

# Positive Evolutionary Selection On the RIG-I-Like Receptor Genes in Mammals

Ana Lemos de Matos<sup>1,2,3</sup>, Grant McFadden<sup>3</sup>, Pedro J. Esteves<sup>1,4\*</sup>

**1** CIBIO - Centro de Investigação em Biodiversidade e Recursos Genéticos/InBio Laboratório Associado, Universidade do Porto, Vairão, Portugal, **2** Departamento de Biologia, Faculdade de Ciências, Universidade do Porto, Porto, Portugal, **3** Department of Molecular Genetics and Microbiology, University of Florida, Gainesville, Florida, United States of America, **4** CESPU, Instituto de Investigação e Formação Avançada em Ciências e Tecnologias da Saúde, Gandra, Portugal

## Abstract

The mammalian RIG-I-like receptors, RIG-I, MDA5 and LGP2, are a family of DExD/H box RNA helicases responsible for the cytoplasmic detection of viral RNA. These receptors detect a variety of RNA viruses, or DNA viruses that express unusual RNA species, many of which are responsible for a great number of severe and lethal diseases. Host innate sentinel proteins involved in pathogen recognition must rapidly evolve in a dynamic arms race with pathogens, and thus are subjected to long-term positive selection pressures to avoid potential infections. Using six codon-based Maximum Likelihood methods, we were able to identify specific codons under positive selection in each of these three genes. The highest number of positively selected codons was detected in MDA5, but a great percentage of these codons were located outside of the currently defined protein domains for MDA5, which likely reflects the imposition of both functional and structural constraints. Additionally, our results support LGP2 as being the least prone to evolutionary change, since the lowest number of codons under selection was observed for this gene. On the other hand, the preponderance of positively selected codons for RIG-I were detected in known protein functional domains, suggesting that pressure has been imposed by the vast number of viruses that are recognized by this RNA helicase. Furthermore, the RIG-I repressor domain, the region responsible for recognizing and binding to its RNA substrates, exhibited the strongest evidence of selective pressures. Branch-site analyses were performed and several species branches on the three receptor gene trees showed evidence of episodic positive selection. In conclusion, by looking for evidence of positive evolutionary selection on mammalian RIG-I-like receptor genes, we propose that a multitude of viruses have crafted the receptors biological function in host defense, specifically for the RIG-I gene, contributing to the innate species-specific resistance/susceptibility to diverse viral pathogens.

**Citation:** Lemos de Matos A, McFadden G, Esteves PJ (2013) Positive Evolutionary Selection On the RIG-I-Like Receptor Genes in Mammals. PLoS ONE 8(11): e81864. doi:10.1371/journal.pone.0081864

**Editor:** Welkin E. Johnson, Boston College, United States of America

**Received:** June 14, 2013; **Accepted:** October 17, 2013; **Published:** November 27, 2013

**Copyright:** © 2013 Lemos de Matos et al. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

**Funding:** The Portuguese Foundation for Science and Technology supported the doctoral fellowship of ALM (SFRH/BD/48566/2008). Two research projects from the Portuguese Foundation for Science and Technology (PTDC/BIA-BEC/103158/2008 and FCT-ANR/BIA-BIC/0043/2012) also funded the study. "Genomics Applied To Genetic Resources" cofinanced by North Portugal Regional Operational Programme 2007/2013 (ON.2 – O Novo Norte), under the National Strategic Reference Framework (NSRF), through the European Regional Development Fund (ERDF) supported this work. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

**Competing interests:** The authors have declared that no competing interests exist.

\* E-mail: pjesteves@cibio.up.pt

## Introduction

The mammalian innate immune system operates as the first line of defense against microbial pathogen invasion [1-3]. This system recognizes infectious agents through a limited number of germline-encoded pattern-recognition receptors (PRRs) predominantly expressed on sentinel cells [2,4-6]. The host PRRs recognize and react with specific microbial components, known as pathogen-associated molecular patterns (PAMPs), which includes bacterial lipopolysaccharides, peptidoglycans, lipoteichoic acids and cell-wall lipoproteins, fungal  $\beta$ -glucan and viral nucleic acids [2,3,5,6]. The host PRRs exhibit distinct

expression patterns and following sensing of their cognate ligands, activate specific signaling pathways that lead to the expression of a variety of inducible self-defense genes involved in the collective inflammatory and immune responses [2]. To date, four different classes of PRRs have been identified, including the cell membrane-associated C-type lectin receptors (CLRs), the Toll-like receptors (TLRs) at the cell surface and at the membrane of intracellular vesicles (endosomes and lysosomes), and the cytoplasmic detection systems for intracellular PAMPs, namely the RIG-I-like receptors (RLRs) and the NOD-like receptors (NLRs) [3,6-8].

The RLRs are a family of DExD/H box RNA helicases critically and exclusively involved in the recognition of “nonself” RNA from actively replicating viruses in the cytoplasm of infected cells [9]. This receptor family consists of three members, the retinoic acid-inducible gene-I (RIG-I/DDX58), the melanoma differentiation associated factor 5 (MDA5/IFIH1) and the laboratory of genetics and physiology 2 (LGP2/DHX58) [10-14]. RIG-I and MDA5 share high sequence similarity and several structural features, including an N-terminal region consisting of tandem caspase activation and recruitment domains (CARDs), a central DExD/H box RNA helicase domain and a C-terminal domain (CTD). The two N-terminally located CARDs function as a signaling and interaction domain with other CARD-containing proteins [13,15,16]. The helicase domain retains catalytic activity to bind and unwind double stranded RNA (dsRNA) in an ATP hydrolysis-dependent manner [10,17]. The CTD plays a predominant role in high-affinity binding with dsRNA, encoding a repressor domain (RD) in RIG-I, but not in MDA5, which harbors an RD-like domain that does not participate in autoregulation [18]. These two RLRs detect a variety of both DNA and RNA viruses, particularly at early phase of viral replication, and signal the production of type I interferons (IFNs) and induction of an anti-viral response [10,17]. The third element of the RLR family, the LGP2 protein, lacks any CARDs but contains the helicase domain and the CTD also harbors a RD. The role of LGP2 in anti-viral immunity is less clear, but it has been suggested in different studies to serve both as a negative and a positive regulator of RIG-I and MDA5 signaling [10,19-21].

Despite the similarities between RIG-I and MDA5, they were shown to play different roles in anti-viral immunity by recognizing and protecting from specific classes of viruses [22]. RIG-I detects preferentially and most effectively short RNA sequences marked with 5'-triphosphate group (5'-ppp) and blunt-end of short double-stranded RNAs (dsRNAs) or single-stranded RNA (ssRNA) hairpins [23-27]. As a key sensor of ssRNA viruses, RIG-I is implicated in the response to Arenaviridae [28], Bunyaviridae [28], Filoviridae [28], Flaviviridae [18,29], Orthomyxoviridae [22,30], Paramyxoviridae [22,28,30,31] and Rhabdoviridae [22,23]. On the other hand, MDA5 is activated by high-molecular-weight poly(I:C) fragments [22,32], and also by long-duplex RNAs from the genomes of dsRNA viruses [30] or dsRNA replication intermediates of positive-strand viruses, such as Caliciviridae [33], Coronaviridae [34] and Picomaviridae [22,32]. Regardless the virus recognition specificity by RIG-I and MDA5, some viruses are redundantly sensed by both RLRs, such as the West Nile virus and the Dengue virus [30,35]. In addition to the extensively described recognition of RNA viruses by RIG-I and MDA5, a role in anti-viral signaling in response to several dsDNA viruses has also been observed. As an RNA sensor, RIG-I does not detect DNA directly; however, not only do many DNA viruses create dsRNA products by virtue of convergent transcriptional units derived from opposite strands, but also the host RNA polymerase III can mediate the transcription of cytoplasmic DNA templates (such as transfected poly dA:dT) into dsRNA containing 5'-triphosphate, which will activate RIG-I and trigger the production of type I IFN [36,37]. Both Epstein-

Barr virus and myxoma virus are detected by RIG-I, while vaccinia virus is sensed by MDA5 [38-40]. It is also likely that the precise RLRs utilized for the sensing of specific viruses also operate within cell-specific contexts as well.

Interaction between host and pathogen results in a dynamic arms race. Whenever pathogens develop strategies to overtake the host immune system, the host proteins involved in pathogen recognition have to respond by evolving to avoid or reduce potential infections. These dynamics result in host-pathogen adaptation and counter-adaptation, which in turns lead to the rapid co-evolution of both parties. Particularly for the host, this accelerated molecular evolution is often reflected in host defense genes that exhibit strong signatures of ongoing diversifying selection [41,42]. Because viruses are responsible for a great number of severe and lethal diseases, together with the important role that RLRs play in mammalian innate immune system, we expect that RIG-I, MDA5 and LGP2 genes may have been under intense selective pressures in all mammals. We have previously demonstrated that one other class of mammalian PRRs, the TLRs, exhibit striking evidence of positive genetic selection as a result of selective pressures exerted by pathogens [43]. Using six different codon-based Maximum Likelihood (ML) methods, we searched for evidence of long-term selective pressures in the three RLR genes present in the available sequenced mammalian genomes and, where possible, pinpoint positively selected residues that might be involved in the host-virus interactions that have shaped their rapid diversification. Specific lineages subject to episodic positive selection have also been identified in the three RLR genes by using two different branch-site models.

## Results

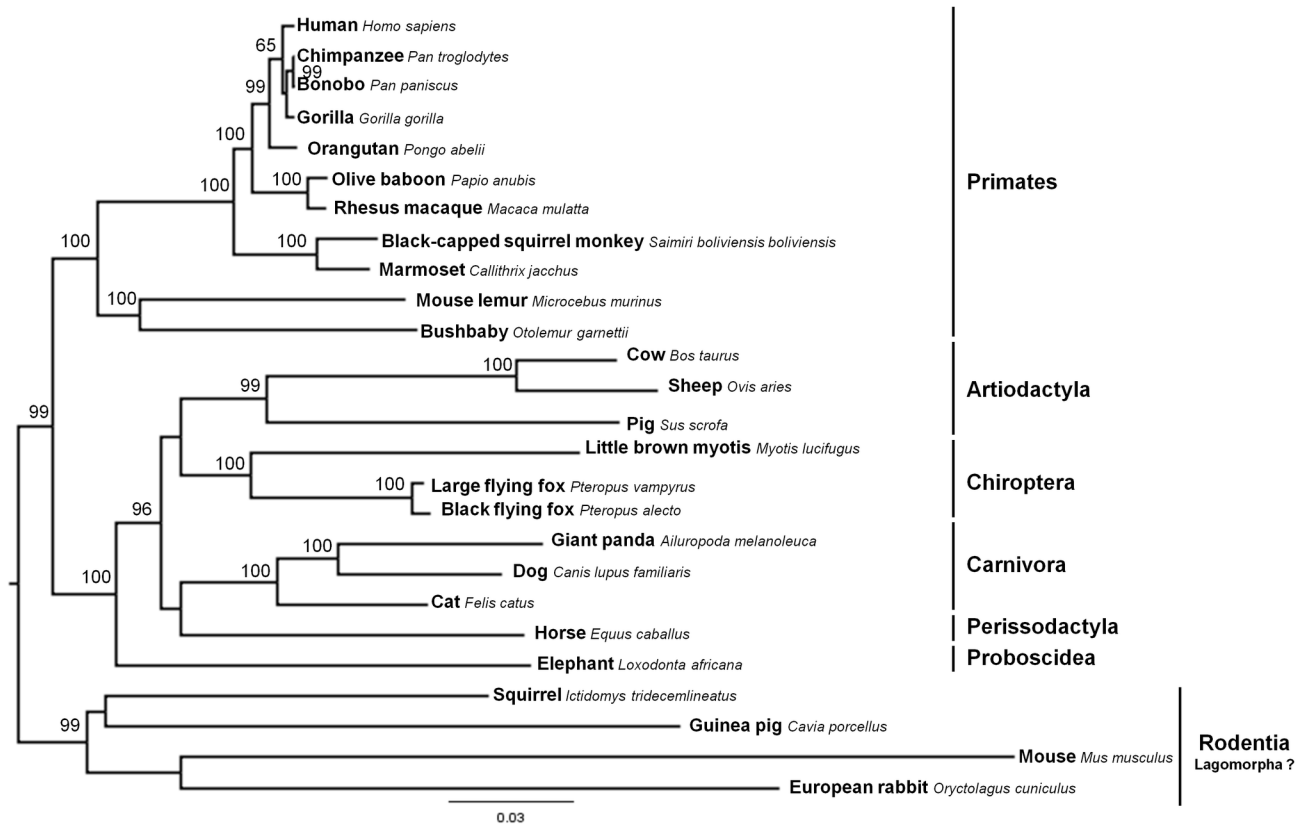
### Mammalian RIG-I-like receptor gene sequences

Publicly available mammalian RIG-I, MDA5 and LGP2 gene sequences were collected from Ensembl and NCBI databases (Table S1) for phylogenetic and selection analyses. The nucleotide coding sequences for each of the three RLR gene orthologous were aligned and are represented in Figure S1 (RIG-I alignment), Figure S2 (MDA5 alignment) and Figure S3 (LGP2 alignment). The translation into deduced protein sequences is also represented in Figure S4 (RIG-I alignment), Figure S5 (MDA5 alignment) and Figure S6 (LGP2 alignment).

The inherent limitations of using solely publicly available mammalian RLRs sequences should be highlighted, although several studies have used the same source of data for general conclusions about other genes in mammals [43-49]. The analyses performed ahead use only an individual representative of each included species and therefore, any drawn conclusions should be carefully considered.

### Phylogenetic reconstruction of mammalian RIG-I-like receptors

ML trees were reconstructed for RIG-I, MDA5 and LGP2 genes after looking for evidence of recombination using the software GARD [50,51] implemented in the Datamonkey web server [52,53].



**Figure 1. Maximum likelihood (ML) phylogenetic tree of RIG-I gene used for codon-based ML analysis.** The GTR+G nucleotide substitution model was employed in mammalian RIG-I gene phylogenetic reconstruction. Bootstrap values >50 are indicated on the branches.

doi: 10.1371/journal.pone.0081864.g001

In mammalian RIG-I phylogenetic reconstruction, the monophyly of six of a total of eight taxonomic orders was observed (Figure 1). However, an interesting fact was registered for the two remaining orders, Rodentia and Lagomorpha, when the European rabbit (order Lagomorpha) grouped with the rodent cluster. When looking carefully at the European rabbit RIG-I deduced protein sequence (Figure S4), a great number of conserved regions between this species and the other mammalian species was observed, with the exception of a region between codons 840 and 879. This 40 amino acid domain region was unique to the European rabbit RIG-I. We originally speculated that this difference might have been the result of a gene conversion event with adjacent genes. However, when we examined the genes that are chromosomally adjacent to European rabbit RIG-I (NDUFB6, TOPORS and FRP), no clear evidence of gene conversion was detected by the software GARD [50,51].

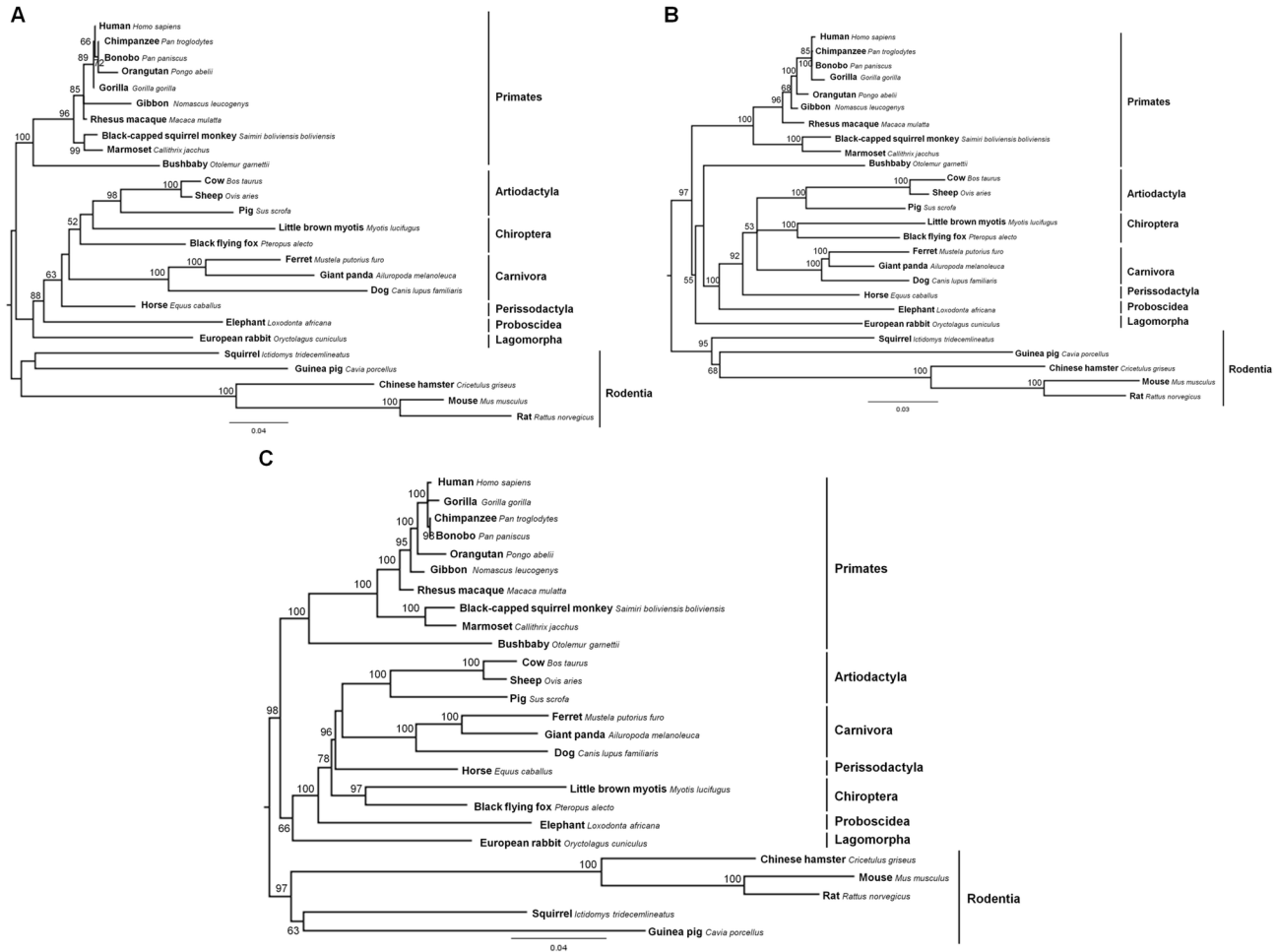
For the mammalian MDA5 gene sequences alignment, a significant recombination breakpoint was detected (nucleotide 903;  $p < 0.01$ ). Therefore, two ML trees were reconstructed for the resulting segments, one for the first 903 nucleotides (Figure 2A) and another ML tree for the remaining 2211 nucleotides (Figure 2B). The gene phylogeny was also reconstructed for

the whole alignment without testing recombination (Figure 2C) to compare its topology with the other two resulting trees. The monophyly of the eight taxonomic orders included in the MDA5 alignment was roughly recovered, with the clear exception of Chiroptera in Figure 2A and Primates in Figure 2B.

Regarding the LGP2 gene, no clear evidence of recombination was detected. The ML tree obtained (Figure 3) supported the monophyly of the nine mammalian orders collected for this gene.

#### Evidence of positive selection in the mammalian RIG-I-like receptors

All the molecular evolutionary analyses in this study were performed for both the complete nucleotide alignments (Figure S1, Figure S2 and Figure S3) and for a trimmed version of the same genes to remove alignment gaps. Figure S7 (RIG-I alignment), Figure S8 (MDA5 alignment) and Figure S9 (LGP2 alignment) correspond to the alignments where gaps present in all sequences, with the exception of one or two, have been removed, while gaps present in only one or two sequences were kept. We observed no significant differences in the results when using one or the other alignment for each gene (data not



**Figure 2. Maximum likelihood (ML) phylogenetic trees of mammalian MDA5 gene used for codon-based ML analysis.** When testing mammalian MDA5 alignment for recombination, one significant breakpoint was detected at nucleotide position 903. (A) A phylogenetic tree was reconstructed for the first 903 nucleotides under the nucleotide substitution model TIM3+G. (B) A second ML tree was inferred for the remaining 2211 nucleotides and under the nucleotide substitution model TIM3+I+G. (C) A tree was also reconstructed for MDA5 total alignment without recombination testing and under the nucleotide substitution model GTR+G. Bootstrap values >50 are indicated on the branches.

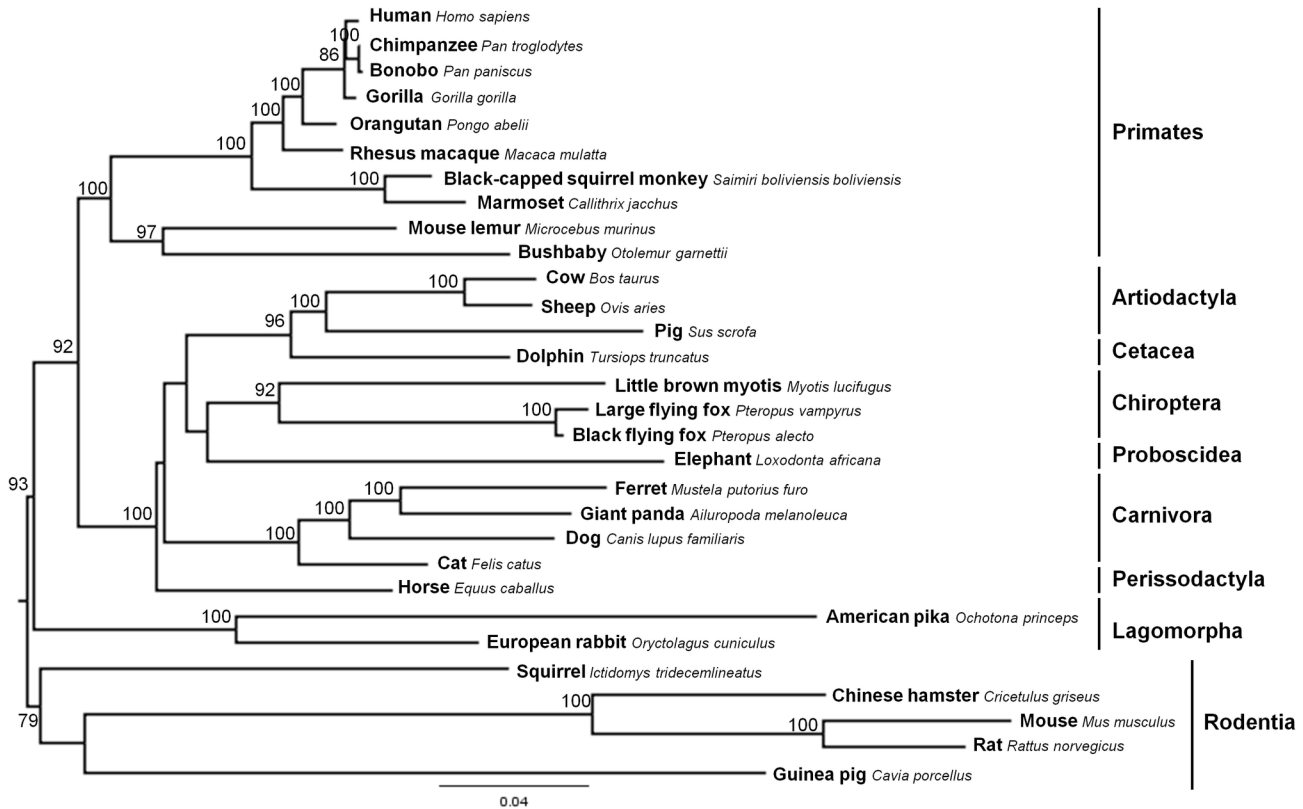
doi: 10.1371/journal.pone.0081864.g002

shown), but ultimately only the results from the trimmed version are presented here.

Evidence for positive selection on mammalian orthologous for RIG-I (Figure S7), MDA5 (Figure S8) and LGP2 (Figure S9) genes was detected using PAML package [54,55] site-specific models M1a versus M2a and M7 versus M8. These models test at the codon level whether a hypothesis that allows for positive selection (models M2a and M8) is a better fit to the data when compared to a null neutral hypothesis (models M1a and M7). Results on the likelihood ratio test (LRT) performed between the likelihood scores of the null neutral and alternative selection models for each gene is indicated in Table 1. Models which allow for positive selection (M2a and M8) gave a significantly better fit to the data for both RIG-I and LGP2 alignments, suggesting that at least some of the codons within

each set of orthologous gene sequences are subject to positive selection [56]. Since a recombination breakpoint was detected on the MDA5 alignment, each resulting segment (identified as 1<sup>st</sup> and 2<sup>nd</sup> segments) was tested individually for PAML package [54,55] site-specific models. Although the comparison between the null neutral site model M1a and the selection site model M2a did not allow for rejection of the null hypothesis of neutral selection, the comparison between the more powerful pair of site-specific models M7 (neutral) and M8 (selection) yielded significant LRTs (Table 1).

The PARRIS method [57] implemented in the Datamonkey web server [52,53] was also applied to each RLR trimmed gene alignment (Figure S7, Figure S8 and Figure S9) to look for evidence of positive selection, but no selective pressures were detected in any of the three genes (Table S2).



**Figure 3. Maximum likelihood (ML) phylogenetic tree of mammalian LGP2 gene used for codon-based ML.** The TPM2uf+I +G nucleotide substitution model was employed in mammalian LGP2 gene phylogenetic reconstruction. Bootstrap values >50 are indicated on the branches.

doi: 10.1371/journal.pone.0081864.g003

For each orthologous gene sequences alignment, the tree length parameter is indicated in Table 1. Higher values of tree length, i.e. the expected number of nucleotide substitutions per codon, correspond to higher sequence divergence [58,59]. The tree length values registered for the three genes fell into an intermediate and realistic level of sequence divergence which confers power to the codon models indicated by the LRT scores and to the Bayes empirical Bayes (BEB) approach for site-specific inference of positive selection [58,60].

Model M8 implemented in the PAML package [54,55] and Datamonkey web server [52,53] SLAC, FEL, REL, MEME and FUBAR methods [61-63] were used to detect specific codons under selection in the three RLR genes. Based on the methodology adopted by other authors and in previous studies [43,47,48,64], only codons identified by at least three of the six used methods are considered to be under positive selection (Table S3). Since the breadth of species included in each alignment is wide, by applying several methods to detect codons under positive selection and by overlapping the results, we should be decreasing the incidence of false positives. A total of sixteen codons for RIG-I (Figure 4), twenty for MDA5 (Figure 5) and ten for LGP2 (Figure 6) were identified as candidate codons under selective pressure. Regarding their location in each corresponding protein, the greatest number of

these codons are located in protein functional domains, more specifically, eleven out of the sixteen RIG-I codons (~ 69%), ten out of the twenty MDA5 codons (50%) and seven out of the ten LGP2 codons (70%). To estimate the percentage of positively selected codons in the three proteins, we used human deduced protein sequences as a reference. Human LGP2 exhibited 1.47% (10/678) of codons under positive selection. Higher values were obtained for human MDA5 and RIG-I, 1.95% (20/1025) and 1.73% (16/925) of codons under selective pressure, respectively.

To detect signatures of episodic positive selection in specific lineages of each RLR orthologous gene sequences alignment we performed two branch-site model analyses. These models allow the selective pressure indicated by the nonsynonymous to synonymous substitution rate ratio  $\omega$  ( $d_n/d_s$ ) to vary both across sites in the gene and across lineages on the tree [65]. Since no biological hypothesis existed to specify *a priori* branches to be examined for positive selection, the branch-site model A implemented in the PAML package [54,55,66] was applied to all species branches on each RLR gene phylogenetic tree. The LRT performed for each branch was significant for  $2\Delta\ln L > 3.84$  [55,66]. Our analyses suggest that nine species branches in RIG-I are under selective pressure (Table 2 and Figure 4B). Branch-site model A was applied to

**Table 1.** RIG-I-like receptors likelihood ratio tests (LRTs) for PAML M1a, M2a, M7 and M8 site models.

PAML site models	InL <sub>null</sub> <sup>a</sup>	InL <sub>alternative</sub> <sup>a</sup>	2ΔInL <sup>b</sup>	p-Value	Tree length
<b>RIG-I</b>					4.60
M1a (nearly neutral) vs. M2a (selection)	-21065.76	-21045.84	39.84	p<0.0001	
M7 (neutral, beta) vs. M8 (selection, beta & ω)	-21053.14	-21016.64	73.00	p<0.0001	
<b>MDA5_1stSegment</b>					6.03
M1a (nearly neutral) vs. M2a (selection)	-7981.98	-7981.98	0.00	n.s.	
M7 (neutral, beta) vs. M8 (selection, beta & ω)	-7970.70	-7966.35	8.70	P<0.02	
<b>MDA5_2ndSegment</b>					3.80
M1a (nearly neutral) vs. M2a (selection)	-14314.15	-14312.61	3.08	n.s.	
M7 (neutral, beta) vs. M8 (selection, beta & ω)	-14284.53	-14272.94	23.18	p<0.0001	
<b>LGP2</b>					6.38
M1a (nearly neutral) vs. M2a (selection)	-18838.31	-18830.88	14.86	p<0.001	
M7 (neutral, beta) vs. M8 (selection, beta & ω)	-18693.92	-18674.22	39.40	p<0.0001	

<sup>a</sup> InL: log-likelihood scores.

<sup>b</sup> 2ΔInL: likelihood ratio test (LRT) to detect positive selection.

n.s. – non-significant

doi: 10.1371/journal.pone.0081864.t001

the two MDA5 trees resultant from recombination testing and, for each tree, positive selection has operated only in two species branches (Table 3 and Figure 5B). For LGP2, a total of six species branches had significant LRTs corresponding to candidate lineages under selection (Table 4 and Figure 6B). Some of the species branches recognized by the branch-site model A were also identified by the branch-site REL method [67] (Table 5) available in the Datamonkey web server [52,53]. For both RIG-I and MDA5, two species branches were simultaneously identified by the two methods, consisting in dog (Calu) and European rabbit (Orcu) branches for RIG-I (Figure 4B) and giant panda (Aime) and Guinea pig (Capo) branches for MDA5 (Figure 5B). Only one LGP2 species branch, corresponding to the giant panda (Aime), was simultaneously identified by the branch-site model A and the branch-site REL method (Figure 6B).

## Discussion

In a human population genetics context, the first study on RLRs evolutionary history and selective footprints has been recently published [68]. Nevertheless, to the best of our knowledge, our study is the first that searches for selective pressure acting on mammalian orthologous of the three RLRs and, in fact, we provide strong evidence of positive selection as well as identify a significant number of codons under probable selective pressures for RIG-I, MDA5 and LGP2. Furthermore,

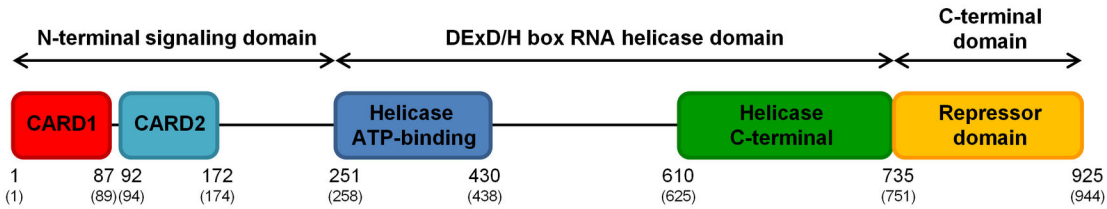
our results on the RIG-I RD in specific hosts suggest that certain viruses might be exerting long-term selective pressures on this gene.

TLRs adaptive evolution has been the most extensively characterized of all the PRRs in several animal groups, such as echinoderms [69], birds [70] and different mammals [43,64,71-74]. Studies on known viral-recognition TLRs (TLR3, 7, 8 and 9) of closely related animal groups, like birds [70], or within species, like humans and chimpanzees [64], demonstrated that this class of PRRs exhibits a background of strong purifying selection to keep their functional integrity, albeit in the birds study [70] significant instances of positive selection acting on a few amino acid sites were identified. Nevertheless, when different ML codon-based methods were applied to detect evidence of acting positive selection in broader groups where a great number of species are included, like primates [64] and mammals [43], most of the viral TLRs exhibited strong evidence of positive selection and specific codons with a high probability of being under selection were identified. Similarly, in our study the codon-based analyses strongly support that the three RLR genes, RIG-I, MDA5 and LGP2, have all been subject to long-term selective pressures during mammalian evolution. Also, we applied several methods that identified specific RLR codons with a high probability of being under selection, which may directly perturb downstream immune responses in a particular host infected by a viral pathogen.

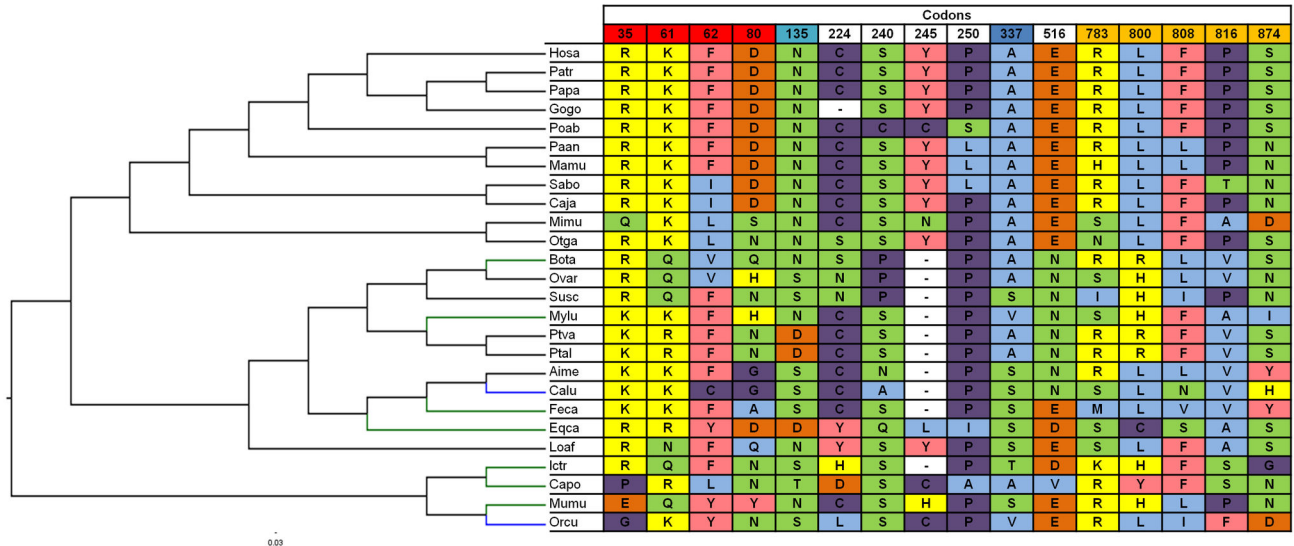
One of the major concerns when using large scale divergent species to infer positive selection acting on a set of orthologous genes and across lineages on the phylogenetic tree is the effect of saturation in synonymous substitutions, since they may saturate quickly as sequences diverge [75,76]. As codon models consider both synonymous and nonsynonymous substitutions, the saturation of the first could cloud the information provided by nonsynonymous substitutions. Nevertheless, the sequence divergence in our study, inferred through RLRs tree length values, fit into intermediate and realistic levels that should confer power to the LRT used to compare nested codon-models and robustness to the branch-site models, and to the BEB approach for codon-specific detection of positive selection [58-60]. Also, in this study the mammalian species collected for each of the three RLR genes were nearly the same, thus this host species spectrum should not influence the codon-based analyses and our observations when comparing the level of selective pressure between genes.

In our study, mammalian MDA5 showed the highest number and percentage of positively selected codons. Nonetheless, the percentage of MDA5 codons under selection located in the known protein functional domains was the lowest. This should reflect the imposition of functional and structural constraints in MDA5 defined domains. On the other hand, we observed that LGP2 is apparently less prone to evolutionary change with the lowest number and percentage of codons under selective pressures. For RIG-I, the greatest number of codons identified as candidates under selective pressures were located in known protein functional domains, which might reveal the pressure imposed by the great number of viruses recognized by this

A



B



C

	Codons															
	35	51	62	80	135	224	240	245	250	337	516	783	800	808	816	874
Hosa	R	K	F	D	N	C	S	Y	P	A	E	R	L	F	P	S
Patr	R	K	F	D	N	C	S	Y	P	A	E	R	L	F	P	S
Papa	R	K	F	D	N	C	S	Y	P	A	E	R	L	F	P	S
Gogo	R	K	F	D	N	-	S	Y	P	A	E	R	L	F	P	S
Poab	R	K	F	D	N	C	C	C	S	A	E	R	L	F	P	S
Paan	R	K	F	D	N	C	S	Y	L	A	E	R	L	L	P	N
Mamu	R	K	F	D	N	C	S	Y	L	A	E	H	L	L	P	N
Sabo	R	K	I	D	N	C	S	Y	L	A	E	R	L	F	T	N
Caja	R	K	I	D	N	C	S	Y	P	A	E	R	L	F	P	N
Mimu	Q	K	L	S	N	C	S	N	P	A	E	S	L	F	A	D
Otga	R	K	L	N	N	S	S	Y	P	A	E	N	L	F	P	S
Bota	R	Q	V	Q	N	S	P	-	P	A	N	R	L	L	V	S
Ovar	R	Q	V	H	S	N	P	-	P	A	N	S	H	L	V	N
Susc	R	Q	F	N	S	N	P	-	P	S	N	I	H	I	P	N
Mylu	K	K	F	H	N	C	S	-	P	V	N	S	H	F	A	I
Ptva	K	R	F	N	D	C	S	-	P	A	N	R	R	F	V	S
Ptal	K	R	F	N	D	C	S	-	P	A	N	R	R	F	V	S
Aime	K	K	F	G	S	C	N	-	P	S	N	R	L	L	V	Y
Calu	K	K	G	G	S	C	A	-	P	S	N	S	L	N	V	H
Feca	K	K	F	A	S	C	S	-	P	S	E	M	L	V	V	Y
Eqca	R	R	Y	D	D	Y	Q	L	I	S	D	S	C	S	A	S
Loaf	R	N	F	Q	N	Y	S	Y	P	S	E	S	L	F	A	S
Ictr	R	Q	F	N	S	H	S	-	P	T	D	K	H	F	S	G
Capo	P	R	L	N	T	D	S	C	A	A	V	R	Y	F	S	N
Mumu	E	Q	Y	Y	N	C	S	H	P	S	E	R	H	L	P	N
Orcu	G	K	Y	N	S	L	S	C	P	V	E	R	L	I	F	D

**Figure 4. Structural representation and identification of positively-selected branches and codons in mammalian RIG-I.** (A) Based on human protein structure, the key domains of RIG-I (<http://www.uniprot.org/uniprot/O95786>) and the corresponding boundaries are schematically represented. Also, the human domain boundaries while in the mammalian RIG-I deduced protein sequences alignment (Figure S4) are shown in brackets. (B) Cladogram of 26 mammalian RIG-I genes collected from Ensembl and NCBI databases. Branch-site analyses were performed to identify specific branches under episodic positive selection. Branches with statistically significant likelihood ratio tests (LRTs) when performing PAML branch-site model A (Table 2) are colored in green; branches simultaneously identified by PAML branch-site model A and Hyphy branch-site REL method (Table 5) are colored in blue. (C) Positively-selected codons are exhibited in the table and numbered according to the mammalian RIG-I deduced protein sequences alignment (Figure S4). Symbol “-” represents a deletion. Colors on the codon numbering row correspond to the RIG-I domain with the same color in the protein structural representation (A). The background colors on the identified sites match different amino acid properties: polar positive (yellow), polar negative (orange), polar neutral (green), non-polar neutral (purple), non-polar aliphatic (blue) and non-polar aromatic (pink). The used abbreviations correspond, by order of appearance, to the following species: Hosa – Human; Patr – Chimpanzee; Papa – Bonobo; Gogo – Gorilla; Poab – Orangutan; Paan – Olive baboon; Mamu – Rhesus macaque; Sabo – Black-capped squirrel monkey; Caja – Marmoset; Mimmu – Mouse lemur; Ptva – Large flying fox; Ptal – Black flying fox; Aime – Giant panda; Calu – Dog; Feca – Cat; Eqca – Horse; Loaf – Elephant; Ictr – Squirrel; Capo – Guinea pig; Mumu – Mouse; Orcu – European rabbit.

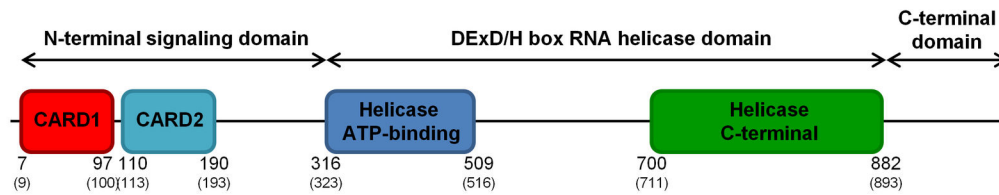
doi: 10.1371/journal.pone.0081864.g004

RLR [13,14]. Vasseur and colleagues [68] came to different conclusions in their study, once they were focused on intra-species (human populations) polymorphisms and on the comparison of nonsynonymous to synonymous rates ratio  $\omega$  ( $d_N/d_S$ ) between human and chimpanzee lineages for the three RLR genes. RIG-I exhibited a stronger evolutionary constraint [68], as attested by its low levels of nucleotide diversity, population differentiation and low tolerance of amino acid-

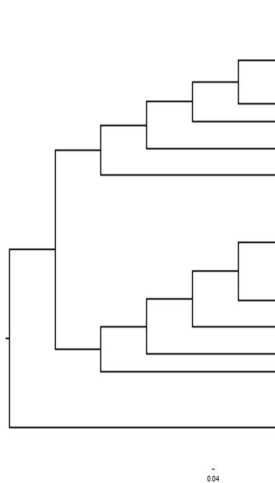
altering variation. It also exhibited a dramatic decay in the  $\omega$  ratio when compared to the other two RLRs [68]. This is the expected outcome in evolutionary studies when using closely related species, or genetic information for population of the same species, which result in a background of strong purifying selection to keep the protein functional integrity. In the same study [68], the strongest signatures of positive selection were found in MDA5 and LGP2 by exhibiting higher  $\omega$  ratios than



A



B



C

	Codons																			
	11	53	59	73	144	149	156	197	212	218	242	305	395	533	673	684	846	871	909	1002
Hosa	E	A	Q	V	E	I	A	V	I	V	G	A	I	D	L	R	H	C	N	V
Gogo	E	A	Q	V	E	I	A	V	I	A	G	A	I	D	L	R	H	C	N	V
Patr	E	A	Q	V	E	I	A	V	-	-	-	A	I	D	L	R	H	C	N	V
Papa	E	A	Q	V	E	I	A	V	I	V	G	A	I	D	L	R	H	C	N	V
Poab	E	A	Q	V	E	I	?	V	I	V	G	A	I	D	V	R	H	C	S	V
Nole	E	A	Q	V	E	I	A	V	-	-	-	A	I	D	L	R	H	C	N	V
Mamu	E	A	Q	V	E	I	A	V	I	D	G	A	I	D	L	R	H	G	N	V
Sabo	E	A	Q	V	E	I	S	V	T	D	G	A	I	D	L	R	P	H	E	D
Caja	E	A	Q	V	E	I	S	V	I	D	G	A	I	D	L	R	S	H	E	D
Otga	S	V	R	E	K	T	S	A	I	E	G	A	T	D	L	R	H	K	A	A
Bota	K	A	Q	N	E	I	S	A	S	E	D	Q	T	N	G	D	Q	R	G	V
Ovar	K	A	Q	N	E	V	S	A	S	E	D	Q	T	N	G	D	Q	R	D	V
Susc	Q	A	N	V	E	G	S	A	S	E	D	Q	T	N	E	I	Q	R	D	A
Mupu	K	V	Q	G	K	N	S	V	T	E	D	E	I	H	E	R	H	H	E	V
Aime	G	V	Q	G	K	N	C	V	P	R	D	E	I	V	E	R	H	H	E	V
Calu	Q	A	Q	A	K	D	S	V	T	E	D	E	I	N	E	R	H	H	E	I
Eqca	K	A	Q	V	V	L	A	V	T	E	D	E	T	D	E	S	H	H	D	V
Mylu	K	A	G	V	V	S	A	S	E	D	E	T	C	G	K	H	R	E	S	
Ptal	K	T	N	V	E	I	A	V	T	E	D	K	M	D	E	R	C	R	E	V
Loaf	K	A	H	S	N	T	S	V	I	R	D	R	M	G	E	R	P	R	D	V
Orcu	K	D	Q	I	E	T	S	A	A	E	D	V	I	D	F	R	C	R	E	V
Crgu	D	A	S	E	K	P	S	C	L	E	D	Q	I	G	V	E	S	R	N	E
Mumu	D	N	S	Q	K	V	S	F	L	R	N	K	I	G	V	E	S	R	D	N
Rano	D	T	S	Q	K	V	S	V	L	K	T	K	I	S	V	E	S	R	S	N
Ictr	K	A	R	V	K	L	T	A	I	E	N	E	I	E	L	K	P	C	D	G
Capo	K	N	K	G	V	I	S	A	I	E	H	K	T	N	T	R	H	H	N	V

**Figure 5. Structural representation and identification of positively-selected branches and codons in mammalian MDA5.** (A) Based on the human protein structure, the key domains of MDA5 (<http://www.uniprot.org/uniprot/Q9BYX4>) and the corresponding boundaries are schematically represented. Also, the human domain boundaries while in the mammalian MDA5 deduced protein sequences alignment (Figure S5) are shown in brackets. (B) Cladogram of 26 mammalian MDA5 genes collected from Ensembl and NCBI databases. Branch-site analyses were performed to identify specific branches episodic under positive selection. Branches with statistically significant likelihood ratio tests (LRTs) when performing PAML branch-site model A (Table 3) are colored in green; branches simultaneously identified by PAML branch-site model A and Hyphy branch-site REL method (Table 5) are colored in blue. (C) Positively-selected codons are exhibited in the table and numbered according to the mammalian MDA5 deduced protein sequences alignment (Figure S5). Symbol “?” represents an undetermined amino acid, while “-” symbolizes a deletion. Colors on the codon numbering row correspond to the MDA5 domain with the same color in the protein structural representation (A). The background colors on the identified sites match different amino acid properties: polar positive (yellow), polar negative (orange), polar neutral (green), non-polar neutral (purple), non-polar aliphatic (blue) and non-polar aromatic (pink). The used abbreviations correspond, by order of appearance, to the following species: Hosa – Human; Gogo – Gorilla; Patr – Chimpanzee; Papa – Bonobo; Poab – Orangutan; Nole – Gibbon; Mamu – Rhesus macaque; Sabo – Black-capped squirrel monkey; Caja – Marmoset; Otga – Bushbaby; Bota – Cow; Ovar – Sheep; Susc – Pig; Mupu – Ferret; Aime – Giant panda; Calu – Dog; Eqca – Horse; Mylu – Little brown myotis; Ptal – Black flying fox; Loaf – Elephant; Orcu – European rabbit; Crgu – Chinese hamster; Mumu – Mouse; Rano – Rat; Ictr – Squirrel; Capo – Guinea pig.

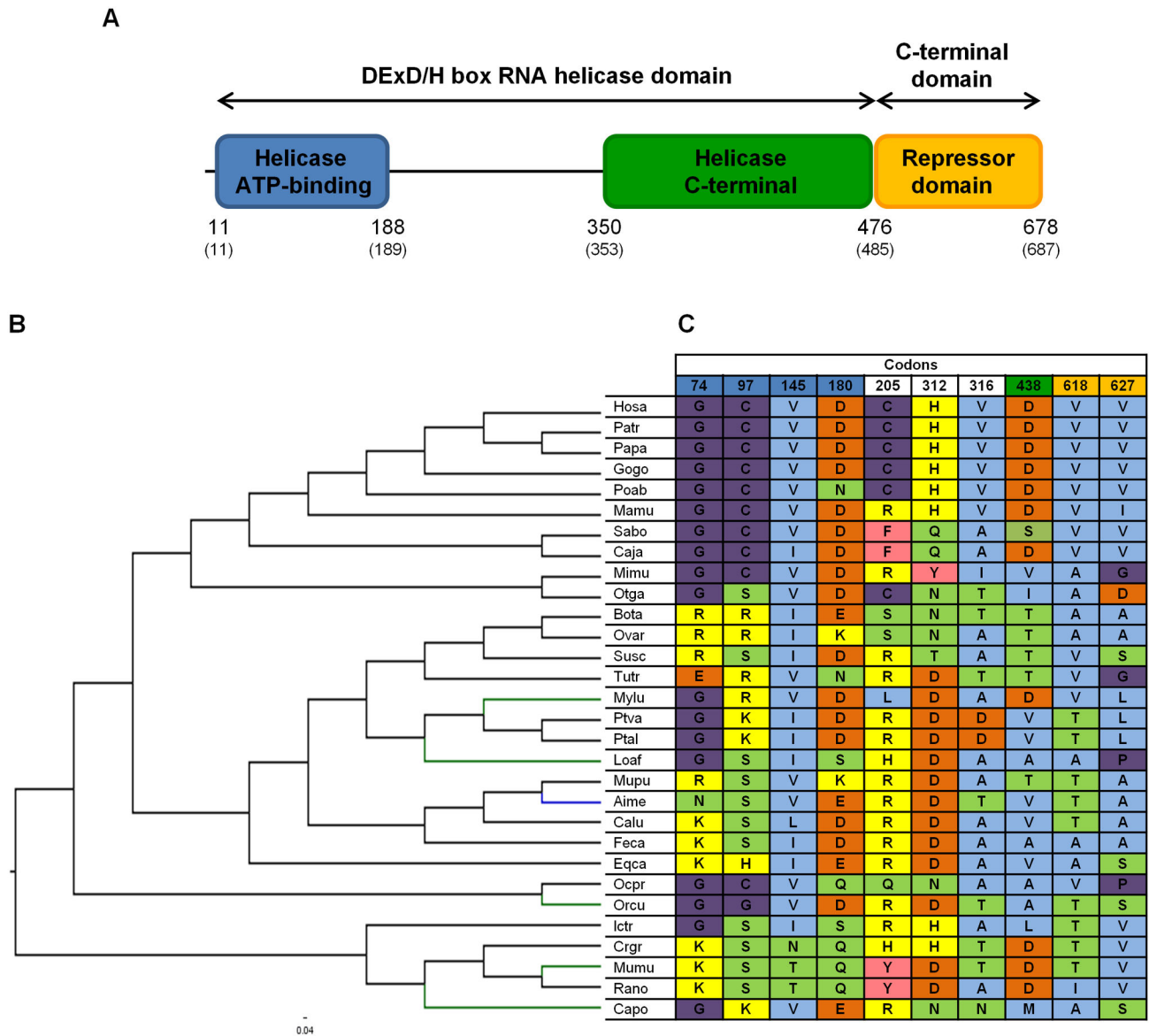
doi: 10.1371/journal.pone.0081864.g005

RIG-I. Besides, MDA5 and LGP2 also appear to have evolved adaptively in specific human populations, presenting a great number of nonsynonymous mutations in both helicase and C-terminal domains [68].

RIG-I and MDA5 contain two N-terminal CARDs [10,17]. The interaction of these domains with an adaptor protein named IPS-1 (also known as MAVS, VISA or CARDIF) is a crucial process to activate a wide range of downstream response

factors, including type I IFNs and other essential anti-viral proteins to induce intracellular immune responses [77]. Interestingly, in our study, the CARDs of both RIG-I and MDA5 concentrated a large number of the deduced codons under selection. Some of these are radical in terms of their physicochemical properties changes across mammalian species (Figure 4 and Figure 5), strengthening the case for positive selection. Since the two CARDs are fundamental for





**Figure 6. Structural representation and identification of positively-selected branches and codons in mammalian LGP2.** (A) Based on human protein structure, the key domains of LGP2 (<http://www.uniprot.org/uniprot/Q96C10>) and the corresponding boundaries are schematically represented. Also, the human domain boundaries while in the mammalian LGP2 deduced protein sequences alignment (Figure S6) are shown in brackets. (B) Cladogram of 30 mammalian LGP2 genes collected from Ensembl and NCBI databases. Branch-site analyses were performed to identify specific branches under episodic positive selection. Branches with statistically significant likelihood ratio tests (LRTs) when performing PAML branch-site model A (Table 4) are colored in green; branch colored in blue has been simultaneously identified by PAML branch-site model A and Hyphy branch-site REL method (Table 5). (C) Positively-selected codons are exhibited in the table and numbered according to the mammalian LGP2 deduced protein sequences alignment (Figure S6). Colors on the codon numbering row correspond to the LGP2 domain with the same color in the protein structural representation (A). The background colors on the identified codons match different amino acid properties: polar positive (yellow), polar negative (orange), polar neutral (green), non-polar neutral (purple), non-polar aliphatic (blue) and non-polar aromatic (pink). The used abbreviations correspond, by order of appearance, to the following species: Hosa – Human; Patr – Chimpanzee; Papa – Bonobo; Gogo – Gorilla; Poab – Orangutan; Mamu – Rhesus macaque; Sabo – Black-capped squirrel monkey; Caja – Marmoset; Mimu – Mouse lemur; Otga – Bushbaby; Bota – Cow; Ovar – Sheep; Susc – Pig; Tutr – Dolphin; Mylu – Little brown myotis; Ptva – Large flying fox; Ptal – Black flying fox; Loaf – Elephant; Mupu – Ferret; Aime – Giant panda; Calu – Dog; Feca – Cat; Eqca – Horse; Ocpr – American pika; Orcu – European rabbit; Ictr – Squirrel; Crgr – Chinese hamster; Mumu – Mouse; Rano – Rat; Capo – Guinea pig.

doi: 10.1371/journal.pone.0081864.g006

**Table 2.** PAML branch-site model A analysis to identify branches under episodic positive selection in RIG-I phylogenetic tree.

Foreground branches <sup>a</sup>	Parameters under null model	lnL <sup>b</sup> (null)	Parameters under alternative model	lnL <sup>b</sup> (alternative)	2ΔlnL <sup>c</sup>	p-Value	Positively selected sites <sup>d</sup>
Bota	$p_0 = 0.536$ $p_1 = 0.294$ $p_{2a} = 0.110$ $p_{2b} = 0.060$ $\omega_0 = 0.090$ $\omega_1 = 1$ $\omega_2 = 1$	-21065.614	$p_0 = 0.637$ $p_1 = 0.349$ $p_{2a} = 0.009$ $p_{2b} = 0.005$ $\omega_0 = 0.090$ $\omega_1 = 1$ $\omega_2 = \mathbf{18.749}$	-21063.247	<b>4.734</b>	<0.05	<b>655F</b> (0.907) <b>656Q</b> (0.996)
Calu	$p_0 = 0.631$ $p_1 = 0.343$ $p_{2a} = 0.016$ $p_{2b} = 0.009$ $\omega_0 = 0.092$ $\omega_1 = 1$ $\omega_2 = 1$	-21069.596	$p_0 = 0.645$ $p_1 = 0.350$ $p_{2a} = 0.003$ $p_{2b} = 0.002$ $\omega_0 = 0.093$ $\omega_1 = 1$ $\omega_2 = \mathbf{70.492}$	-21065.810	<b>7.572</b>	<0.01	none
Capo	$p_0 = 0.617$ $p_1 = 0.335$ $p_{2a} = 0.031$ $p_{2b} = 0.017$ $\omega_0 = 0.090$ $\omega_1 = 1$ $\omega_2 = 1$	-21068.600	$p_0 = 0.641$ $p_1 = 0.347$ $p_{2a} = 0.008$ $p_{2b} = 0.004$ $\omega_0 = 0.092$ $\omega_1 = 1$ $\omega_2 = \mathbf{e}$	-21065.132	<b>6.936</b>	<0.01	none
Eqca	$p_0 = 0.648$ $p_1 = 0.352$ $p_{2a} = 0$ $p_{2b} = 0$ $\omega_0 = 0.093$ $\omega_1 = 1$ $\omega_2 = 1$	-21069.711	$p_0 = 0.643$ $p_1 = 0.351$ $p_{2a} = 0.004$ $p_{2b} = 0.002$ $\omega_0 = 0.092$ $\omega_1 = 1$ $\omega_2 = \mathbf{47.106}$	-21067.540	<b>4.342</b>	<0.05	none
Feca	$p_0 = 0.613$ $p_1 = 0.333$ $p_{2a} = 0.035$ $p_{2b} = 0.019$ $\omega_0 = 0.092$ $\omega_1 = 1$ $\omega_2 = 1$	-21068.674	$p_0 = 0.643$ $p_1 = 0.351$ $p_{2a} = 0.004$ $p_{2b} = 0.002$ $\omega_0 = 0.092$ $\omega_1 = 1$ $\omega_2 = \mathbf{265.180}$	-21064.248	<b>8.852</b>	<0.005	<b>44W</b> (0.943)
Ictr	$p_0 = 0.611$ $p_1 = 0.331$ $p_{2a} = 0.038$ $p_{2b} = 0.020$ $\omega_0 = 0.091$ $\omega_1 = 1$ $\omega_2 = 1$	-21068.446	$p_0 = 0.644$ $p_1 = 0.348$ $p_{2a} = 0.005$ $p_{2b} = 0.003$ $\omega_0 = 0.093$ $\omega_1 = 1$ $\omega_2 = \mathbf{16.527}$	-21066.281	<b>4.330</b>	<0.05	none
Mumu	$p_0 = 0.648$ $p_1 = 0.352$ $p_{2a} = 0$ $p_{2b} = 0$ $\omega_0 = 0.093$ $\omega_1 = 1$ $\omega_2 = 1$	-21069.711	$p_0 = 0.646$ $p_1 = 0.346$ $p_{2a} = 0.005$ $p_{2b} = 0.003$ $\omega_0 = 0.093$ $\omega_1 = 1$ $\omega_2 = \mathbf{f}$	-21067.171	<b>5.080</b>	<0.025	none
Mylu	$p_0 = 0.648$ $p_1 = 0.352$ $p_{2a} = 0$ $p_{2b} = 0$ $\omega_0 = 0.093$ $\omega_1 = 1$ $\omega_2 = 1$	-21069.711	$p_0 = 0.645$ $p_1 = 0.351$ $p_{2a} = 0.002$ $p_{2b} = 0.001$ $\omega_0 = 0.093$ $\omega_1 = 1$ $\omega_2 = \mathbf{172.548}$	-21066.025	<b>7.372</b>	<0.01	none
Orcu	$p_0 = 0.599$ $p_1 = 0.320$ $p_{2a} = 0.053$ $p_{2b} = 0.028$ $\omega_0 = 0.089$ $\omega_1 = 1$ $\omega_2 = 1$	-21064.127	$p_0 = 0.637$ $p_1 = 0.336$ $p_{2a} = 0.017$ $p_{2b} = 0.009$ $\omega_0 = 0.092$ $\omega_1 = 1$ $\omega_2 = \mathbf{38.747}$	-21046.263	<b>35.728</b>	<0.001	<b>849T</b> (0.989) <b>851C</b> (0.997) <b>854S</b> (0.958) <b>857H</b> (0.988) <b>861G</b> (0.982) <b>898V</b> (0.989)

<sup>a</sup> Species names on the foreground branches: Bota – Cow; Calu – Dog; Capo – Guinea pig; Eqca – Horse; Feca – Cat; Ictr – Squirrel; Mumu – Mouse; Mylu – Little brown myotis; Orcu – European rabbit.

<sup>b</sup> lnL: log-likelihood scores.

<sup>c</sup> 2ΔlnL: likelihood ratio test (LRT) to detect positive selection.

<sup>d</sup> Positively selected sites: posterior probabilities >0.90 in the BEB (Bayes Empirical Bayes) analyses.

<sup>e</sup>  $\omega_2$  parameter varied for different ( $\omega$ ) values:  $\omega_2$  (2) = 242.957;  $\omega_2$  (3) = 340.801;  $\omega_2$  (4) = 982.380.

<sup>f</sup>  $\omega_2$  parameter varied for different ( $\omega$ ) values:  $\omega_2$  (2) = 131.879;  $\omega_2$  (3) = 246.814;  $\omega_2$  (4) = 289.634.

doi: 10.1371/journal.pone.0081864.t002

downstream RIG-I and MDA5 signaling, which implies functional constraints, the observed variability across species can be perceived as a great structural plasticity for mammalian CARDs.

The helicase domain in the RLR family is generally described as exhibiting affinity for dsRNA [78]. The existence of six highly conserved sequence motifs within this domain is a characteristic of the helicase superfamily 2 which integrates DExD/H box RNA helicases. Also, different aspects of helicase functions have been assigned to specific motifs [79]. Bamming and Horvath [11] compared the amino acid sequences of the three human RLR helicase domains with the established consensus sequences of helicase families elements and, despite slight differences, the sequences in individual motifs are highly conserved within RIG-I, MDA5 and LGP2. Indeed, in our study the six helicase motifs of the three proteins were evolutionary conserved (data not shown) in the mammalian species collected. Minor alterations occur in some species, but the extent of those differences concerning the involvement in

substrate interaction, signal transduction and/or the whole anti-viral response profile, is difficult to predict.

RIG-I RD is responsible for recognizing and binding to its RNA substrates in a 5'-triphosphate (5'-ppp)-dependent manner. Besides, binding studies clearly established that the pppRNA binding site resides within the RD [14,26,80]. The function described for RIG-I RD makes our current results worthy of note, since the RD is the RIG-I domain that exhibits the strongest evidence of trans-acting selective pressures (Figure 4). Whether these differences play a role in RIG-I activation after binding to the RNAs from different viral pathogens that infect distinct mammalian hosts is a complex question. Nevertheless, we can assume that the RD variability in mammals is related to the fact that RIG-I senses a large variety of viruses [13,14].

The performance of branch-site models in our study imposes a careful interpretation of data, since only one representative element of each species was included. Still, some branches of the three RLR phylogenetic trees exhibited evidence of positive selection. The two species under episodic positive selection on

**Table 3.** PAML branch-site model A analysis to identify branches under episodic positive selection in MDA5 phylogenetic trees.

Foreground branches <sup>a</sup>	Parameters under null model	lnL <sup>b</sup> (null)	Parameters under alternative model	lnL <sup>b</sup> (alternative)	2ΔlnL <sup>c</sup>	p-Value	Positively selected sites <sup>d</sup>
<b>1<sup>st</sup> Segment</b>							
Aime	$p_0 = 0.506$ $p_1 = 0.377$ $p_{2a} = 0.067$ $p_{2b} = 0.050$ $\omega_0 = 0.086$ $\omega_1 = 1$ $\omega_2 = 1$	-7981.227	$p_0 = 0.559$ $p_1 = 0.418$ $p_{2a} = 0.013$ $p_{2b} = 0.010$ $\omega_0 = 0.086$ $\omega_1 = 1$ $\omega_2 = 27.801$	-7976.648	<b>9.158</b>	<0.005	<b>295P</b> (0.995) <b>281K</b> (0.972) <b>284F</b> (0.993) <b>291P</b> (0.996) <b>293L</b> (0.989) <b>297I</b> (0.998)
Capo	$p_0 = 0.488$ $p_1 = 0.364$ $p_{2a} = 0.085$ $p_{2b} = 0.063$ $\omega_0 = 0.077$ $\omega_1 = 1$ $\omega_2 = 1$	-7973.674	$p_0 = 0.543$ $p_1 = 0.409$ $p_{2a} = 0.027$ $p_{2b} = 0.021$ $\omega_0 = 0.078$ $\omega_1 = 1$ $\omega_2 = 27.756$	-7964.534	<b>18.280</b>	<0.001	<b>358G</b> (0.958) <b>593S</b> (1) <b>594S</b> (1) <b>595L</b> (0.990)
<b>2<sup>nd</sup> Segment</b>							
Capo	$p_0 = 0.727$ $p_1 = 0.205$ $p_{2a} = 0.053$ $p_{2b} = 0.015$ $\omega_0 = 0.070$ $\omega_1 = 1$ $\omega_2 = 1$	-14272.950	$p_0 = 0.759$ $p_1 = 0.209$ $p_{2a} = 0.025$ $p_{2b} = 0.007$ $\omega_0 = 0.073$ $\omega_1 = 1$ $\omega_2 = 998.992$	-14255.172	<b>35.556</b>	<0.001	<b>593S</b> (1) <b>594S</b> (1) <b>595L</b> (0.990)
Crggr	$p_0 = 0.778$ $p_1 = 0.222$ $p_{2a} = 0$ $p_{2b} = 0$ $\omega_0 = 0.074$ $\omega_1 = 1$ $\omega_2 = 1$	-14281.640	$p_0 = 0.778$ $p_1 = 0.219$ $p_{2a} = 0.003$ $p_{2b} = 0.001$ $\omega_0 = 0.075$ $\omega_1 = 1$ $\omega_2 = e$	-14277.837	<b>7.606</b>	<0.01	none

<sup>a</sup> Species names on the foreground branches: Aime – Giant panda; Capo – Guinea pig; Crggr – Chinese hamster.

<sup>b</sup> lnL: log-likelihood scores.

<sup>c</sup> 2ΔlnL: likelihood ratio test (LRT) to detect positive selection.

<sup>d</sup> Positively selected sites: posterior probabilities >0.90 in the BEB (Bayes Empirical Bayes) analyses.

<sup>e</sup>  $\omega_2$  parameter varied for different ( $\omega$ ) values:  $\omega_2$  (2) = 999.000;  $\omega_2$  (3) = 832.570;  $\omega_2$  (4) = 681.973.

doi: 10.1371/journal.pone.0081864.t003

**Table 4.** PAML branch-site model A analysis to identify branches under episodic positive selection in LGP2 phylogenetic tree.

Foreground branches <sup>a</sup>	Parameters under null model	lnL <sup>b</sup> (null)	Parameters under alternative model	lnL <sup>b</sup> (alternative)	2ΔlnL <sup>c</sup>	p-Value	Positively selected sites <sup>d</sup>
Aime	$p_0 = 0.729$ $p_1 = 0.182$ $p_{2a} = 0.070$ $p_{2b} = 0.018$ $\omega_0 = 0.082$ $\omega_1 = 1$ $\omega_2 = 1$	-18816.537	$p_0 = 0.794$ $p_1 = 0.199$ $p_{2a} = 0.005$ $p_{2b} = 0.001$ $\omega_0 = 0.082$ $\omega_1 = 1$ $\omega_2 = 86.683$	-18810.272	<b>12.530</b>	<0.001	<b>227L</b> (0.913) <b>519C</b> (0.999) <b>520N</b> (0.995)
Capo	$p_0 = 0.794$ $p_1 = 0.198$ $p_{2a} = 0.006$ $p_{2b} = 0.002$ $\omega_0 = 0.083$ $\omega_1 = 1$ $\omega_2 = 1$	-18821.176	$p_0 = 0.795$ $p_1 = 0.197$ $p_{2a} = 0.006$ $p_{2b} = 0.001$ $\omega_0 = 0.083$ $\omega_1 = 1$ $\omega_2 = 12.697$	-18819.037	<b>4.278</b>	<0.05	<b>176N</b> (0.929) <b>267S</b> (0.979)
Loaf	$p_0 = 0.787$ $p_1 = 0.196$ $p_{2a} = 0.014$ $p_{2b} = 0.003$ $\omega_0 = 0.083$ $\omega_1 = 1$ $\omega_2 = 1$	-18820.429	$p_0 = 0.792$ $p_1 = 0.196$ $p_{2a} = 0.009$ $p_{2b} = 0.002$ $\omega_0 = 0.083$ $\omega_1 = 1$ $\omega_2 = 423.742$	-18812.759	<b>15.340</b>	<0.001	<b>421S</b> (0.996) <b>569S</b> (0.985)
Mimu	$p_0 = 0.801$ $p_1 = 0.199$ $p_{2a} = 0$ $p_{2b} = 0$ $\omega_0 = 0.083$ $\omega_1 = 1$ $\omega_2 = 1$	-18821.530	$p_0 = 0.797$ $p_1 = 0.198$ $p_{2a} = 0.004$ $p_{2b} = 0.001$ $\omega_0 = 0.083$ $\omega_1 = 1$ $\omega_2 = 775.269$	-18819.489	<b>4.082</b>	<0.05	none
Mylu	$p_0 = 0.769$ $p_1 = 0.190$ $p_{2a} = 0.033$ $p_{2b} = 0.008$ $\omega_0 = 0.082$ $\omega_1 = 1$ $\omega_2 = 1$	-18819.788	$p_0 = 0.795$ $p_1 = 0.193$ $p_{2a} = 0.009$ $p_{2b} = 0.002$ $\omega_0 = 0.083$ $\omega_1 = 1$ $\omega_2 = 11.149$	-18817.547	<b>4.482</b>	<0.05	<b>646A</b> (0.980)
Orcu	$p_0 = 0.779$ $p_1 = 0.193$ $p_{2a} = 0.022$ $p_{2b} = 0.006$ $\omega_0 = 0.083$ $\omega_1 = 1$ $\omega_2 = 1$	-18820.688	$p_0 = 0.797$ $p_1 = 0.197$ $p_{2a} = 0.005$ $p_{2b} = 0.001$ $\omega_0 = 0.083$ $\omega_1 = 1$ $\omega_2 = 63.328$	-18817.321	<b>6.734</b>	<0.01	<b>605P</b> (0.961)

<sup>a</sup> Species names on the foreground branches: Aime – Giant panda; Capo – Guinea pig; Loaf – Elephant; Mimu – Mouse lemur; Mylu – Little brown myotis; Orcu – European rabbit.

<sup>b</sup> lnL: log-likelihood scores.

<sup>c</sup> 2ΔlnL: likelihood ratio test (LRT) to detect positive selection.

<sup>d</sup> Positively selected sites: posterior probabilities >0.90 in the BEB (Bayes Empirical Bayes) analyses.

doi: 10.1371/journal.pone.0081864.t004

RIG-I phylogenetic tree, the domestic dog and the European rabbit, are susceptible hosts of two viruses recognized by RIG-I, rabies virus (Rhabdoviridae family) and myxoma virus

(Poxviridae family), respectively [23,39]. Such results suggest that these lethal pathogens, and possibly other re-occurring viral infections in these specific hosts, might be exerting long-

**Table 5.** Hyphy branch-site REL analysis to identify RIG-I-like receptor species branches subject to episodic diversifying selection.

Branch <sup>a</sup>	$\omega^+{}^b$	p-Value
<b>RIG-I</b>		
Calu	57.89	0.032
Orcu	30.23	< 0.0001
<b>MDA5</b>		
Aime	11.46	0.004
Capo	3334.49	< 0.0001
<b>LGP2</b>		
Aime	196.66	0.043

<sup>a</sup> Species names on the branches: Aime – Giant panda; Calu – Dog; Capo – Guinea pig; Orcu – European rabbit.

<sup>b</sup>  $\omega^+$ : strength and extent (proportion of sites) of selection along each branch.

doi: 10.1371/journal.pone.0081864.t005

term selective pressures on the susceptible host RIG-I gene. Therefore, the changes on RIG-I sequences across species, with special focus on the RD as suggested above, should be the result of a co-evolutionary process between species-specific infecting viruses and this host RNA sensor protein.

By detecting the extension of acting positive selection on mammalian RLRs, this study provides further insights into their biological functions in host defense against viral pathogens in general. Differences in these genes across mammalian species may consequently impact downstream immune responses and, as a result, contribute to the species-specific resistance/susceptibility profiles against many diverse viral pathogens.

## Materials and Methods

### Mammalian RIG-I-like receptor gene sequences

The coding region of the three RLR genes, RIG-I, MDA5 and LGP2, were collected for different mammalian species from NCBI (<http://www.ncbi.nlm.nih.gov>) and Ensembl (<http://www.ensembl.org/index.html>) databases (Table S1). Each set of mammalian orthologous gene sequences was aligned with ClustalW [81] implemented in BioEdit v7.0.9 [82]. The nucleotide sequences alignment corresponding to each gene coding region is represented in Figure S1 (RIG-I alignment), Figure S2 (MDA5 alignment) and Figure S3 (LGP2 alignment). Translation into protein sequences was performed using also BioEdit [82]. Figure S4, Figure S5 and Figure S6 represent the alignments of the deduced protein sequences for RIG-I, MDA5 and LGP2, respectively. For the evolutionary analyses, representative alignment gaps in Figure S1, Figure S2 and Figure S3 had to be removed: gaps present in all sequences, with the exception of one or two, have been removed, while gaps present in only one or two sequences were kept. Figure S7 (RIG-I alignment), Figure S8 (MDA5 alignment) and Figure S9 (LGP2 alignment) correspond to trimmed versions of the nucleotide sequences alignment of each RLR gene.

### Phylogenetic reconstruction analyses

The nucleotide sequences alignment of each gene was firstly tested for recombination, as this biological process can mislead phylogenetic and positive selection analyses [83]. We used the software GARD (Genetic Algorithm for Recombination Detection) [50,51], implemented in the Datamonkey web server [52,53], to detect possible recombination breakpoints on each alignment. The nucleotide substitution model for each phylogenetic reconstruction was indicated by the Akaike Information Criterion (AIC) implemented in jModelTest v0.1.1 [84].

Regarding the RIG-I gene one breakpoint was identified, but it was not supported by the Kishino-Hasegawa test. Therefore, the complete alignment was used for the gene phylogeny reconstruction and GTR+G nucleotide substitution model was indicated as the best-fitting model. On the other hand, the software GARD found evidence of two breakpoints in the MDA5 gene alignment. However, only one of the breakpoints (nucleotide 903) reflected a significant topological incongruence (Kishino-Hasegawa test,  $p < 0.01$ ), suggesting that the multiple tree model can be preferred over the single tree model. We reconstructed MDA5 phylogeny for the first 903 nucleotides of the mammalian alignment as also for the remaining 2211 nucleotides. To compare the different MDA5 trees topology, we also used the complete alignment (no recombination testing) to reconstruct the gene phylogeny and GTR+G nucleotide substitution model was indicated by the AIC as the best-fit. For the MDA5 segments which resulted from recombination detection, the best-fitting nucleotide substitution models were TIM3+G (first segment) and TIM3+I+G (second segment). Finally, we found no evidence of recombination for the LGP2 gene alignment. The best-fitting nucleotide substitution model determined for this alignment was the TPM2uf+I+G model.

ML phylogenetic reconstruction was performed for the three genes using GARLI v2.0 (Genetic Algorithm for Rapid Likelihood Inference) [85]. The analyses were performed with 1,000,000 generations and 1,000 bootstrap searches. ML trees were displayed using FigTree v1.3.1 (<http://tree.bio.ed.ac.uk/>).

### Molecular evolutionary analyses

Codon substitution models implemented in the CODEML program in PAML v4.4 (Phylogenetic Analysis by Maximum Likelihood) package [54,55] were applied to the trimmed alignments of RIG-I (Figure S7), MDA5 (Figure S8) and LGP2 (Figure S9). The codon frequency model F3x4 was fitted to all the alignments. Two pairs of site-specific models were used, M1a (nearly neutral) versus M2a (selection) and M7 (neutral, beta) versus M8 (selection, beta &  $\omega$ ). In these comparisons, M1a and M7 neutral models (null hypothesis) do not admit positive selection, while M2a and M8 alternative models allow positive selection. A LRT with 2 degrees of freedom was performed, where a significant LRT demonstrates that the selection model fits better than the neutral model [56,86,87]. From the HyPhy software available on the Datamonkey web server [52,53], the PARRIS method [57] was also applied to detect if a proportion of sites in each RLR alignment evolved with  $\omega$  ( $d_N/d_S$ ) > 1.

Six different codon-based ML methods were applied to detect codons under positive selection on mammalian RIG-I, MDA5 and LGP2 trimmed alignments, and based on the methodology adopted by other authors and in previous studies [43,47,64], only codons identified by at least three of the six used methods were considered to be under positive selection. Model M8 from PAML package [54,55] was one of the codon-based methods used to detect codons under positive selection, and a Bayes empirical Bayes (BEB) approach was employed to detect codons with a posterior probability >90% of being under selection [88]. Five other methods, using HyPhy software implemented in the Datamonkey web server [52,53], were also applied to detect sites under selection for the three genes: the Single Likelihood Ancestor Counting (SLAC) method, the Fixed Effect Likelihood (FEL) method, the Random Effect Likelihood (REL) method [61] and the recently described Mixed Effects Model of Evolution (MEME) [62] and Fast Unbiased Bayesian AppRoximation (FUBAR) [63] methods. To avoid a high false-positive rate [61], sites with  $p$ -values <0.1 for SLAC, FEL and MEME models, Bayes Factor >50 for REL model and a posterior probability >0.90 for FUBAR were accepted as candidates for selection.

Two branch-site models allowing  $\omega$  ratios to vary both among lineages and amino acid sites were performed: the PAML branch-site model A [66] and the Hyphy branch-site REL method [67]. When performing PAML branch-site model A [66], every species branch was analyzed as a foreground branch independently. For each analysis of a foreground branch, the remaining lineages were denominated as background branches. In branch-site model A, three  $\omega$  ratios are assumed for foreground ( $0 < \omega_0 < 1$ ,  $\omega_1 = 1$ ,  $\omega_2 > 1$ ) and two  $\omega$  ratios for background ( $0 < \omega_0 < 1$ ,  $\omega_1 = 1$ ). The null model is the same as model A, but  $\omega_2 = 1$  is fixed [66]. The BEB approach was also used to calculate the posterior probability of a specific codon site and to identify those most likely to be under positive selection (posterior probability >90%) [88]. On the other hand, the branch-site REL method [67] was applied to identify branches where a proportion of sites evolved under episodic diversifying selection.

## Supporting Information

**Figure S1. Mammalian RIG-I nucleotide coding region sequences alignment.** RIG-I nucleotide coding region sequences for twenty-six mammalian species were collected from Ensembl and NCBI databases, and aligned with ClustalW implemented in BioEdit. The symbol “.” represents the same nucleotide as the reference sequence of human RIG-I gene, “?” symbolizes an undetermined nucleotide and “-” represents a gap or deletion in the alignment. The used abbreviations correspond, by order of appearance, to the following species: Hosa – Human; Patr – Chimpanzee; Papa – Bonobo; Gogo – Gorilla; Poab – Orangutan; Paan – Olive baboon; Mamu – Rhesus macaque; Sabo – Black-capped squirrel monkey; Caja – Marmoset; Mimu – Mouse lemur; Otga – Bushbaby; Bota – Cow; Ovar – Sheep; Susc – Pig; Mylu – Little brown myotis; Ptva – Large flying fox; Ptal – Black flying fox; Aime – Giant panda; Calu – Dog; Feca – Cat; Eqca – Horse; Loaf –

Elephant; Ictr – Squirrel; Capo – Guinea pig; Mumu – Mouse; Orcu – European rabbit. (TIF)

**Figure S2. Mammalian MDA5 nucleotide coding region sequences alignment.** MDA5 nucleotide coding region sequences for twenty-six mammalian species were collected from Ensembl and NCBI databases, and aligned with ClustalW implemented in BioEdit. The symbol “.” represents the same nucleotide as the reference sequence of human MDA5 gene, “?” symbolizes an undetermined nucleotide and “-” represents a gap or deletion in the alignment. The used abbreviations correspond, by order of appearance, to the following species: Hosa – Human; Gogo – Gorilla; Patr – Chimpanzee; Papa – Bonobo; Poab – Orangutan; Nole – Gibbon; Mamu – Rhesus macaque; Sabo – Black-capped squirrel monkey; Caja – Marmoset; Otga – Bushbaby; Bota – Cow; Ovar – Sheep; Susc – Pig; Mupu – Ferret; Aime – Giant panda; Calu – Dog; Eqca – Horse; Mylu – Little brown myotis; Ptal – Black flying fox; Loaf – Elephant; Orcu – European rabbit; Crgr – Chinese hamster; Mumu – Mouse; Rano – Rat; Ictr – Squirrel; Capo – Guinea pig. (TIFF)

**Figure S3. Mammalian LGP2 nucleotide coding region sequences alignment.** LGP2 nucleotide coding region sequences for thirty mammalian species were collected from Ensembl and NCBI databases, and aligned with ClustalW implemented in BioEdit. The symbol “.” represents the same nucleotide as the reference sequence of human LGP2 gene and “-” symbolizes a gap or deletion in the alignment. The used abbreviations correspond, by order of appearance, to the following species: Hosa – Human; Patr – Chimpanzee; Papa – Bonobo; Gogo – Gorilla; Poab – Orangutan; Mamu – Rhesus macaque; Sabo – Black-capped squirrel monkey; Caja – Marmoset; Mimu – Mouse lemur; Otga – Bushbaby; Bota – Cow; Ovar – Sheep; Susc – Pig; Tutr – Dolphin; Mylu – Little brown myotis; Ptva – Large flying fox; Ptal – Black flying fox; Loaf – Elephant; Mupu – Ferret; Aime – Giant panda; Calu – Dog; Feca – Cat; Eqca – Horse; Ocpr – American pika; Orcu – European rabbit; Ictr – Squirrel; Crgr – Chinese hamster; Mumu – Mouse; Rano – Rat; Capo – Guinea pig. (TIF)

**Figure S4. Mammalian RIG-I deduced protein sequences alignment.** RIG-I deduced protein sequences for twenty-six mammalian species were collected from Ensembl and NCBI databases, and aligned with ClustalW implemented in BioEdit. The symbol “.” represents the same codon as the reference sequence of human RIG-I protein, “?” symbolizes an undetermined codon and “-” represents a gap or deletion in the alignment. The used abbreviations correspond, by order of appearance, to the following species: Hosa – Human; Patr – Chimpanzee; Papa – Bonobo; Gogo – Gorilla; Poab – Orangutan; Paan – Olive baboon; Mamu – Rhesus macaque; Sabo – Black-capped squirrel monkey; Caja – Marmoset; Mimu – Mouse lemur; Otga – Bushbaby; Bota – Cow; Ovar – Sheep; Susc – Pig; Mylu – Little brown myotis; Ptva – Large flying fox;

Ptal – Black flying fox; Aime – Giant panda; Calu – Dog; Feca – Cat; Eqca – Horse; Loaf – Elephant; Ictr – Squirrel; Capo – Guinea pig; Mumu – Mouse; Orcu – European rabbit. (TIFF)

**Figure S5. Mammalian MDA5 deduced protein sequences alignment.** MDA5 deduced protein sequences for twenty-six mammalian species were collected from Ensembl and NCBI databases, and aligned with ClustalW implemented in BioEdit. The symbol “.” represents the same codon as the reference sequence of human MDA5 protein, “?” symbolizes an undetermined codon and “-” represents a gap or deletion in the alignment. The used abbreviations correspond, by order of appearance, to the following species: Hosa – Human; Gogo – Gorilla; Patr – Chimpanzee; Papa – Bonobo; Poab – Orangutan; Nole – Gibbon; Mamu – Rhesus macaque; Sabo – Black-capped squirrel monkey; Caja – Marmoset; Otga – Bushbaby; Bota – Cow; Ovar – Sheep; Susc – Pig; Mupu – Ferret; Aime – Giant panda; Calu – Dog; Eqca – Horse; Mylu – Little brown myotis; Ptal – Black flying fox; Loaf – Elephant; Orcu – European rabbit; Crgr – Chinese hamster; Mumu – Mouse; Rano – Rat; Ictr – Squirrel; Capo – Guinea pig. (TIFF)

**Figure S6. Mammalian LGP2 deduced protein sequences alignment.** LGP2 deduced protein sequences for thirty mammalian species were collected from Ensembl and NCBI databases, and aligned with ClustalW implemented in BioEdit. The symbol “.” represents the same codon as the reference sequence of human LGP2 protein and “-” symbolizes a gap or deletion in the alignment. The used abbreviations correspond, by order of appearance, to the following species: Hosa – Human; Patr – Chimpanzee; Papa – Bonobo; Gogo – Gorilla; Poab – Orangutan; Mamu – Rhesus macaque; Sabo – Black-capped squirrel monkey; Caja – Marmoset; Mimu – Mouse lemur; Otga – Bushbaby; Bota – Cow; Ovar – Sheep; Susc – Pig; Tutr – Dolphin; Mylu – Little brown myotis; Ptva – Large flying fox; Ptal – Black flying fox; Loaf – Elephant; Mupu – Ferret; Aime – Giant panda; Calu – Dog; Feca – Cat; Eqca – Horse; Ocpr – American pika; Orcu – European rabbit; Ictr – Squirrel; Crgr – Chinese hamster; Mumu – Mouse; Rano – Rat; Capo – Guinea pig. (TIF)

**Figure S7. Mammalian RIG-I nucleotide trimmed sequences alignment.** RIG-I nucleotide trimmed sequences for twenty-six mammalian species were collected from Ensembl and NCBI databases, and aligned with ClustalW implemented in BioEdit. The symbol “.” represents the same nucleotide as the reference sequence of human RIG-I gene, “?” symbolizes an undetermined nucleotide and “-” represents a gap or deletion in the alignment. The used abbreviations correspond, by order of appearance, to the following species: Hosa – Human; Patr – Chimpanzee; Papa – Bonobo; Gogo – Gorilla; Poab – Orangutan; Paan – Olive baboon; Mamu – Rhesus macaque; Sabo – Black-capped squirrel monkey; Caja – Marmoset; Mimu – Mouse lemur; Otga – Bushbaby; Bota – Cow; Ovar – Sheep; Susc – Pig; Mylu – Little brown myotis;

Ptva – Large flying fox; Ptal – Black flying fox; Aime – Giant panda; Calu – Dog; Feca – Cat; Eqca – Horse; Loaf – Elephant; Ictr – Squirrel; Capo – Guinea pig; Mumu – Mouse; Orcu – European rabbit. (TIF)

**Figure S8. Mammalian MDA5 nucleotide trimmed sequences alignment.** MDA5 nucleotide trimmed sequences for twenty-six mammalian species were collected from Ensembl and NCBI databases, and aligned with ClustalW implemented in BioEdit. The symbol “.” represents the same nucleotide as the reference sequence of human MDA5 gene, “?” symbolizes an undetermined nucleotide and “-” represents a gap or deletion in the alignment. The used abbreviations correspond, by order of appearance, to the following species: Hosa – Human; Gogo – Gorilla; Patr – Chimpanzee; Papa – Bonobo; Poab – Orangutan; Nole – Gibbon; Mamu – Rhesus macaque; Sabo – Black-capped squirrel monkey; Caja – Marmoset; Otga – Bushbaby; Bota – Cow; Ovar – Sheep; Susc – Pig; Mupu – Ferret; Aime – Giant panda; Calu – Dog; Eqca – Horse; Mylu – Little brown myotis; Ptal – Black flying fox; Loaf – Elephant; Orcu – European rabbit; Crgr – Chinese hamster; Mumu – Mouse; Rano – Rat; Ictr – Squirrel; Capo – Guinea pig. (TIFF)

**Figure S9. Mammalian LGP2 nucleotide trimmed sequences alignment.** LGP2 nucleotide trimmed sequences for thirty mammalian species were collected from Ensembl and NCBI databases, and aligned with ClustalW implemented in BioEdit. The symbol “.” represents the same nucleotide as the reference sequence of human LGP2 gene and “-” symbolizes a gap or deletion in the alignment. The used abbreviations correspond, by order of appearance, to the following species: Hosa – Human; Patr – Chimpanzee; Papa – Bonobo; Gogo – Gorilla; Poab – Orangutan; Mamu – Rhesus macaque; Sabo – Black-capped squirrel monkey; Caja – Marmoset; Mimu – Mouse lemur; Otga – Bushbaby; Bota – Cow; Ovar – Sheep; Susc – Pig; Tutr – Dolphin; Mylu – Little brown myotis; Ptva – Large flying fox; Ptal – Black flying fox; Loaf – Elephant; Mupu – Ferret; Aime – Giant panda; Calu – Dog; Feca – Cat; Eqca – Horse; Ocpr – American pika; Orcu – European rabbit; Ictr – Squirrel; Crgr – Chinese hamster; Mumu – Mouse; Rano – Rat; Capo – Guinea pig. (TIF)

**Table S1. List of mammalian species and gene accession numbers used in this study.** (PDF)

**Table S2. RIG-I, MDA5 and LGP2 likelihood ratio test (LRT) for PARRIS analysis from HyPhy software.** (PDF)

**Table S3. Positively-selected codon positions for RIG-I, MDA5 and LGP2 determined by six different methods.** (PDF)



## Author Contributions

Conceived and designed the experiments: ALM GM PJE.  
Performed the experiments: ALM. Analyzed the data: ALM.

Contributed reagents/materials/analysis tools: ALM PJE. Wrote the manuscript: ALM GM PJE.

## References

- Hoffmann JA (2003) The immune response of *Drosophila*. *Nature* 426: 33-38. doi:10.1038/nature02021. PubMed: 14603309.
- Akira S, Uematsu S, Takeuchi O (2006) Pathogen recognition and innate immunity. *Cell* 124: 783-801. doi:10.1016/j.cell.2006.02.015. PubMed: 16497588.
- Kawai T, Akira S (2009) The roles of TLRs, RLRs and NLRs in pathogen recognition. *Int Immunol* 21: 317-337. doi:10.1093/intimm/dxp017. PubMed: 19246554.
- Janeway CA Jr. (1989) Approaching the asymptote? Evolution and revolution in immunology. *Cold Spring Harb Symp Quant Biol* 54: 1-13.
- Medzhitov R, Janeway C Jr. (2000) Innate immunity. *N Engl J Med* 343: 338-344. doi:10.1056/NEJM200008033430506. PubMed: 10922424.
- Medzhitov R (2007) Recognition of microorganisms and activation of the immune response. *Nature* 449: 819-826. doi:10.1038/nature06246. PubMed: 17943118.
- van Vliet SJ, Garcia-Vallejo JJ, van Kooyk Y (2008) Dendritic cells and C-type lectin receptors: coupling innate to adaptive immune responses. *Immunol Cell Biol* 86: 580-587. doi:10.1038/ncb.2008.55. PubMed: 18679407.
- Eisenächer K, Krug A (2012) Regulation of RLR-mediated innate immune signaling—it is all about keeping the balance. *Eur J Cell Biol* 91: 36-47. doi:10.1016/j.ejcb.2011.01.011. PubMed: 21481967.
- Kawai T, Akira S (2006) Innate immune recognition of viral infection. *Nat Immunol* 7: 131-137. doi:10.1038/ni1303. PubMed: 16424890.
- Yoneyama M, Kikuchi M, Matsumoto K, Imaizumi T, Miyagishi M et al. (2005) Shared and unique functions of the DExD/H-box helicases RIG-I, MDA5, and LGP2 in antiviral innate immunity. *J Immunol* 175: 2851-2858. PubMed: 16116171.
- Bamming D, Horvath CM (2009) Regulation of signal transduction by enzymatically inactive antiviral RNA helicase proteins MDA5, RIG-I, and LGP2. *J Biol Chem* 284: 9700-9712. doi:10.1074/jbc.M807365200. PubMed: 19211564.
- Kato H, Takahashi K, Fujita T (2011) RIG-I-like receptors: cytoplasmic sensors for non-self. *RNA - Immunol Rev* 243: 91-98. doi:10.1111/j.1600-065X.2011.01052.x.
- Loo YM, Gale M Jr. (2011) Immune signaling by RIG-I-like receptors. *Immunology* 34: 680-692. doi:10.1016/j.immuni.2011.05.003. PubMed: 21616437.
- Dixit E, Kagan JC (2013) Intracellular Pathogen Detection by RIG-I-Like Receptors. *Adv Immunol* 117: 99-125. doi:10.1016/B978-0-12-410524-9.00004-9. PubMed: 23611287.
- Bruns AM, Horvath CM (2012) Activation of RIG-I-like receptor signal transduction. *Crit Rev Biochem Mol Biol* 47: 194-206. doi:10.3109/10409238.2011.630974. PubMed: 22066529.
- Schmidt A, Rothenfusser S, Hopfner KP (2012) Sensing of viral nucleic acids by RIG-I: from translocation to translation. *Eur J Cell Biol* 91: 78-85. doi:10.1016/j.ejcb.2011.01.015. PubMed: 21496944.
- Yoneyama M, Kikuchi M, Natsukawa T, Shinobu N, Imaizumi T et al. (2004) The RNA helicase RIG-I has an essential function in double-stranded RNA-induced innate antiviral responses. *Nat Immunol* 5: 730-737. doi:10.1038/ni1087. PubMed: 15208624.
- Saito T, Hirai R, Loo YM, Owen D, Johnson CL et al. (2007) Regulation of innate antiviral defenses through a shared repressor domain in RIG-I and LGP2. *Proc Natl Acad Sci U S A* 104: 582-587. doi:10.1073/pnas.0606699104. PubMed: 17190814.
- Rothenfusser S, Goutagny N, DiPerna G, Gong M, Monks BG et al. (2005) The RNA helicase Lgp2 inhibits TLR-independent sensing of viral replication by retinoic acid-inducible gene-1. *J Immunol* 175: 5260-5268. PubMed: 16210631.
- Venkataraman T, Valdes M, Elsbey R, Kakuta S, Caceres G et al. (2007) Loss of DExD/H box RNA helicase LGP2 manifests disparate antiviral responses. *J Immunol* 178: 6444-6455. PubMed: 17475874.
- Satoh T, Kato H, Kumagai Y, Yoneyama M, Sato S et al. (2010) LGP2 is a positive regulator of RIG-I- and MDA5-mediated antiviral responses. *Proc Natl Acad Sci U S A* 107: 1512-1517. doi:10.1073/pnas.0912986107. PubMed: 20080593.
- Kato H, Takeuchi O, Sato S, Yoneyama M, Yamamoto M et al. (2006) Differential roles of MDA5 and RIG-I helicases in the recognition of RNA viruses. *Nature* 441: 101-105. doi:10.1038/nature04734. PubMed: 16625202.
- Hornung V, Ellegast J, Kim S, Brzózka K, Jung A et al. (2006) 5'-Triphosphate RNA is the ligand for RIG-I. *Science* 314: 994-997. doi:10.1126/science.1132505. PubMed: 17038590.
- Pichlmair A, Schulz O, Tan CP, Näsund TI, Liljestrom P et al. (2006) RIG-I-mediated antiviral responses to single-stranded RNA bearing 5'-phosphates. *Science* 314: 997-1001. doi:10.1126/science.1132998. PubMed: 17038589.
- Schlee M, Roth A, Hornung V, Hagmann CA, Wimmenauer V et al. (2009) Recognition of 5' triphosphate by RIG-I helicase requires short blunt double-stranded RNA as contained in panhandle of negative-strand virus. *Immunology* 31: 25-34. doi:10.1016/j.immuni.2009.05.008. PubMed: 19576794.
- Baum A, Sachidanandam R, Garcia-Sastre A (2010) Preference of RIG-I for short viral RNA molecules in infected cells revealed by next-generation sequencing. *Proc Natl Acad Sci U S A* 107: 16303-16308. doi:10.1073/pnas.1005077107. PubMed: 20805493.
- Berke IC, Li Y, Modis Y (2013) Structural basis of innate immune recognition of viral RNA. *Cell Microbiol* 15: 386-394. doi:10.1111/cmi.12061.
- Habjan M, Andersson I, Klingström J, Schumann M, Martin A et al. (2009) Processing of genome 5' termini as a strategy of negative-strand RNA viruses to avoid RIG-I-dependent interferon induction. *PLOS ONE* 3: e2032. doi:10.1371/journal.pone.0002032. PubMed: 18446221.
- Sumpter R Jr., Loo YM, Foy E, Li K, Yoneyama M et al. (2005) Regulating intracellular antiviral defense and permissiveness to hepatitis C virus RNA replication through a cellular RNA helicase, RIG-I. *J Virol* 79: 2689-2699. doi:10.1128/JVI.79.5.2689-2699.2005. PubMed: 15708988.
- Loo YM, Fornek J, Crochet N, Bajwa G, Perwitasari O et al. (2008) Distinct RIG-I and MDA5 signaling by RNA viruses in innate immunity. *J Virol* 82: 335-345. doi:10.1128/JVI.01080-07. PubMed: 17942531.
- Plumet S, Herschke F, Bourhis JM, Valentin H, Longhi S et al. (2007) Cytosolic 5'-triphosphate ended viral leader transcript of measles virus as activator of the RIG-I-mediated interferon response. *PLOS ONE* 2: e279. doi:10.1371/journal.pone.0000279. PubMed: 17356690.
- Gittin L, Barchet W, Gilfillan S, Cella M, Beutler B et al. (2006) Essential role of mda-5 in type I IFN responses to polyriboinosinic:polyribocytidylic acid and encephalomyocarditis picornavirus. *Proc Natl Acad Sci U S A* 103: 8459-8464. doi:10.1073/pnas.0603082103. PubMed: 16714379.
- McCartney SA, Thackray LB, Gittin L, Gilfillan S, Virgin HW et al. (2008) MDA-5 recognition of a murine norovirus. *PLoS Pathog* 4: e1000108. PubMed: 18636103.
- Roth-Cross JK, Bender SJ, Weiss SR (2008) Murine coronavirus mouse hepatitis virus is recognized by MDA5 and induces type I interferon in brain macrophages/microglia. *J Virol* 82: 9829-9838. doi:10.1128/JVI.01199-08. PubMed: 18667505.
- Fredericksen BL, Keller BC, Fornek J, Katze MG, Gale M Jr. (2008) Establishment and maintenance of the innate antiviral response to West Nile Virus involves both RIG-I and MDA5 signaling through IPS-1. *J Virol* 82: 609-616. doi:10.1128/JVI.01305-07. PubMed: 17977974.
- Ablasser A, Bauernfeind F, Hartmann G, Latz E, Fitzgerald KA et al. (2009) RIG-I-dependent sensing of poly(dA:dT) through the induction of an RNA polymerase III-transcribed RNA intermediate. *Nat Immunol* 10: 1065-1072. doi:10.1038/ni.1779. PubMed: 19609254.
- Chiu YH, Macmillan JB, Chen ZJ (2009) RNA polymerase III detects cytosolic DNA and induces type I interferons through the RIG-I pathway. *Cell* 138: 576-591. doi:10.1016/j.cell.2009.06.015. PubMed: 19631370.
- Samanta M, Iwakiri D, Kanda T, Imaizumi T, Takada K (2006) EB virus-encoded RNAs are recognized by RIG-I and activate signaling to induce type I IFN. *EMBO J* 25: 4207-4214. doi:10.1038/sj.emboj.7601314. PubMed: 16946700.
- Wang F, Gao X, Barrett JW, Shao Q, Bartee E et al. (2008) RIG-I mediates the co-induction of tumor necrosis factor and type I interferon elicited by myxoma virus in primary human macrophages. *PLoS Pathog* 4: e1000099. PubMed: 18617992.

40. Pichlmair A, Schulz O, Tan CP, Rehwinkel J, Kato H et al. (2009) Activation of MDA5 requires higher-order RNA structures generated during virus infection. *J Virol* 83: 10761-10769. doi:10.1128/JVI.00770-09. PubMed: 19656871.
41. Jiggins FM, Hurst GD (2003) The evolution of parasite recognition genes in the innate immune system: purifying selection on *Drosophila melanogaster* peptidoglycan recognition proteins. *J Mol Evol* 57: 598-605. doi:10.1007/s00239-003-2506-6. PubMed: 14738318.
42. Meyerson NR, Sawyer SL (2011) Two-stepping through time: mammals and viruses. *Trends Microbiol* 19: 286-294. doi:10.1016/j.tim.2011.03.006. PubMed: 21531564.
43. Areal H, Abrantes J, Esteves PJ (2011) Signatures of positive selection in Toll-like receptor (TLR) genes in mammals. *BMC Evol Biol* 11: 368. doi:10.1186/1471-2148-11-368. PubMed: 22185391.
44. Han MV, Demuth JP, McGrath CL, Casola C, Hahn MW (2009) Adaptive evolution of young gene duplicates in mammals. *Genome Res* 19: 859-867. doi:10.1101/gr.085951.108. PubMed: 19411603.
45. Machado JP, Johnson WE, O'Brien SJ, Vasconcelos V, Antunes A (2011) Adaptive evolution of the matrix extracellular phosphoglycoprotein in mammals. *BMC Evol Biol* 11: 342. doi:10.1186/1471-2148-11-342. PubMed: 22103247.
46. Grayson P, Civetta A (2012) Positive Selection and the Evolution of izumo Genes in Mammals. *International - Journal of Evolutionary Biology*: 7: 958164. PubMed: 22957301.
47. Lemos de Matos A, Liu J, McFadden G, Esteves PJ (2013) Evolution and divergence of the mammalian SAMD9/SAMD9L gene family. *BMC Evol Biol* 13: 121. doi:10.1186/1471-2148-13-121. PubMed: 23758988.
48. Neves F, Abrantes J, Steinke JW, Esteves PJ (2013) Maximum-likelihood approaches reveal signatures of positive selection in IL genes in mammals. *Innate. Immun.*
49. Pinheiro A, Woof JM, Abi-Rached L, Parham P, Esteves PJ (2013) Computational analyses of an evolutionary arms race between mammalian immunity mediated by immunoglobulin A and its subversion by bacterial pathogens. *PLOS ONE* (in press).
50. Kosakovsky Pond SL, Posada D, Gravenor MB, Woelk CH, Frost SD (2006) Automated phylogenetic detection of recombination using a genetic algorithm. *Mol Biol Evol* 23: 1891-1901. doi:10.1093/molbev/msl051. PubMed: 16818476.
51. Kosakovsky Pond SL, Posada D, Gravenor MB, Woelk CH, Frost SD (2006) GARD: a genetic algorithm for recombination detection. *Bioinformatics* 22: 3096-3098. doi:10.1093/bioinformatics/btl474. PubMed: 17110367.
52. Pond SL, Frost SD (2005) Datamonkey: rapid detection of selective pressure on individual sites of codon alignments. *Bioinformatics* 21: 2531-2533. doi:10.1093/bioinformatics/bti320. PubMed: 15713735.
53. Delport W, Poon AFY, Frost SDW, Kosakovsky Pond SL (2010) Datamonkey 2010: a suite of phylogenetic analysis tools for evolutionary biology. *Bioinformatics* 26: 2455-2457. doi:10.1093/bioinformatics/btq429. PubMed: 20671151.
54. Yang Z (1997) PAML: a program package for phylogenetic analysis by maximum likelihood. *Comput Appl Biosci* 13: 555-556. PubMed: 9367129.
55. Yang Z (2007) PAML 4: phylogenetic analysis by maximum likelihood. *Mol Biol Evol* 24: 1586-1591. doi:10.1093/molbev/msm088. PubMed: 17483113.
56. Yang Z (1998) Likelihood ratio tests for detecting positive selection and application to primate lysozyme evolution. *Mol Biol Evol* 15: 568-573. doi:10.1093/oxfordjournals.molbev.a025957. PubMed: 9580986.
57. Scheffler K, Martin DP, Seoighe C (2006) Robust inference of positive selection from recombining coding sequences. *Bioinformatics* 22: 2493-2499. doi:10.1093/bioinformatics/btl427. PubMed: 16895925.
58. Anisimova M, Bielawski JP, Yang Z (2001) Accuracy and power of the likelihood ratio test in detecting adaptive molecular evolution. *Mol Biol Evol* 18: 1585-1592. doi:10.1093/oxfordjournals.molbev.a003945. PubMed: 11470850.
59. Gharib WH, Robinson-Rechavi M (2013) The Branch-Site Test of Positive Selection Is Surprisingly Robust but Lacks Power under Synonymous Substitution Saturation and Variation in GC. *Mol Biol Evol* 30: 1675-1686. doi:10.1093/molbev/mst062. PubMed: 23558341.
60. Scheffler K, Seoighe C (2005) A Bayesian model comparison approach to inferring positive selection. *Mol Biol Evol* 22: 2531-2540. doi:10.1093/molbev/msi250. PubMed: 16120799.
61. Kosakovsky Pond SL, Frost SD (2005) Not so different after all: a comparison of methods for detecting amino acid sites under selection. *Mol Biol Evol* 22: 1208-1222. doi:10.1093/molbev/msi105. PubMed: 15703242.
62. Murrell B, Wertheim JO, Moola S, Weighill T, Scheffler K et al. (2012) Detecting individual sites subject to episodic diversifying selection. *PLoS Genet* 8: e1002764. PubMed: 22807683.
63. Murrell B, Moola S, Mabona A, Weighill T, Sheward D et al. (2013) FUBAR: a fast, unconstrained bayesian approximation for inferring selection. *Mol Biol Evol* 30: 1196-1205. doi:10.1093/molbev/mst030. PubMed: 23420840.
64. Wlasiuk G, Nachman MW (2010) Adaptation and constraint at Toll-like receptors in primates. *Mol Biol Evol* 27: 2172-2186. doi:10.1093/molbev/msq104. PubMed: 20410160.
65. Anisimova M, Yang Z (2007) Multiple hypothesis testing to detect lineages under positive selection that affects only a few sites. *Mol Biol Evol* 24: 1219-1228. doi:10.1093/molbev/msm042. PubMed: 17339634.
66. Zhang J, Nielsen R, Yang Z (2005) Evaluation of an improved branch-site likelihood method for detecting positive selection at the molecular level. *Mol Biol Evol* 22: 2472-2479. doi:10.1093/molbev/msi237. PubMed: 16107592.
67. Kosakovsky Pond SL, Murrell B, Fourment M, Frost SDW, Delport W et al. (2011) A random effects branch-site model for detecting episodic diversifying selection. *Mol Biol Evol*, 28: 3033-43. PubMed: 21670087.
68. Vasseur E, Patin E, Laval G, Pajon S, Fornarino S et al. (2011) The selective footprints of viral pressures at the human RIG-I-like receptor family. *Hum Mol Genet* 20: 4462-4474. doi:10.1093/hmg/ddr377. PubMed: 21865300.
69. Buckley KM, Rast JP (2012) Dynamic evolution of toll-like receptor multigene families in echinoderms. *Front Immunol* 3: 136. PubMed: 22679446.
70. Alcaide M, Edwards SV (2011) Molecular evolution of the toll-like receptor multigene family in birds. *Mol Biol Evol* 28: 1703-1715. doi:10.1093/molbev/msq351. PubMed: 21239391.
71. Barreiro LB, Ben-Ali M, Quach H, Laval G, Patin E et al. (2009) Evolutionary dynamics of human Toll-like receptors and their different contributions to host defense. *PLoS Genet* 5: e1000562. PubMed: 19609346.
72. Jann OC, Werling D, Chang JS, Haig D, Glass EJ (2008) Molecular evolution of bovine Toll-like receptor 2 suggests substitutions of functional relevance. *BMC Evol Biol* 8: 288. doi:10.1186/1471-2148-8-288. PubMed: 18937834.
73. Nakajima T, Ohtani H, Satta Y, Uno Y, Akari H et al. (2008) Natural selection in the TLR-related genes in the course of primate evolution. *Immunogenetics* 60: 727-735. doi:10.1007/s00251-008-0332-0. PubMed: 18810425.
74. Wlasiuk G, Khan S, Switzer WM, Nachman MW (2009) A history of recurrent positive selection at the toll-like receptor 5 in primates. *Mol Biol Evol* 26: 937-949. doi:10.1093/molbev/msp018. PubMed: 19179655.
75. Smith JM, Smith NH (1996) Synonymous nucleotide divergence: what is "saturation"? *Genetics* 142: 1033-1036. PubMed: 8849908.
76. Seo TK, Kishino H (2008) Synonymous substitutions substantially improve evolutionary inference from highly diverged proteins. *Syst Biol* 57: 367-377. doi:10.1080/10635150802158670. PubMed: 18570032.
77. Kawai T, Takahashi K, Sato S, Coban C, Kumar H et al. (2005) IPS-1, an adaptor triggering RIG-I- and Mda5-mediated type I interferon induction. *Nat Immunol* 6: 981-988. doi:10.1038/ni1243. PubMed: 16127453.
78. Takahashi K, Yoneyama M, Nishihori T, Hirai R, Kumeta H et al. (2008) Nonself RNA-sensing mechanism of RIG-I helicase and activation of antiviral immune responses. *Mol Cell* 29: 428-440. doi:10.1016/j.molcel.2007.11.028. PubMed: 18242112.
79. Fairman-Williams ME, Guenther UP, Jankowsky E (2010) SF1 and SF2 helicases: family matters. *Curr Opin Struct Biol* 20: 313-324. doi:10.1016/j.sbi.2010.03.011. PubMed: 20456941.
80. Cui S, Eisenächer K, Kirchhofer A, Brzózka K, Lammens A et al. (2008) The C-terminal regulatory domain is the RNA 5'-triphosphate sensor of RIG-I. *Mol Cell* 29: 169-179. doi:10.1016/j.molcel.2007.10.032. PubMed: 18243112.
81. Thompson JD, Higgins DG, Gibson TJ (1994) CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Res* 22: 4673-4680. doi:10.1093/nar/22.22.4673. PubMed: 7984417.
82. Hall T (1999) BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. *Nucleic Acids Symp Ser* 41: 95-98.
83. Posada D, Crandall KA (2002) The effect of recombination on the accuracy of phylogeny estimation. *J Mol Evol* 54: 396-402. PubMed: 11847565.
84. Posada D (2008) jModelTest: phylogenetic model averaging. *Mol Biol Evol* 25: 1253-1256. doi:10.1093/molbev/msn083. PubMed: 18397919.
85. Zwickl DJ (2006) Genetic algorithm approaches for the phylogenetic analysis of large biological sequence datasets under the maximum likelihood criterion [PhD Thesis]. Austin: University of Texas.

86. Nielsen R, Yang Z (1998) Likelihood models for detecting positively selected amino acid sites and applications to the HIV-1 envelope gene. *Genetics* 148: 929-936. PubMed: 9539414.
87. Yang Z, Nielsen R, Goldman N, Pedersen AM (2000) Codon-substitution models for heterogeneous selection pressure at amino acid sites. *Genetics* 155: 431-449. PubMed: 10790415.
88. Yang Z, Wong WS, Nielsen R (2005) Bayes empirical bayes inference of amino acid sites under positive selection. *Mol Biol Evol* 22: 1107-1118. doi:10.1093/molbev/msi097. PubMed: 15689528.