al. 2009).

### Cell Culturing and Transfection

The HEK 293 immortalized HEK cell line, expressing TLR4 was a gift from Prof Paul Moynagh at National University of Ireland, Maynooth. These cells were maintained in Dulbecco's modified eagle's medium (DMEM) supplemented with 10% (v/v) fetal bovine serum (FBS), and 2% (v/v) penicillin–streptomycin. Cells were incubated at 37 °C at in a CO<sub>2</sub> incubator, and split every 3 days for a maximum of 20 passages.

The murine embryonic 3T3 cells were a gift from Dr Tomas Moore's lab at University College Cork, and maintained in DMEM supplemented with 10% (v/v) FBS, 2 mM L-glutamine, 100 U/ml penicillin–streptomycin. Cells were incubated at 37 °C at in a CO<sub>2</sub> incubator and split when they reached 70–80% confluency, after approximately 3 days, for a maximum of 20 passages. Both cell lines were transfected with plasmid DNA using lipofectamine 2000 (ThermoFisher scientific, 1668027). Briefly, 4 E + 04 of either HEK 293-TLR4 cells, or murine 3T3 cells were seeded in a 96-well plate and allowed to grow overnight to reach a 75–85% confluency. Approximately 50 ng of purified plasmid DNA was incubated with lipofectamine at a ratio of 1:4 (DNA [μg]: lipofectamine [μl]) for 20 min at room temperature and added to the cells. Cells were incubated overnight at 37 °C at in a CO<sub>2</sub> incubator and analyzed the following day.

### RANTES ELISA Assay

Transfected HEK 293-TLR4 and murine 3T3 cells were analyzed for the expression and secretion of the cytokine RANTES. At the indicated time post-transfection, the supernatants from both HEK 293 and 3T3 cells were collected and 50 μl (undiluted) was analyzed using the DuoSet ELISA human RANTES (R&D systems DY278) and murine RANTES (R&D systems DY478) respectively in a 96-well format. As a control for RANTES production, mock-transfected cells were stimulated with 100 ng/ml LPS (Enzo life science, ALX-581-007-L002) for 10 h prior to ELISA analysis.

### Quantitative RT-PCR

To verify the expression of each transgene, total RNA was extracted from transfected cells and quantitative RT-PCR was undertaken. For RNA extraction, approximately 1 E + 06 both HEK 293-TLR4 and murine 3T3 cells were seeded in 6-well plates, transfected using lipofectamine as before, and at the indicated timepoints cells were harvested. Total RNA was prepared using the Macherey-Nagel RNA isolation kit (11922402), and cDNA was prepared using Biotin Tetro-cDNA synthesis kit (BIO-65042). Transcript abundance was determined using Applied Biosystem Fast SYBR green reagent (4385616). **Supplementary table S3, Supplementary Material**

online contains a list of the primer sequences employed in this study.

## Supplementary Material

Supplementary data are available at *Genome Biology and Evolution* online.

## Acknowledgements

This work was funded by Science Foundation Ireland Research Frontiers Programme grant (EOB2673) awarded to M.J.O'C. M.J.O'C. would also like to acknowledge the Fulbright commission for support. The authors wish to acknowledge the Irish Centre for High-End Computing (ICHEC) for the provision of computational facilities and support.

## Author Contributions

M.J.O'C. conceived of the study and designed all computational analyses. MJO'C, EMH and AEW carried out the computational analyses. M.J.O'C., E.M.H., and C.E.L. designed the immunological assays and E.M.H. carried out wet bench experiments with assistance from K.F.K. M.J.O'C., E.M.H., and N.G.I. carried out the structural modeling. E.M.H. and M.J.O'C. drafted the manuscript, all authors contributed to revising, reading, and critiquing the manuscript.

## Data Availability

The data underlying this article are available in the public domain ([www.ensembl.org](http://www.ensembl.org)) and can be accessed with the accession numbers provided in the text and supplementary files.

## Literature Cited

- Adrian J, Bonsignore P, Hammer S, Frickey T, Hauck C. 2019. Adaptation to host-specific bacterial pathogens drives rapid evolution of a human innate immune receptor. *Curr Biol*. 29(4):616–630.
- Ashkenazy H, et al. 2012. FastML: a web server for probabilistic reconstruction of ancestral sequences. *Nucleic Acids Res*. 40(Web Server issue):W580–E584.
- Baldwin MW, et al. 2014. Sensory biology. Evolution of sweet taste perception in hummingbirds by transformation of the ancestral umami receptor. *Science* 345(6199):929–933.
- Barreiro LB, Marioni JC, Blekhan R, Stephens M, Gilad Y. 2010. Functional comparison of innate immune signaling pathways in primates. *PLoS Genet*. 6(12):e1001249.
- Behrman EL, et al. 2018. Rapid seasonal evolution in innate immunity of wild *Drosophila melanogaster*. *Proc Biol Sci*. 285:1870.
- Bickelmann C, et al. 2015. The molecular origin and evolution of dim-light vision in mammals. *Evolution* 69(11):2995–3003.
- Castro-Fernandez V, et al. 2017. Reconstructed ancestral enzymes reveal that negative selection drove the evolution of substrate specificity in ADP-dependent kinases. *J Biol Chem*. 292(38):15598–15610.
- Dehouck Y, et al. 2009. Fast and accurate predictions of protein stability changes upon mutations using statistical potentials and neural networks: PoPMuSiC-2.0. *Bioinformatics* 25(19):2537–2543.
- Delano WL. 2002. The PyMOL molecular graphics system. Version 1.2r3pre. Palo Alto (CA): Schrödinger, LLC.
- Ekman S, Andersen HL, Wedin M. 2008. The limitations of ancestral state reconstruction and the evolution of the ascus in the Lecanorales (lichenized Ascomycota). *Syst Biol*. 57(1):141–156.
- Fiser A, Sali A. 2003. Modeller: generation and refinement of homology-based protein structure models. *Methods Enzymol*. 374:461–491.
- Fitzgerald KA, et al. 2003a. IKKepsilon and TBK1 are essential components of the IRF3 signaling pathway. *Nat Immunol*. 4(5):491–496.
- Fitzgerald KA, et al. 2003b. LPS-TLR4 signaling to IRF-3/7 and NF-kappaB involves the toll adapters TRAM and TRIF. *J Exp Med*. 198(7):1043–1055.
- Fornarino S, et al. 2011. Evolution of the TIR domain-containing adaptors in humans: swinging between constraint and adaptation. *Mol Biol Evol*. 28(11):3087–3097.
- Furukawa R, Toma W, Yamazaki K, Akanuma S. 2020. Ancestral sequence reconstruction produces thermally stable enzymes with mesophilic enzyme-like catalytic properties. *Sci Rep*. 10(1):15493.
- Goriely A, McVean GA, et al. 2005. Gain-of-function amino acid substitutions drive positive selection of FGFR2 mutations in human spermatogonia. *Proc Natl Acad Sci USA*. 102(17):6051–6056.
- Harms MJ, Thornton JW. 2010. Analyzing protein structure and function using ancestral gene reconstruction. *Curr Opin Struct Biol*. 20(3):360–366.
- Hatzipetros T, et al. 2014. C57BL/6J congenic Prp-TDP43A315T mice develop progressive neurodegeneration in the myenteric plexus of the colon without exhibiting key features of ALS. *Brain Res*. 1584:59–72.
- Hedges SB, Dudley J, Kumar S. 2006. TimeTree: a public knowledge-base of divergence times among organisms. *Bioinformatics* 22(23):2971–2972.
- Heinz S, et al. 2003. Species-specific regulation of Toll-like receptor 3 genes in men and mice. *J Biol Chem*. 278(24):21502–21509.
- Iwami KI, et al. 2000. Cutting edge: naturally occurring soluble form of mouse Toll-like receptor 4 inhibits lipopolysaccharide signaling. *J Immunol*. 165(12):6682–6686.
- Kammerer R, Popp T, Härtle S, Singer BB, Zimmermann W. 2007. Species-specific evolution of immune receptor tyrosine based activation motif-containing CEACAM1-related immune receptors in the dog. *BMC Evol Biol*. 7:196.
- Langlet C, et al. 2010. PKC-alpha controls MYD88-dependent TLR/IL-1R signaling and cytokine production in mouse and human dendritic cells. *Eur J Immunol*. 40(2):505–515.
- Lazzaro BP, Clarke A. 2013. Chapter: "Rapid evolution of innate immune response genes". In: Singh RS, Xu J, Kulathinal RJ, editors. *Rapidly evolving genes and genetic systems*. 1st ed. Oxford: Oxford University Press. p. 203–210.
- Letunic I, Bork P. 2011. Interactive tree of life v2: online annotation and display of phylogenetic trees made easy. *Nucleic Acids Res*. 39(Web Server issue):W475–W478.
- Levasseur A, et al. 2006. Tracking the connection between evolutionary and functional shifts using the fungal lipase/feruloyl esterase A family. *BMC Evol Biol*. 6:92.
- Loughran NB, et al. 2012. Functional consequence of positive selection revealed through rational mutagenesis of human myeloperoxidase. *Mol Biol Evol*. 29(8):2039–2046.
- Loytynoja A, Goldman N. 2008. Phylogeny-aware gap placement prevents errors in sequence alignment and evolutionary analysis. *Science* 320(5883):1632–1635.
- McTaggart SJ, Obbard DJ, Conlon C, Little TJ. 2012. Immune genes undergo more adaptive evolution than non-immune system genes in *Daphnia pulex*. *BMC Evol Biol*. 12:63.
- Mestas J, Hughes CCW. 2004. Of mice and not men: differences between mouse and human immunology. *J Immunol*. 172(5):2731–2737.

- Mitchell WM, Nicodemus CF, Carter WA, Horvath JC, Strayer DR. 2014. Discordant biological and toxicological species response to TLR3 activation. *Am J Pathol.* 184(4):1062–1072.
- Morgan CC, et al. 2013. Heterogeneous models place the root of the placental mammal phylogeny. *Mol Biol Evol.* 30(9):2145–2156.
- Nevin Gerek Z, Kumar S, Banu Ozkan S. 2013. Structural dynamics flexibility informs function and evolution at a proteome scale. *Evol Appl.* 6(3):423–433.
- Obbard DJ, Welch JJ, Kim K-W, Jiggins FM. 2009. Quantifying adaptive evolution in the *Drosophila* immune system. *PLoS Genet.* 5(10):e1000698.
- O'Maille PE, et al. 2008. Quantitative exploration of the catalytic landscape separating divergent plant sesquiterpene synthases. *Nat Chem Biol.* 4(10):617–623.
- O'Neill LA, Bowie AG. 2007. The family of five: TIR-domain-containing adaptors in Toll-like receptor signalling. *Nat Rev Immunol.* 7(5):353–364.
- Penn O, et al. 2010. GUIDANCE: a web server for assessing alignment confidence scores. *Nucleic Acids Res.* 38(Web Server issue):W23–W28.
- Phillips PC. 2008. Epistasis—the essential role of gene interactions in the structure and evolution of genetic systems. *Nat Rev Genet.* 9(11):855–867.
- Randall RN, Radford CE, Roof KA, Natarajan DK, Gaucher EA. 2016. An experimental phylogeny to benchmark ancestral sequence reconstruction. *Nat Commun.* 7:12847.
- Rangarajan A, Weinberg RA. 2003. Opinion: comparative biology of mouse versus human cells: modelling human cancer in mice. *Nat Rev Cancer.* 3(12):952–959.
- Rehli M. 2002. Of mice and men: species variations of Toll-like receptor expression. *Trends Immunol.* 23(8):375–378.
- Sellers RS. 2017. Translating mouse models: immune variation and efficacy testing. *Toxicol Pathol.* 45(1):134–145.
- Sango K, et al. 1995. Mouse models of Tay-Sachs and Sandhoff diseases differ in neurologic phenotype and ganglioside metabolism. *Nat Genet.* 11(2):170–176.
- Thornton JW. 2004. Resurrecting ancient genes: experimental analysis of extinct molecules. *Nat Rev Genet.* 5(5):366–375.
- Ullah MO, Sweet MJ, Mansell A, Kellie S, Kobe B. 2016. TRIF-dependent TLR signaling, its functions in host defense and inflammation, and its potential as a therapeutic target. *J Leukoc Biol.* 100(1):27–45.
- Vasseur E, Quintana-Murci L. 2013. The impact of natural selection on health and disease: uses of the population genetics approach in humans. *Evol Appl.* 6(4):596–607.
- Voordeckers K, et al. 2012. Reconstruction of ancestral metabolic enzymes reveals molecular mechanisms underlying evolutionary innovation through gene duplication. *PLoS Biol.* 10(12):e1001446.
- Webb AE, et al. 2015. Adaptive evolution as a predictor of species-specific innate immune response. *Mol Biol Evol.* 32(7):1717–1729.
- Webb AE, Walsh TA, O'Connell MJ. 2017. VESPA: very large-scale evolutionary and selective pressure analyses. *PeerJ Preprints.* 5:e1895v2.
- Yang Z. 2007. PAML 4: phylogenetic analysis by maximum likelihood. *Mol Biol Evol.* 24(8):1586–1591.
- Yang Z, Nielsen R. 2002. Codon-substitution models for detecting molecular adaptation at individual sites along specific lineages. *Mol Biol Evol.* 19(6):908–917.
- Zhang SY, et al. 2013. TLR3 immunity to infection in mice and humans. *Curr Opin Immunol.* 25(1):19–33.

Associate editor: Andrea Betancourt