

RESEARCH ARTICLE

# Evolution of a Novel Antiviral Immune-Signaling Interaction by Partial-Gene Duplication

Bryan Korithoski<sup>1</sup>, Oralia Kolaczowski<sup>1</sup>, Krishanu Mukherjee<sup>2</sup>, Reema Kola<sup>1,3</sup>, Chandra Earl<sup>4</sup>, Bryan Kolaczowski<sup>1\*</sup>

**1** Department of Microbiology and Cell Science, Institute of Food and Agricultural Sciences, University of Florida, Gainesville, FL, United States of America, **2** Whitney Laboratory for Marine Bioscience, University of Florida, St. Augustine, FL, United States of America, **3** Department of Food Science and Human Nutrition, University of Florida, Gainesville, FL, United States of America, **4** Department of Biology, University of Florida, Gainesville, FL, United States of America

\* [bryank@ufl.edu](mailto:bryank@ufl.edu)



OPEN ACCESS

**Citation:** Korithoski B, Kolaczowski O, Mukherjee K, Kola R, Earl C, Kolaczowski B (2015) Evolution of a Novel Antiviral Immune-Signaling Interaction by Partial-Gene Duplication. *PLoS ONE* 10(9): e0137276. doi:10.1371/journal.pone.0137276

**Editor:** Pierre Boudinot, INRA, FRANCE

**Received:** March 2, 2015

**Accepted:** August 16, 2015

**Published:** September 10, 2015

**Copyright:** © 2015 Korithoski et al. 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 use, distribution, and reproduction in any medium, provided the original author and source are credited.

**Data Availability Statement:** All relevant data are within the paper and its Supporting Information files.

**Funding:** This research was supported by the National Institute of Allergy and Infectious Diseases of the National Institutes of Health under award number R21AI101571, grant number FLA-MCS-005262 from the USDA National Institute of Food and Agriculture, and by the University of Florida. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health or the University of Florida. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

## Abstract

The RIG-like receptors (RLRs) are related proteins that identify viral RNA in the cytoplasm and activate cellular immune responses, primarily through direct protein-protein interactions with the signal transducer, IPS1. Although it has been well established that the RLRs, RIG-I and MDA5, activate IPS1 through binding between the twin caspase activation and recruitment domains (CARDs) on the RLR and a homologous CARD on IPS1, it is less clear which specific RLR CARD(s) are required for this interaction, and almost nothing is known about how the RLR-IPS1 interaction evolved. In contrast to what has been observed in the presence of immune-modulating K63-linked polyubiquitin, here we show that—in the absence of ubiquitin—it is the first CARD domain of human RIG-I and MDA5 (CARD1) that binds directly to IPS1 CARD, and not the second (CARD2). Although the RLRs originated in the earliest animals, both the IPS1 gene and the twin-CARD domain architecture of RIG-I and MDA5 arose much later in the deuterostome lineage, probably through a series of tandem partial-gene duplication events facilitated by tight clustering of RLRs and IPS1 in the ancestral deuterostome genome. Functional differentiation of RIG-I CARD1 and CARD2 appears to have occurred early during this proliferation of RLR and related CARDs, potentially driven by adaptive coevolution between RIG-I CARD domains and IPS1 CARD. However, functional differentiation of MDA5 CARD1 and CARD2 occurred later. These results fit a general model in which duplications of protein-protein interaction domains into novel gene contexts could facilitate the expansion of signaling networks and suggest a potentially important role for functionally-linked gene clusters in generating novel immune-signaling pathways.

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

## Introduction

The evolution of novel molecular interactions can have profound implications for the organism, including changing how it develops, responds to environmental challenges and interacts with symbionts and pathogens. Understanding how new molecular interactions arise, establish themselves as distinct from existing interactions, and change over evolutionary time therefore provides a reductionist foundation from which to begin answering fundamental questions regarding the evolution of organismal complexity and novelty [1].

The RIG-like receptors (RLRs) RIG-I and MDA5 are cytoplasmic proteins that identify viral RNA and initiate cellular immune responses by interacting with the signal-transducing protein, IPS1 [2–4]. The RLR-IPS1 interaction initiates a series of signaling events that results in production of type I interferons and proinflammatory cytokines. RLR-IPS1 immune signaling is mediated by a direct physical interaction between twin N-terminal caspase activation and recruitment domains (CARDs) on the RLR and the N-terminal CARD of IPS1 [5, 6]. The third RLR, LGP2, lacks CARD signaling domains and appears to regulate antiviral immune responses through various mechanisms [7–9].

All three RLRs and IPS1 have been confidently identified in a wide variety of vertebrate genomes, suggesting that the RLR-IPS1 signaling pathway was present in the ancestral vertebrate. We have recently shown that full-length RLRs were present in the earliest metazoa [10], and IPS1 has been tentatively identified in the echinoderm, *Strongylocentrotus purpuratus*, suggesting that the RLR-IPS1 interaction may pre-date the vertebrate lineage [11]. Although the CARD domains of RLRs and IPS1 appear homologous [12], very little is known about how the RLR-IPS1 interaction evolved.

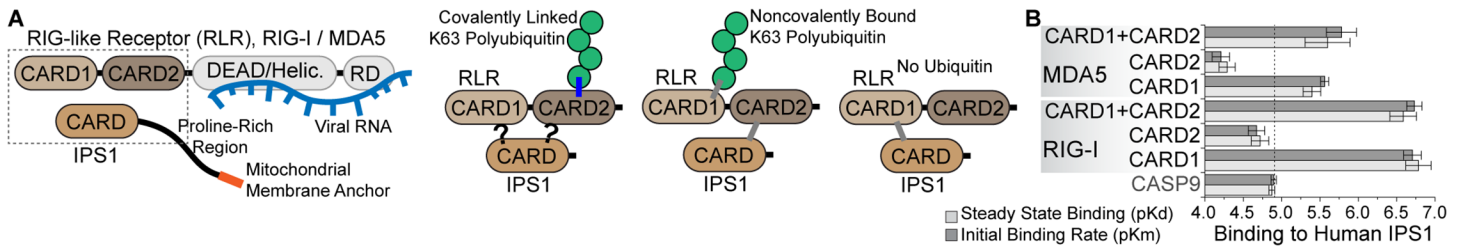
Here we use a combination of genomics, phylogenetics, structural modeling and molecular-functional characterization of ancestral and extant RLR and IPS1 CARD domains to determine when and how direct RLR-IPS1 interactions arose and initially diversified. In contrast to what has been observed in the presence of immune-modulating polyubiquitin chains [13], we find that the first CARD domains of human RIG-I and MDA5 bind directly to IPS1 CARD in the absence of ubiquitin. Evolutionary analysis suggests that RLR and IPS1 CARDs diversified in early deuterostomes, probably through a series of tandem, partial-gene duplication events. Functionally distinct RLR-IPS1 signaling pathways emerged coincident with early deuterostome CARD proliferation, potentially driven by adaptive coevolution between RLR and IPS1 CARDs. Although functionally distinct RIG-I CARD1 and CARD2 domains evolved early, functional specificity of MDA5 CARD domains appears to have arisen at a later time.

## Results and Discussion

### Human RLR CARD1s Interact Directly with IPS1 CARD when Ubiquitin Is Absent

Analyses of cellular immune-signaling in mammals have demonstrated convincingly that the twin N-terminal CARD domains of the RIG-like receptors (RLRs), RIG-I and MDA5, interact directly with the N-terminal CARD domain of IPS1 to initiate antiviral immune responses [5, 6]. Further studies have suggested that K63-linked polyubiquitin chains—either covalently linked to CARD2 or bound noncovalently—may potentiate RLR-based immune signaling [14, 15]. Recent structural and mutational studies suggest that—in the presence of noncovalently bound polyubiquitin—RIG-I CARD2 interacts directly with IPS1's CARD domain, whereas RIG-I CARD1 helps to form a homotetrameric RLR complex [13, 16] (Fig 1A).

We expressed and purified human RIG-I CARD1 (amino acid positions 1–87), CARD2 (88–172), and CARD1+2 domains as well as MDA5 CARD1 (1–96) CARD2 (97–204) and



**Fig 1. In the absence of ubiquitin, the RIG-like receptors (RLRs), RIG-I and MDA5, bind their signaling partner, IPS1, via a direct interaction between the first RLR CARD domain (CARD1) and the CARD domain of IPS1.** (A) After binding viral RNA, RIG-I and MDA5 interact directly with IPS1 via CARD-CARD interactions. K63 Polyubiquitin chains—either covalently linked or noncovalently bound to RLR CARDS—can potentiate RLR-IPS1 signaling. In the case of noncovalent polyubiquitin binding, studies suggest RIG-I CARD2 interacts with IPS1 (bottom middle). Our results suggest that, in the absence of ubiquitin, RLR CARD1 interacts with IPS1 (bottom right). (B) We measured the kinetics of human RIG-I and MDA5 CARD1, CARD2 and CARD1+2 domains bound to IPS1 CARD in vitro (see [Materials and Methods](#)) (S1 Fig). We plot the mean and standard error in  $-\log$ -transformed steady-state dissociation ( $pK_d$ ) and initial binding rate ( $pK_m$ ), with longer bars indicating tighter binding (S2 Fig). Human CASP9 CARD was used as a negative control to indicate an approximate level of nonspecific binding (dotted line); CASP9 is the CARD domain most closely related to RLR/IPS1 CARDS (Fig 3), and direct interactions between RLRs and CASP9 have not been reported.

doi:10.1371/journal.pone.0137276.g001

CARD1+2 domains and measured their binding kinetics to human IPS1 CARD (1–92) in vitro (S1 Fig). In the absence of polyubiquitin, we found that RIG-I and MDA5 CARD1 and CARD1+2 domains bound to IPS1 CARD with high affinity, whereas CARD2 domains did not (Fig 1B and S2 Fig). Human RIG-I CARD1 bound IPS1 with the strongest affinity (dissociation constant,  $K_d = 3.31e^{-8}$  M; initial binding rate,  $K_m = 3.89e^{-8}$  M), ~70-times stronger than human caspase9 CARD, which we used as a negative control ( $p < 0.006$ ). Human MDA5 CARD1 also bound IPS1 better than the negative control (MDA5 CARD1  $K_d = 8.13e^{-7}$  M,  $K_m = 5.50e^{-7}$  M,  $p < 0.013$ ), but not as strongly as RIG-I CARD1 ( $p < 0.004$ ). Interestingly, we found that neither RIG-I nor MDA5 CARD2 domains bound IPS1 CARD more tightly than the negative control ( $p > 0.16$ ). However, the presence of CARD2 domains did not interfere with RLR-IPS1 binding by CARD1, as both RIG-I and MDA5 CARD1+2 domains bound IPS1 with the same affinity as their respective CARD1 domains ( $p > 0.37$ ).

Our analyses used flag-tagged IPS1 proteins to facilitate binding measurements (see [Materials and Methods](#)). Although unlikely, the presence of the flag polypeptide could alter IPS1's binding kinetics in vitro. However, similar results were obtained using either a maltose-binding-protein IPS1 tag (S3 Fig) or flag-tagged RLR proteins bound to untagged IPS1 tag (S3 Fig), arguing against this interpretation.

We divided RLR CARD1+CARD2 sequences into CARD1 and CARD2 using functional domains annotated in the Pfam, CDD and SMART databases (see [Materials and Methods](#)). Dividing CARD1+2 into individual functional domains that do not precisely map to structural domain boundaries could introduce binding artifacts by altering the folding of individual functional domains, compared to the CARD1+2 multidomain construct. To examine this possibility, we measured IPS1 CARD binding of structurally-determined RIG-I CARD1 (amino acid positions 1–93) and CARD2 (102–188) as well as MDA5 CARD1 (1–101) and CARD2 (113–204) constructs (S1 Fig). We found no difference in the pattern of which RLR CARDS bound IPS1, whether CARD boundaries were identified using sequence or structural data. Structurally-determined RIG-I and MDA5 CARD1 (excluding the CARD1-CARD2 linker) bound IPS1 better than the negative control ( $p < 0.021$  for RIG-I;  $p < 0.034$  for MDA5), while CARD2 domains (excluding the interdomain linker) did not ( $p > 0.41$ ) (S5 Fig), suggesting that artifacts related to domain boundary identification or inclusion/exclusion of the CARD1-CARD2 linker are unlikely to have a major impact on our results.

## Protein-Protein Docking Predicts the RIG-I CARD1 –IPS1 Interface

In the presence of ubiquitin, structural and mutational studies suggest that RIG-I CARD2 interacts with IPS1 CARD through a variety of interfaces to form a large-scale helical complex [13]. Although no structure of an RLR CARD bound to IPS1 in the absence of ubiquitin has been determined, early studies of IPS1 CARD crystalized in isolation suggested that IPS1 might bind its signaling partners through an  $\alpha 1$ - $\alpha 3$ - $\alpha 4$  interface [12]. In fact, the  $\alpha 1$ - $\alpha 3$ - $\alpha 4$  surface of IPS1 was observed to interact with the  $\alpha 2$ - $\alpha 3$  surface of RIG-I CARD2 in the helical complex [13].

We predicted the structural interfaces between human RLR CARD1s and IPS1 CARD using protein-protein docking (see [Materials and Methods](#)). We found that a novel antiparallel  $\alpha 1$ - $\alpha 3$ - $\alpha 4$  binding interface between RIG-I CARD1 and IPS1 CARD was consistently recovered in our docking analyses and was supported by additional manual examination and mutation experiments (Fig 2). Four of the top 5 (7 of the top 10) RIG-I–IPS1 complexes predicted by ClusPro [17] contained the antiparallel  $\alpha 1$ - $\alpha 3$ - $\alpha 4$  arrangement, and all of the top 20 predicted complexes utilized the  $\alpha 1$ - $\alpha 3$ - $\alpha 4$  surface of IPS1 (Fig 2A). Using Dot2 [18] as an alternative docking protocol, all of the top 5 RIG-I–IPS1 complexes had the antiparallel  $\alpha 1$ - $\alpha 3$ - $\alpha 4$  interface, as did 9 of the top 10 and 19 of the top 20 (Fig 2B). Manual examination of the antiparallel  $\alpha 1$ - $\alpha 3$ - $\alpha 4$  RIG-I–IPS1 complex suggests a high degree of shape and electrostatic complementarity (Fig 2C), which has been observed for other CARD-CARD complexes [19].

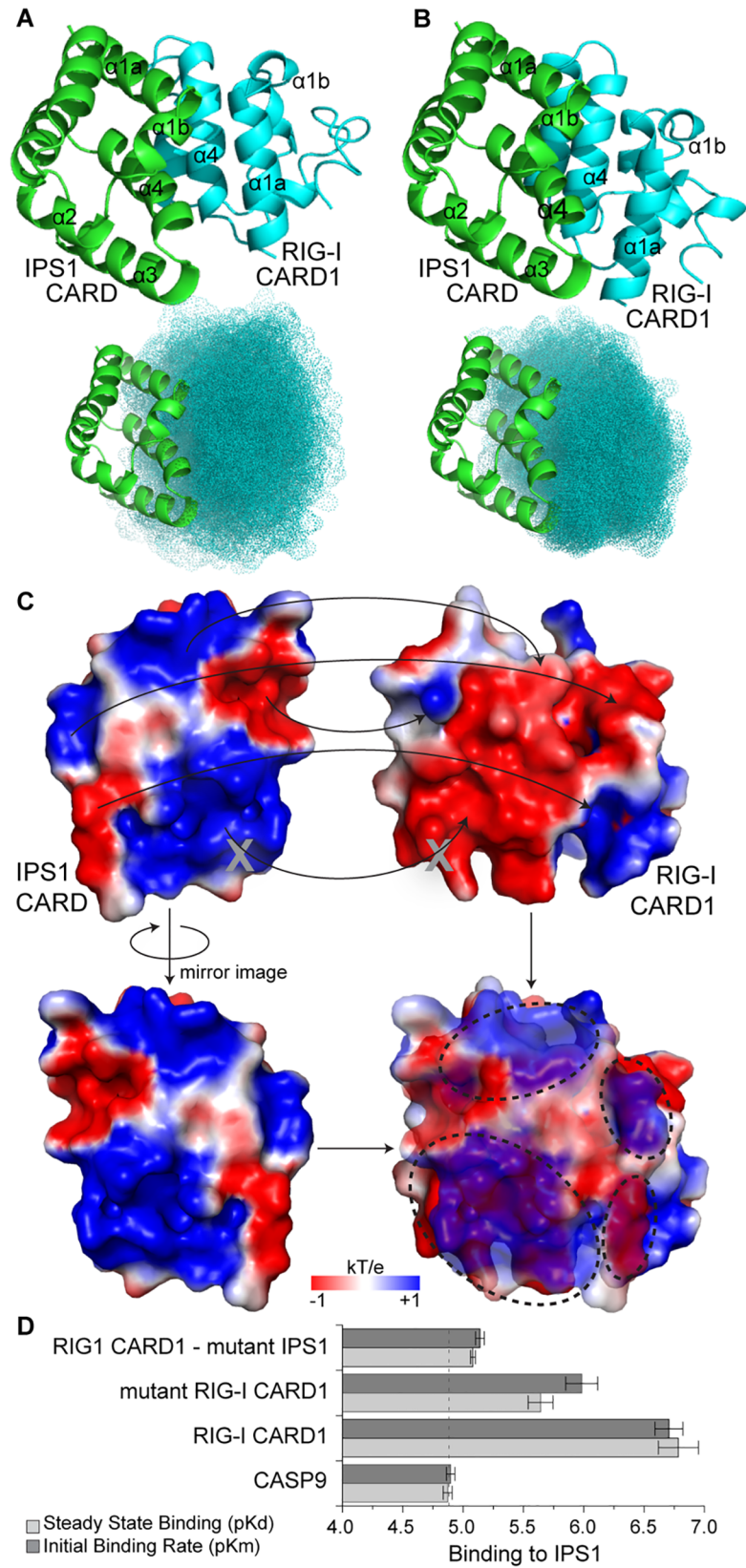
We identified a pair of residue clusters (RIG-I CARD1 Glu66,67 and IPS1 Arg66,67) that were predicted to interact in the RIG-I–IPS1 interface and were located on a loop between helices 4 and 5, minimizing the potential impact of mutations at these residues on overall protein stability and structure (Fig 2C). Mutating RIG-I CARD1 Glu66,67 to Arg disrupted the protein's binding to wild-type IPS1 (10-fold increase in Kd, 6.9-fold increase in Km,  $p < 0.033$ ) (Fig 2D and S6 Fig). Mutating IPS1 Arg66,67 to Glu had an even larger effect on RIG-I–IPS1 binding, increasing Kd by 34.6-fold and Km by 43.8-fold ( $p < 0.004$ ) (Fig 2D and S6 Fig). Similar mutations did not strongly affect RIG-I–IPS1 interactions in the presence of ubiquitin chains [16], suggesting that ubiquitin likely alters the RIG-I–IPS1 interface.

The MDA5-IPS1 complex was less consistently recovered. Although the top-scoring ClusPro complex had the same antiparallel  $\alpha 1$ - $\alpha 3$ - $\alpha 4$  interface that we observed for RIG-I, only 3 of the top 5 (7 of the top 10) complexes had this arrangement (S7 Fig). MDA5-IPS1 docking using the alternative Dot2 protocol only recovered a similar antiparallel  $\alpha 1$ - $\alpha 3$ - $\alpha 4$  interface in 12 of the top 20 complexes (S7 Fig). We therefore consider prediction of the MDA5-IPS1 complex ambiguous.

Protein-protein docking is a challenging problem, and error rates have historically been notoriously high [20]. However, recent advances have made dramatic improvements in accuracy, particularly for cases such as RIG-I–IPS1 binding, in which the unbound and bound structures are expected to be highly similar [13, 16, 21]. Although the results of mutation experiments were consistent with the predicted antiparallel  $\alpha 1$ - $\alpha 3$ - $\alpha 4$  RIG-I–IPS1 complex, detailed structural studies will be required to determine the precise RIG-I–IPS1 interface when ubiquitin is not present.

## RLR and IPS1 CARD Domains Proliferated in Early Deuterostomes

We inferred the phylogeny of RLR and IPS1 CARD domains by maximum likelihood (ML) and Bayesian analyses, using CARD sequences from caspases 9 and 2 as outgroups. The resulting consensus tree (Fig 3A) recovered individual deuterostome IPS1, MDA5 CARD1, MDA5 CARD2, RIG-I CARD1 and RIG-I CARD2 clades with  $> 90\%$  ML SH-like aLRT and  $> 95\%$  Bayesian posterior probability, suggesting that RLR and IPS1 CARD domains diversified in



**Fig 2. Structure-based protein-protein docking suggests RIG-I CARD1 interacts with IPS1 CARD through a novel antiparallel  $\alpha 1$ - $\alpha 3$ - $\alpha 4$  interface.** We used ClusPro (A) and Dot2 (B) to predict the most likely structural orientation of human RIG-I CARD1 bound to IPS1 (see [Materials and Methods](#)). We show the best-scoring complex (top panels) as well as the 20 top-scoring orientations (bottom panels) from each analysis. (C) We calculated the surface electrostatic charge (kT/e) of human RIG-I CARD1 and IPS1 CARD (see [Materials and Methods](#)). Top panel displays IPS1 (left) and RIG-I CARD1 (right) in the same antiparallel  $\alpha 1$ - $\alpha 3$ - $\alpha 4$  orientation as A and B. Arrows indicate regions of inferred shape and electrostatic complementarity. Gray Xs indicate locations of mutations examined in D. In the bottom panel, IPS1 is shown in mirror image and made 35% transparent; regions of electrostatic complementarity appear purple in the overlay (also highlighted by dotted ovals). (D) We created mutant RIG-I CARD1 (Glu66,67—Arg) and IPS1 (Arg66,67—Glu) and measured the affinities with which mutant proteins bind their wild-type partners. We plot the mean and standard error—log steady-state dissociation (pKd) and initial binding rate (pKm) of each interaction ([S6 Fig](#)). Wild-type RIG-I CARD1 and CASP9 interacting with wild-type IPS1 were used as positive and negative controls, respectively ([Fig 1B](#)).

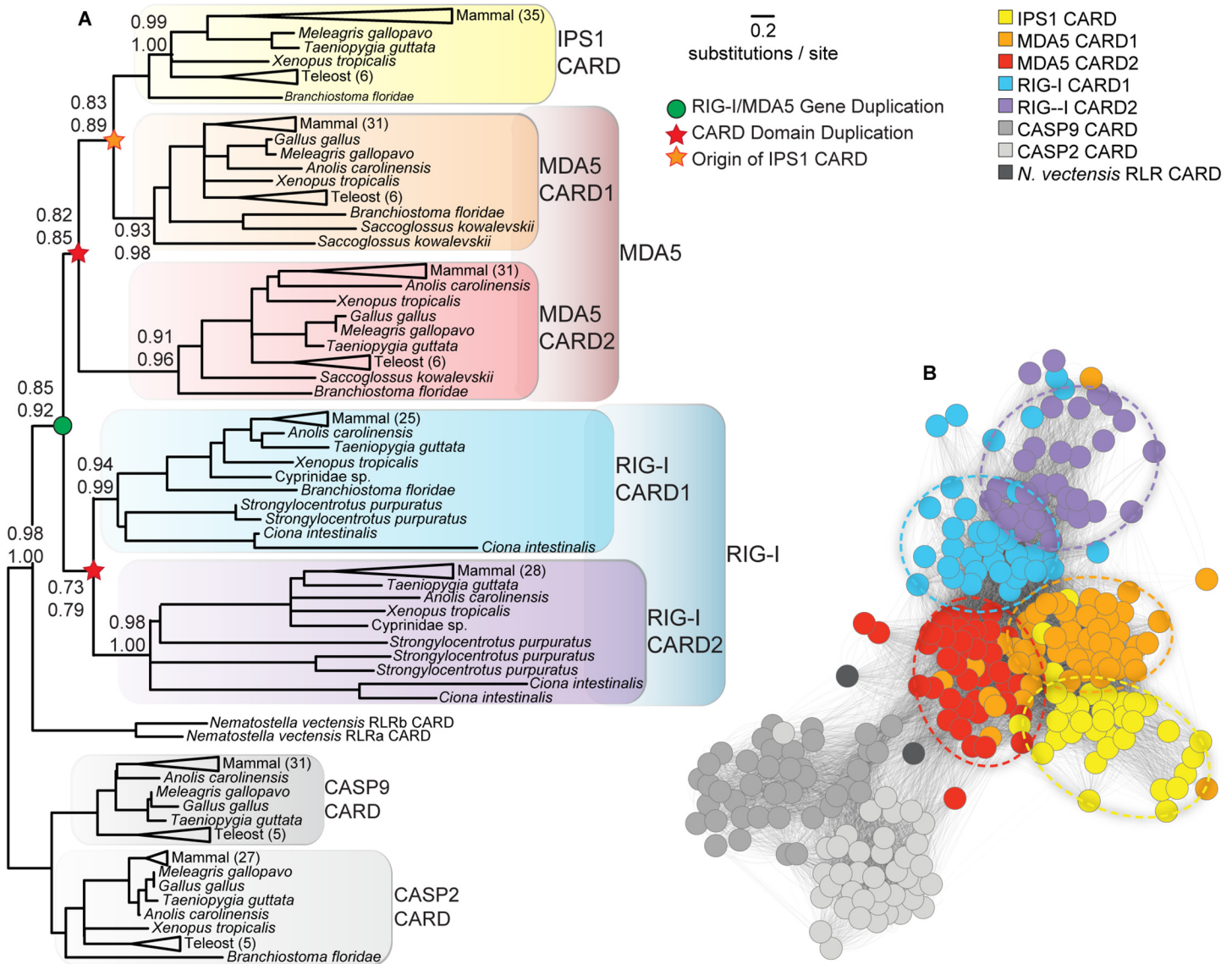
doi:10.1371/journal.pone.0137276.g002

deuterostomes before the vertebrate lineage. Both the ML and Bayesian phylogenies inferred that RIG-I and MDA5 CARD domains duplicated independently following the RIG-I/MDA5 gene duplication, suggesting that the CARD1+CARD2 domain architecture arose independently in RIG-I and MDA5, with IPS1 originating from a later duplication of MDA5 CARD1. The alternative phylogeny, in which CARD domain duplication precedes the RIG-I/MDA5 gene duplication ([S8 Fig](#)), was rejected as a plausible explanation of our sequence data, with marginal significance (SH-test  $p = 0.02$ ). However, the statistical support in favor of the maximum-likelihood branching pattern is weak. Removal of fast-evolving non-vertebrate sequences improved inferred support for the consensus branching pattern but did not fully resolve the tree ([S9 Fig](#)).

RLR and IPS1 CARD domains are small and fast-evolving, making it difficult to resolve the branching pattern with high confidence using traditional techniques. We therefore used a variety of alternative approaches to more thoroughly assess the reliability of our phylogenetic inference. First, we used a sequence-similarity network to cluster RLR and IPS1 CARD domains independent of phylogenetic inference [[22](#), [23](#)]. The results of this analysis are largely consistent with the consensus phylogeny ([Fig 3B](#)). The caspase 9–2 sequences form a well-defined group that is highly diverged from RLR and IPS1 CARDS, with *Nematostella* RLR CARDS falling between the caspases and the other RLR-IPS1s. The RIG-I CARD domains cluster together in the network, and the MDA5 CARDS form a tight cluster with IPS1 CARD domains, reinforcing the branching pattern observed from phylogenetic analysis.

Second, we constructed hidden Markov models (HMMs) from independent alignments of deuterostome RIG-I—CARD1, RIG-I—CARD2, MDA5-CARD1, MDA5-CARD2 and IPS1-CARD sequences [[24](#), [25](#)], each of which formed a monophyletic clade in the consensus tree with strong support ([Fig 3A](#)). We calculated pairwise distances between the HMMs and inferred the neighbor-joining tree from these distances [[26](#), [27](#)], which recovered the same branching pattern as the ML phylogeny based on protein sequence data ([S10 Fig](#)). Although these analyses cannot completely resolve the branching pattern explaining RLR-IPS1 CARD proliferation, that the same branching pattern is consistently recovered using a variety of different inference methodologies argues against strong methodological bias.

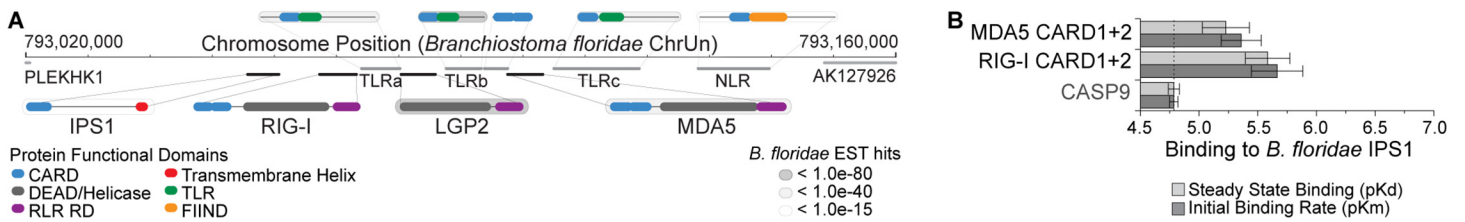
Although caspase outgroups rooted our RLR-IPS1 CARD phylogeny with strong support, the large evolutionary distance between caspase and RLR-IPS1 sequences invokes the possible specter of long-branch attraction. To address this possibility, we used gene/species tree reconciliation to identify the most likely RLR-IPS1 rooting in the absence of outgroup data [[28–30](#)]. The resulting rooting is consistent with that obtained using caspase outgroups ([S11 Fig](#)). Although this analysis cannot completely rule out long-branch attraction, it does argue against such an interpretation.



**Fig 3. RIG-like receptor (RLR) and IPS1 CARD domains diversified in early deuterostomes.** (A) We inferred the consensus phylogeny of RLR and IPS1 CARD domains using maximum likelihood and Bayesian inference (see [Materials and Methods](#)). We report statistical confidence supporting key nodes on the phylogeny using maximum-likelihood SH-like aLRT (top) and Bayesian posterior probability (bottom); branch lengths are scaled to substitutions/site. Inferred gene and CARD duplication events are indicated by circles and stars, respectively. (B) We calculated BLAST e-values between all pairs of RLR and IPS1 CARD protein sequences and visualized sequence similarity using a force-directed network layout implemented in Cytoscape v2.8 [22, 23]. Nodes in the network represent individual sequences and are colored by CARD type. Edges indicate pairwise e-value calculations; e-values  $> 10^{-3}$  are not shown, to improve clarity. Nodes close to one another are more similar, based on e-value calculations, than nodes far away from one another. The central cluster of each RLR and IPS1 CARD domain is indicated by a dotted oval.

doi:10.1371/journal.pone.0137276.g003

It has been suggested that gene conversion events can homogenize the sequences of duplicate genes, potentially misleading evolutionary analyses [31]. We used pairwise comparisons of closely-related mammalian RLR and IPS1 genes to investigate the extent of gene conversion among RLR and IPS1 CARD domains [32]. We found very little evidence for RLR-IPS1 gene conversion (S1 Table). Only two regions of mammalian MDA5 CARD2 were identified as having undergone potential gene conversion events ( $p < 0.03$ ), and neither of these was significant after Bonferroni correction for multiple tests ( $p > 0.4$ ).



**Fig 4. Deuterostome RIG-like receptors (RLRs) and IPS1 cluster with other potential CARD-signaling immune receptors in the *Branchiostoma floridae* genome.** (A) We plot the physical location of each identified RLR and IPS1 gene along the assembled *B. floridae* chromosome (bottom). Other genes in the cluster are shown along the top. Functional domains are indicated by colors. Genes with significant BLAST hits to the *B. floridae* expressed sequence tag (EST) database are shaded according to e-value. (B) We measured the kinetics of *B. floridae* RIG-I CARD1+2 and MDA5 CARD1+2 domains binding to *B. floridae* IPS1 CARD in vitro (see [Materials and Methods](#)) (S1 Fig). We plot the mean and standard error of  $-\log$ -transformed steady-state dissociation (pKd) and initial binding rate (pKm), with human CASP9 used as a negative control. Dotted vertical line indicates approximate level of non-specific binding, with longer bars indicating tighter binding (S12 Fig).

doi:10.1371/journal.pone.0137276.g004

We conclude that the lack of phylogenetic resolution in the RLR-IPS1 CARD tree likely reflects the ambiguous information present in available sequence data due to fast-evolving, short sequences. Although we could not detect any evidence of methodological bias, we remain cautious in our interpretation of the precise branching pattern explaining the diversification of RLR and IPS1 CARD domains. However, the conclusion that RLR-IPS1 CARDs proliferated in deuterostomes prior to the vertebrate origin appears robust.

### RLR and IPS1 Genes Are Tightly Clustered in *Branchiostoma floridae*

Our phylogenetic analysis suggests that RLR and IPS1 CARD domains proliferated in early deuterostomes (Fig 3). Insight into a plausible mechanism for generating this series of partial-gene duplications comes from examining where RLR and IPS1 genes are located in the *Branchiostoma floridae* genome, a cephalochordate that diverged from other deuterostomes shortly before the vertebrate lineage [33].

We found that all three RLRs (RIG-I, MDA5 and LGP2), as well as IPS1, three possible toll-like receptors and one possible nod-like receptor are all located within a short 100-kb region of the assembled *B. floridae* chromosome (Fig 4A). Inferred coding sequences encoded proteins having domain architectures equivalent to those found in human proteins, including the twin N-terminal CARD domains of RIG-I and MDA5. The *B. floridae* IPS1 gene encoded an N-terminal CARD domain as well as an unstructured, proline-rich region and a predicted C-terminal transmembrane helix, all of which is characteristic of vertebrate IPS1. We found near-exact matches to some of our inferred RLR genes in expressed sequence tag (EST) databases, arguing that these genes are likely functional in *B. floridae* and not pseudogenes or gene modeling errors (Fig 4A).

We measured the binding of *B. floridae* RIG-I CARD1+2 and MDA5 CARD1+2 domains to the *B. floridae* IPS1 CARD domain and found that RIG-I bound IPS1 significantly more tightly than a negative control (.54-fold lower Kd, 6.1-fold lower Km,  $p < 0.044$ ) (Fig 4B and S12 Fig). The difference in binding of *B. floridae* MDA5 CARD1+2 vs. the negative control was not significant ( $p > 0.07$ ). However, these results do suggest that *B. floridae* has functional RLR genes that are tightly clustered in its genome—along with IPS1 and other immune-system genes—and at least some of these RLRs are likely to signal through IPS1.

Interestingly, all the genes present in this cluster encode proteins with likely immune-signaling functions, and all of them contain homologous CARD domains expected to be involved in CARD-CARD signaling interactions. This suggests a plausible mechanism by which a series of tandem, partial-gene duplication events may have given rise to the diversity of RLR-IPS1



CARD sequences observed in deuterostomes. A dual-CARD open reading frame just upstream of *B. floridae* MDA5 that is 98% identical to the MDA5 CARD1+2 protein sequence further reinforces this conclusion and suggests that tandem partial-gene duplications may be ongoing.

Although most of the sequenced genomes of other pre-vertebrate deuterostomes have not been assembled to chromosomes, we did observe some evidence for possible genomic clustering of at least some of the RLRs in *Ciona intestinalis* and other deuterostome genomic scaffolds (S13 Fig). Additionally, recombination is expected to break down gene clusters over large evolutionary timescales, whereas the de novo generation of functionally-linked gene clusters is much less likely [34]. Together, these results suggest that RLRs and IPS1 were probably clustered in the ancestral deuterostome.

Tandem gene and partial-gene duplications—facilitated by unequal crossing over and other mechanisms—have been widely observed in animal and plant genomes [35]. Our results suggest that the tight clustering of RLR and IPS1 genes in the ancestral deuterostome genome may have facilitated CARD domain proliferation by bringing similar sequences into close physical proximity, allowing unequal crossing over or other mechanisms to generate multiple rounds of tandem partial-gene duplications.

## Coevolution between RLRs and IPS1 Differentiated Early Deuterostome RLR-IPS1 Signaling Interactions

Examination of RLR-IPS1 binding in pre-vertebrate deuterostomes can suggest possible ancestral molecular functions but cannot explicitly test these hypotheses. To directly measure the interactions of early deuterostome proteins, we reconstructed ancestral sequences (S1 Fig) using phylogenetic techniques that explicitly incorporate uncertainty about the tree topology (see Materials and Methods). Consistent with previous findings that ancestral reconstruction is highly robust to phylogenetic uncertainty [36], we found that ancestral deuterostome RLR and IPS1 CARD domains were reconstructed with high confidence, even though the tree topology was ambiguous (S14 Fig). Across all ancestral sequence reconstructions, over 55% of residues were reconstructed with posterior probability > 0.95; over 60% were reconstructed with posterior probability > 0.90, and over 70% were reconstructed with posterior probability > 0.80. Furthermore, only 91/960 (<9.5%) reconstructed positions had an alternative reconstruction with posterior probability > 0.3, and the majority of these (93%) were in the same general biochemical class as the maximum-likelihood reconstructed residue. Together, these results suggest that ancestral sequences were reconstructed with high confidence and low ambiguity, despite the ambiguity in the underlying phylogeny.

High statistical confidence in reconstructed ancestral sequences does not necessarily imply that the sequences are accurate. To better characterize the expected error rates of reconstructed ancestral sequences, we simulated protein sequence data along the maximum-likelihood phylogeny using the best-fit evolutionary model and ML parameter estimates. We found that ancestral reconstruction error rates under this scenario were < 0.05 for all nodes examined in this study (< 6 errors/sequence) (S14 Fig). When amino-acid residues with similar biochemical properties were considered equivalent, error rates were reduced to < 0.02 (S14 Fig).

When the assumed evolutionary model differs markedly from the actual molecular-evolutionary process, phylogenetic analyses such as ancestral sequence reconstruction can produce unreliable results. Recent studies have suggested that changes in state frequencies across the phylogeny can lead to errors in ancestral sequence reconstruction [37]. Although we did observe differences in amino-acid residue patterns across RLR and IPS1 CARD sequences (S10 Fig), none of the amino-acid frequency distributions differed significantly from the frequency distribution across all sequences (chi-squared  $p > 0.74$ ), suggesting that amino-acid frequencies

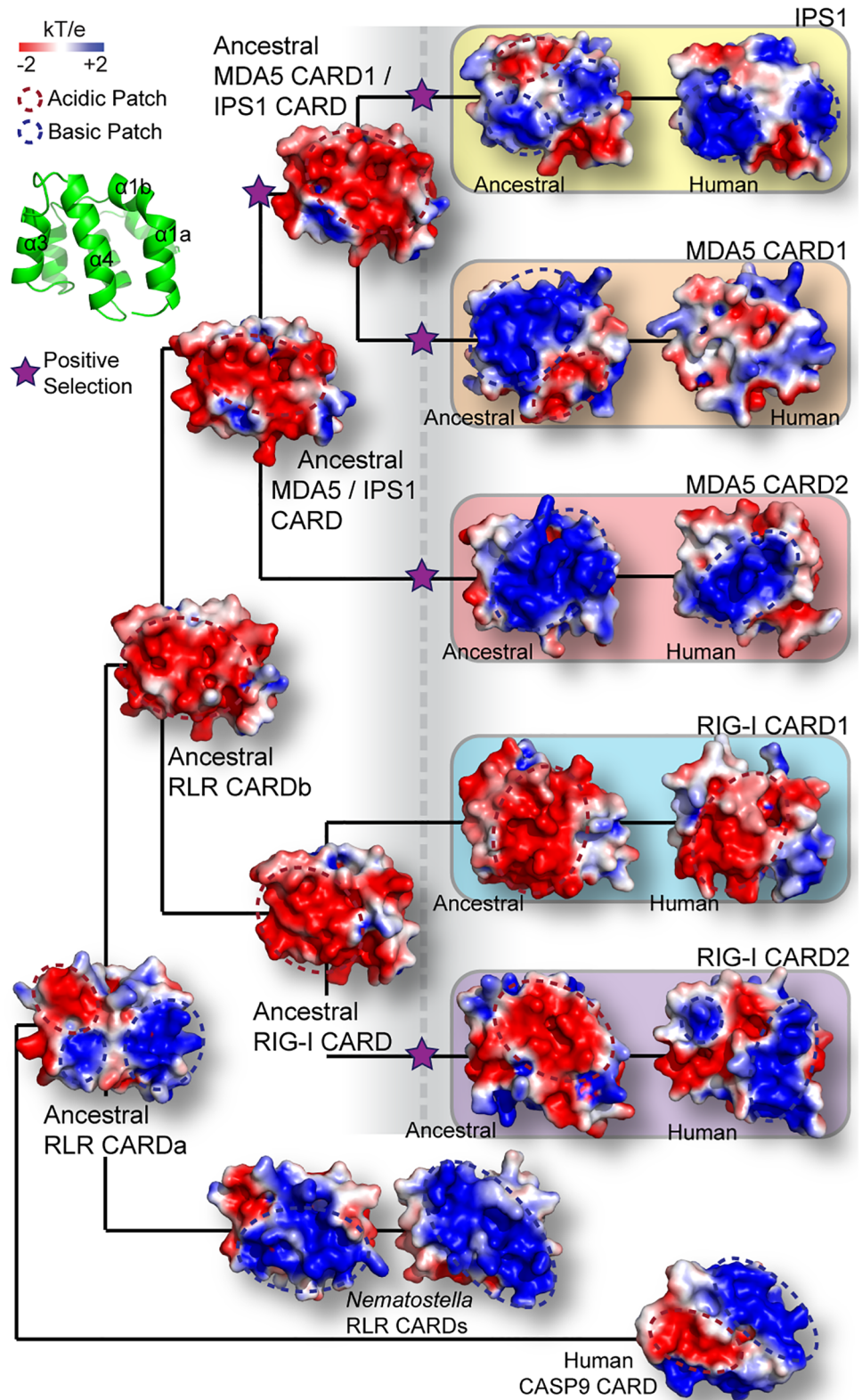
are roughly stationary across the tree (S14 Fig). Although it is impossible to completely rule out errors due to model misspecification, we observed no evidence for strong model violations or other artifacts that might affect ancestral sequence reconstruction.

Structural modeling of ancestral RLR and IPS1 CARD domains identified large changes in electrostatic charge across the surfaces of these molecules (Fig 5). Interestingly, major evolutionary shifts in electrostatic charge distribution across ancestral CARD domains coincided with inferences of significant protein-coding adaptation on those branches of the phylogeny, suggesting a model in which positive selection may have driven structural changes in RLR and IPS1 CARDS early in deuterostome evolution (Fig 5). Specifically, the predicted  $\alpha 1$ - $\alpha 3$ - $\alpha 4$  RLR-IPS1 interaction interface was inferred as strongly acidic in all ancestral CARD domains before the deuterostome proliferation (ancRLR CARDb, ancRIG-I CARD, ancMDA5/IPS1 CARD, and ancMDA5 CARD1 / IPS1 CARD) (Fig 5). This acidic  $\alpha 1$ - $\alpha 3$ - $\alpha 4$  surface remained largely acidic in the ancestral and human RIG-I CARD1 domains but appears to have acquired large basic patches in other RLR and IPS1 CARDS, after the deuterostome proliferation (boxed ancestral and human CARDS) (Fig 5). On every branch across which we observed the acquisition of basic patches on the  $\alpha 1$ - $\alpha 3$ - $\alpha 4$  surface, we also inferred protein-coding adaptation by nonsynonymous/synonymous substitution analysis (see [Materials and Methods](#)) (Fig 5). That the distribution of basic patches across the  $\alpha 1$ - $\alpha 3$ - $\alpha 4$  surface differs markedly among different CARD domains reinforces the phylogenetic conclusion that these basic patches were independently acquired.

Although nonsynonymous/synonymous substitution comparisons are widely used to characterize protein-coding adaptation and are generally considered robust [38–41], concerns have been raised that such analyses may suffer from high false-positive error rates under some conditions [42, 43]. We examined potential false-positive error rates by simulating codon sequence data along the maximum-likelihood RLR-IPS1 CARD phylogeny under three different empirically-derived evolutionary models (S15 Fig). False-positive error rates were highest when sequences evolved neutrally across the entire tree but were never greater than 0.05. Error rates were lower (mean = 0.009, SE = 0.0044) when sequences evolved under an empirical model with all parameters estimated from our sequence data and intermediate (mean = 0.024, SE = 0.006) when sequences evolved under an empirical model but were released from selective constraint on the branch being tested. That false-positive error rates were low across a variety of scenarios suggests that our analysis of protein-coding adaptation is likely reliable, although it is impossible to completely rule out potential model violations.

As CARD-CARD signaling interactions are known to be facilitated by surface charge complementarity [19], we hypothesized that the structural changes we observed were responsible for establishing the specific CARD-CARD interactions facilitating ancestral RLR-IPS1 binding in deuterostomes.

To examine this hypothesis, we expressed and purified ancestral RLR and IPS1 CARD domains and measured their binding affinities *in vitro*. Binding to human caspase 9 CARD was used as a negative control. Examining the initial establishment of RLR-IPS1 CARD-CARD interactions is complicated, as IPS1 CARD appears to have arisen via the duplication of MDA5 CARD1 into a new gene context, and the precise branching pattern by which RLR-IPS1 CARD domains proliferated cannot be unambiguously determined (Figs 3 and 5). We focused on characterizing the potential ancestral RLR-IPS1 CARD-CARD interactions before the deuterostome proliferation (dotted vertical line) (Fig 5), at which time each RLR had a single CARD domain (ancRIG-I CARD and ancMDA5/IPS1 CARD) (Fig 5), and the ancestral MDA5 CARD1 / IPS1 CARD was present as a potential interaction partner. We compared results using these pre-deuterostome-proliferation sequences to results obtained using ancestral RLR and IPS1 CARD sequences reconstructed after the deuterostome proliferation to determine



**Fig 5. RIG-like receptor (RLR) and IPS1 CARD domains structurally diverged in early deuterostomes, and this structural divergence was associated with protein-coding adaptation.** We inferred the 3-dimensional structures of ancestral RLR and IPS1 CARD domains at key points on the RLR-IPS1 CARD

phylogeny (see [Materials and Methods](#)) (Fig 3). We plot the electrostatic surface potential (kT/e) across the  $\alpha 1$ - $\alpha 3$ - $\alpha 4$  surface of each CARD domain, with large acidic and basic patches indicated by dotted red and blue ovals, respectively. We additionally used phylogenetic techniques to infer the presence of adaptive protein-coding substitutions on each branch of the phylogeny (see [Materials and Methods](#)). Stars indicate significant support for protein-coding adaptation on the indicated branch ( $p < 0.05$  after correcting for multiple tests). Large dotted vertical line indicates the approximate time of RLR-IPS1 CARD proliferation in early deuterostomes.

doi:10.1371/journal.pone.0137276.g005

how initial RLR-IPS1 CARD-CARD interactions may have established themselves as these domains proliferated.

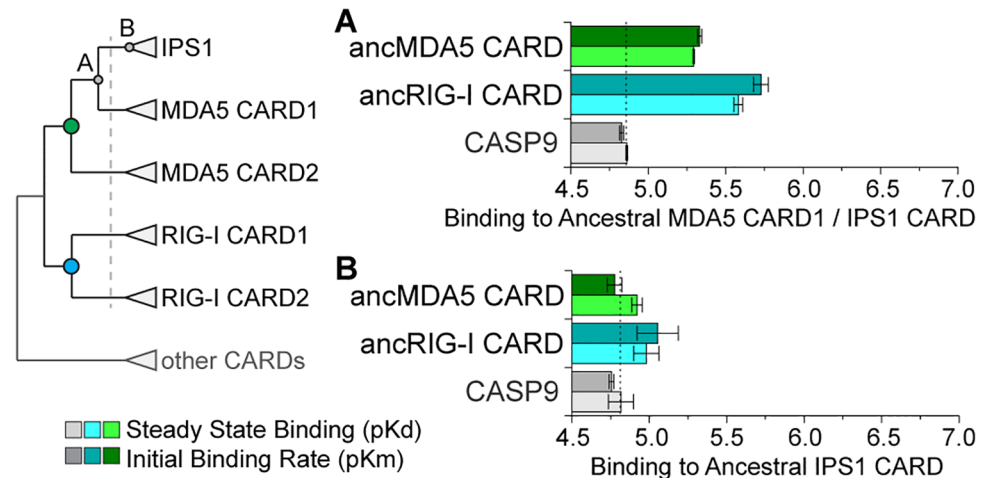
We found that, prior to the deuterostome proliferation, ancestral RIG-I and MDA5 CARD domains (ancRIG-I CARD and ancMDA5/IPS1 CARD) (Fig 5) bound the ancestral MDA5 CARD1 / IPS1 CARD with high affinity ( $p < 0.0016$ ) (Fig 6A and S16 Fig), suggesting that, prior to RLR-IPS1 CARD proliferation, the potential for an ancestral RLR CARD-IPS1 CARD interaction existed. Partial-gene duplication of MDA5 CARD1 into an IPS1-like architecture could have recruited this potential to establish a novel protein-protein signaling interaction without requiring any changes in protein sequence.

However, neither of the ancestral RLR CARDS before the deuterostome proliferation bound the ancestral IPS1 CARD after it diverged from MDA5 CARD1 (ancIPS1 CARD) (Fig 5), suggesting that, after the divergence of IPS1 from the RLRs, substitutions accrued in IPS1 CARD that prevented it from binding to its potential ancestral signaling partners ( $p > 0.15$ ) (Fig 6B and S16 Fig).

After the deuterostome proliferation, we found that compensatory substitutions occurred in RIG-I and MDA5 CARD domains that differentially re-established their ability to bind the diverged IPS1 CARD (Fig 7). Interestingly, all the diverged RIG-I and MDA5 CARD domains bound the ancestral MDA5 CARD1 / IPS1 CARD better than the negative control, suggesting that compensatory substitutions in RLR CARD domains did not interfere with their ability to bind their potential pre-deuterostome-proliferation partner ( $p < 0.018$ ) (Fig 7A and S17 Fig). However, after the deuterostome proliferation, only the ancestral RIG-I CARD1 and combined MDA5 CARD1+2 bound the diverged IPS1 CARD better than the control ( $p < 0.003$  for RIG-I CARD1,  $p < 0.04$  for MDA5 CARD1+2) (Fig 7B and S18 Fig).

In the case of the ancestral deuterostome RIG-I, we observed the same functional differentiation that we observed in human RIG-I (Fig 1B). The ancestral RIG-I CARD1 bound ancestral IPS1 CARD ( $K_d = 5.89e^{-6}$  M,  $K_m = 1.15e^{-6}$  M,  $p < 0.003$ ); ancRIG-I CARD2 did not ( $p > 0.31$ ), and the combined RIG-I CARD1+2 domain bound similarly to CARD1, alone ( $p > 0.09$ ) (Fig 7B and S18 Fig). In contrast, although the ancestral MDA5 CARD1+2 bound ancestral IPS1 CARD ( $K_d = 5.89e^{-6}$  M,  $K_m = 6.61e^{-6}$  M,  $p < 0.04$ ), ancestral MDA5 CARD1 and ancMDA5 CARD2, alone, were not significantly better than the negative control ( $p > 0.06$ ) (Fig 7B and S18 Fig), suggesting that the similar functional differentiation of MDA5 CARD1 and CARD2 that we observed in human MDA5 (Fig 1B) must have occurred later in the MDA5 lineage.

Together, these results suggest that early coevolution between RLRs and IPS1 was responsible for differentiating RLR-IPS1 CARD-CARD interactions in deuterostomes. Before the proliferation of RLR and IPS1 CARDS in deuterostomes, both RIG-I and MDA5 CARD domains had the potential to bind an IPS1-like signaling partner (Fig 6A). After the deuterostome divergence, all RLR CARDS retained their capacity to bind the ancestral MDA5 CARD1 / IPS1 CARD (Fig 7A), but only RIG-I CARD1 and the combined MDA5 CARD1+2 could bind the diverged IPS1 (Fig 7B). The changes that accrued in IPS1 CARD as it diverged from the RLR CARD domains also interfered with its capacity to bind its potential ancestral partners (Fig 6A), suggesting that coevolution between IPS1 and RLR CARD domains was responsible for



**Fig 6. Ancestral RIG-like receptor (RLR) CARDs bound the IPS1 CARD progenitor before the proliferation of RLR-IPS1 CARD domains in early deuterostomes, but did not bind the ancestral IPS1 CARD after it diverged from RLR CARDs.** We resurrected ancestral RLR and IPS1 CARD domains and measured their binding kinetics in vitro (see [Materials and Methods](#)) (S1 Fig). The simplified phylogeny at left indicates the ancestral proteins examined in this set of experiments, with the dotted vertical line indicating the approximate time of the RLR-IPS1 CARD proliferation in deuterostomes (Figs 3 and 5). Colored circles indicate ancestral RLR CARDs, while gray circles indicate the ancestral IPS1 CARD progenitor before (A) and after (B) the deuterostome proliferation. We show the mean and standard error in  $-\log$ -transformed steady-state dissociation (pKd) and initial binding rate (pKm) of in vitro kinetics assays, with dotted vertical line indicating approximate nonspecific binding (indicated by CASP9 negative control) and longer bars indicating tighter binding (S16 Fig).

doi:10.1371/journal.pone.0137276.g006

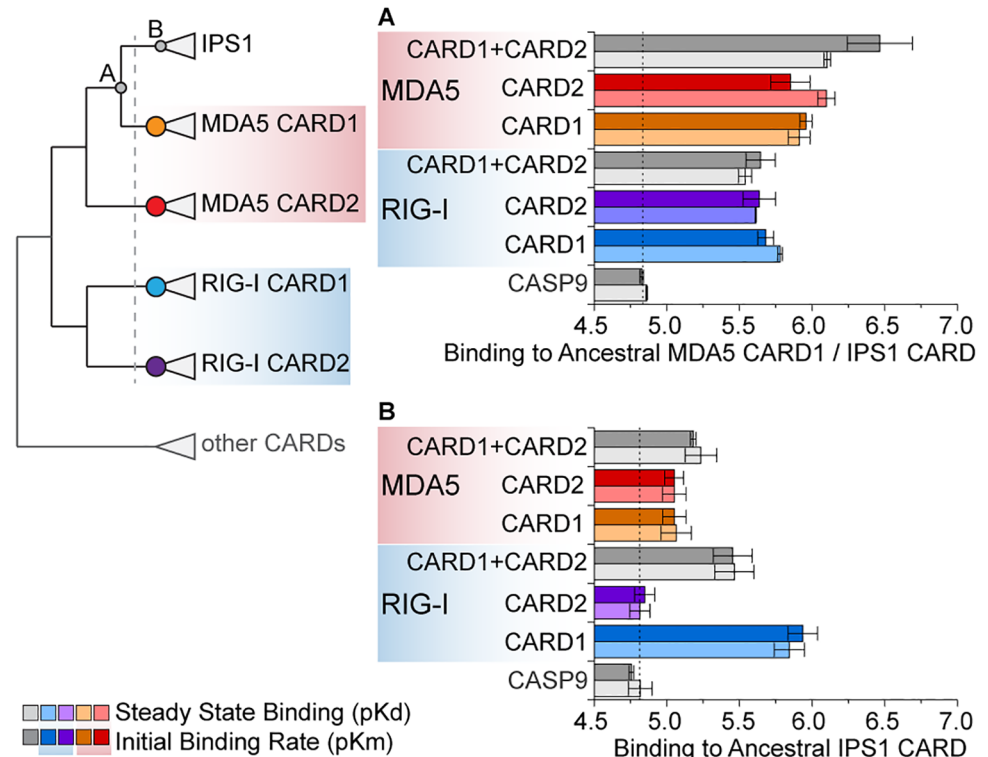
differentiating RLR-IPS1 interactions. Whether molecular adaptation drove the establishment of these interactions or was a result of selection for other functions is unclear. The implications of these molecular interactions for the evolution of cellular immune signaling are also unclear, as interactions with ubiquitinating enzymes and binding of free polyubiquitin could also have regulated how direct RLR-IPS1 CARD-CARD interactions functioned in early cellular immunity.

## Conclusion

The importance of the RIG-like receptors (RIG-I, MDA5 and LGP2) to antiviral immunity is well established, but we still lack a precise understanding of the specific biochemical interactions responsible for RLR-based immune signaling. An emerging picture—based on detailed analyses of human and other vertebrate RLRs—suggests that RIG-I and MDA5 activate immune responses primarily through direct interactions with IPS1, mediated by homologous RLR and IPS1 CARD domains and modulated by K63-linked polyubiquitin chains and ubiquitinating enzymes [5, 6, 14, 15].

Although recent studies have highlighted the importance of ubiquitinating enzymes and binding of free polyubiquitin chains for RLR-based immune signaling [14, 15], the mechanisms by which ubiquitin may facilitate immune activation remain unclear. Ubiquitin chains might enhance RLR binding to specific signaling partners, facilitate localization of RLRs to the mitochondrial membrane, generate or stabilize oligomeric signaling complexes, occlude binding sites used by negative regulators or function through other mechanisms.

Structural and mutational studies have suggested that—in the presence of free K63-linked polyubiquitin—it is the CARD2 domain of RIG-I that directly binds IPS1 CARD [13, 16]. However, here we demonstrate that—in the absence of ubiquitin—RIG-I and MDA5 CARD1



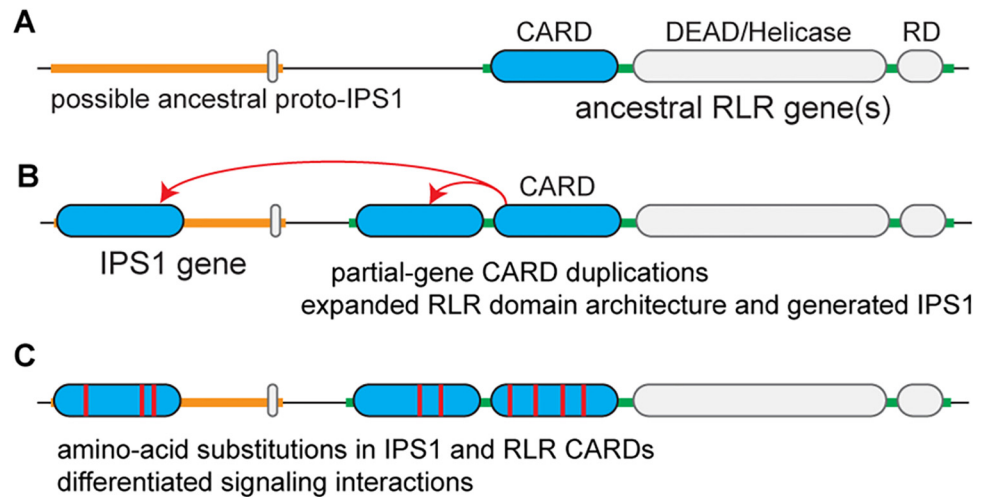
**Fig 7. Functional differentiation of ancestral RIG-I CARD1 and CARD2 domains occurred after the proliferation of RLR-IPS1 CARDS in deuterostomes, due to coevolution between RIG-I and IPS1 CARDS.** Ancestral RLR and IPS1 CARD domains were resurrected, and we measured the kinetics of RIG-I and MDA5 CARDS bound to IPS1 CARD before (A) and after (B) the RLR-IPS1 CARD proliferation in early deuterostomes (dotted vertical line on simplified CARD phylogeny) (Figs 3 and 5). We plot the mean and standard error of  $-\log$ -transformed steady-state dissociation (pKd) and initial binding rate (pKm) when ancestral RLR CARD domains bind to IPS1 CARD before (A) and after (B) the deuterostome proliferation (see [Materials and Methods](#)). The approximate level of nonspecific binding is indicated by a vertical dotted line (inferred using human CASP9), and longer bars indicate tighter binding (S17 and S18 Figs).

doi:10.1371/journal.pone.0137276.g007

domains are primarily responsible for mediating direct binding to IPS1 CARD. Our results do not contradict the body of structural and mutational work showing that RIG-I CARD2 interacts directly with IPS1 in the presence of free ubiquitin. Rather, we feel our results suggest that free ubiquitin may modulate the RLR-IPS1 interface. Covalently linked polyubiquitin may further modulate this interface, although further studies will be required to examine these hypotheses.

Our evolutionary analyses characterize how the RLR-IPS1 interaction—absent ubiquitin—changed over time. The consequences of these changes in RLR molecular function for the evolution of cellular immune signaling remain unclear, as free ubiquitin and ubiquitinating enzymes are likely to have played a role in the evolution of RLR-IPS1 immune signaling. Larger evolutionary studies examining the complete RLR-IPS1-ubiquitin signaling system will be required to understand how the evolution of specific molecular-functional aspects of this system impacted organismal evolution.

In contrast to the relatively large body of work examining RLR-IPS1 interactions in mammals and other vertebrates, very little is known about how RLRs and IPS1 evolved. Earlier studies suggested that RLRs arose and diversified in vertebrates through complex evolutionary processes [44, 45]. In our view, these early studies suffered from incomplete sequence data.



**Fig 8. Our results suggest a model in which partial-gene duplication and subsequent coevolution generated RLR-IPS1 interactions.** (A) Prior to the deuterostome proliferation of RLR and IPS1 CARD domains, ancestral RLR(s) functioned through an unknown immune-signaling mechanism. (B) Partial-gene duplications of RLR CARD(s) in deuterostomes generated the canonical RIG-I—MDA5 2-CARD domain architecture as well as the IPS1 gene, creating a new potential immune-signaling partnership. (C) Coevolution between RLR and IPS1 CARD domains differentiated RLR-IPS1 signaling interactions. Note that this model is parsimonious, in that it only posits the necessary duplications of CARD signaling domains. More complex scenarios are also possible, such as the complete duplication of one of the RLRs, followed by loss of RNA-binding domains to produce IPS1. Whether RLR-IPS1 evolution followed our simple parsimonious model or a more complicated scenario cannot be determined, given current data.

doi:10.1371/journal.pone.0137276.g008

Our previous work and this study argue that RLRs arose in early metazoa and diversified through a series of gene duplication events [10]. Our phylogenetic analyses—which appear robust to major error—further suggest that IPS1 arose much later than the RLRs, probably in early deuterostomes. This finding raises the question, how did the RLRs function before the origin of their canonical signaling partner? Examination of RLR-based immune signaling in early animals will likely be required to answer this question and is expected to further illuminate the mechanisms driving the evolution of animal innate immunity.

How novel protein-protein signaling interactions arise and establish themselves as distinct from existing interactions has been a long-standing question in molecular evolution, but we have had very little direct information to address this question. Our examination of RLR-IPS1 interactions across evolutionary history suggests a general model in which the duplication of individual domains involved in existing protein-protein interactions into new gene contexts could facilitate the immediate establishment of new functional interactions (Fig 8A and 8B). Subsequent coevolution between these new interaction partners could then differentiate the newly-established signaling interaction from other existing interactions (Fig 8C). Our work also suggests that gene clusters may play an important role in facilitating the generation of novel protein-protein interactions by bringing different genes into close physical proximity, allowing unequal crossing-over and other localized processes to duplicate functional domains into new gene contexts.

## Materials and Methods

### Binding Kinetics Analysis

CARD proteins were expressed in *E. coli* Rosetta 2(DE3)pLysS cells using pET-22b(+) constructs, which were verified by Sanger sequencing (S1 Fig). 500-mL cultures were inoculated

with 10-mL suspended cells and incubated for 3 hours at 37° C. Cells were lysed using mechanical disruption with MP Biomedicals Fastprep 24, and cellular material was filtered by centrifugation. CARD proteins were purified under non-denaturing conditions by His-affinity purification using an elution buffer containing 50 mM sodium phosphate, 300 mM sodium chloride and 150 mM imidazole at pH 7.4. Resulting purified protein was dialyzed overnight at 4° C in 1000 mL PBS buffer using a 3500 MWCO dialysis membrane to remove salts and other unwanted components. Proteins were visualized by SDS-page stained with 1% coomassie to confirm purity, and protein concentrations were measured via a linear-transformed Bradford assay [46]. IPS1 CARDS were FLAG-tagged on the N-terminus to facilitate binding kinetics assays. RLR CARDS were not FLAG-tagged.

We measured IPS1-RLR CARD-CARD binding using a label-free in vitro kinetics assay at pH 7.0 [47]. IPS1 CARD proteins at 2.5 µg/mL concentration were immobilized to a series of 8 anti-FLAG probes for 5 minutes, until saturation was observed. Probes were washed in 1x Kinetics Buffer (ForteBio) and then exposed in parallel to RLR CARDS at 0.001–100 µM concentrations in 1x Kinetics Buffer for 60 minutes, at which point binding curve saturation was observed, and all probes were allowed to dissociate in Kinetics Buffer for an additional 60 minutes. All kinetics buffers included BSA to control for non-specific binding. CARD-CARD association and dissociation were measured as the nanometer shift in laser wavelength reflected through each probe over time, sampled every 3 milliseconds. Results from a parallel probe run at 0 µM RLR CARD concentration were subtracted to account for system fluctuation over time.

Steady-state molecular binding at saturation was measured as the change in laser wavelength averaged over the last 500 samples during CARD-CARD association. We additionally fit association-dissociation curves to the kinetics data in order to estimate the initial rates of CARD-CARD binding. Measurements were taken from 3 experimental replicates, each of which produced 7 normalized association-dissociation plots across RLR CARD concentrations.

We estimated RLR CARD concentrations at which ½-maximal steady-state CARD-CARD binding ( $K_d$ ) and ½-maximal initial binding rate ( $K_m$ ) were achieved using two different approaches, which produced qualitatively equivalent results. First, we estimated  $K_d$  and  $K_m$  values separately using the data generated from each experimental replicate and then calculated average  $K_d$  and  $K_m$  values across replicates. Second, we calculated the average and standard-error of each of the 7 concentration-response data points across the 3 experimental replicates and estimated  $K_d$  and  $K_m$  values from these ‘average’ data. Results from the first approach are presented as bar plots and discussed in the main text. Results from the second approach are displayed as concentration-response graphs in supplementary figures. In either case,  $K_d$  and  $K_m$  were estimated by fitting a one-site binding curve using nonlinear regression. All curve fits were inferred with  $r^2 > 0.91$ . Differences in  $K_d$  and  $K_m$  were assessed using two-tailed, two-sample  $t$  tests, assuming unequal variances between samples. For visualization, we report  $-\log_{10}$ -transformed  $K_d$  ( $pK_d$ ) and  $K_m$  ( $pK_m$ ).

## Sequence Identification and Alignment

Full-length RLR and IPS1 protein sequences were identified by sequence search of the NCBI RefSeq and Ensembl databases [48, 49], using human RLRs and IPS1 as queries and a BLASTP  $e$ -value cutoff of  $10^{-10}$  [50]. Protein sequences from well-curated vertebrate genomes were then used to refine RLR and IPS1 gene models in the genomes of less reliably-curated deuterostomes by homology-based gene identification using FGENSEH+ [51].



RLR and IPS1 CARD domains were identified by sequence search of the Pfam database [52], the Conserved Domain Database [53] and the SMART database [54]. Domain boundaries were further refined by aligning RLR and IPS1 sequences to those whose CARD domains have been structurally determined (PDB IDs: 4A2W, 4A2Q, 2LWD, 2LWE, 4NQK, 2VGQ, 3YGS, 3J6J, 4P4H).

CARD domain sequences were aligned using the G-INS-i algorithm in Mafft v6.850b [55] and using Muscle v3.8.31, with default parameters [56]. Alignments were further manually refined by comparing the sequence alignment to a structural alignment of the above PDB IDs, produced using the CEAlign algorithm available in PyMOL v1.7 [57].

## Phylogenetic Analysis

The maximum-likelihood (ML) phylogeny was inferred using PhyML v3.0 [58], with the best-fit evolutionary model selected by AIC in ProtTest v3 [59]. We used the LG substitution model [60], estimating the residue frequencies from the alignment data and allowing for an 8-category discrete-gamma approximation of among-site rate variation. The tree topology was inferred by heuristic search using a combination of SPR and NNI rearrangements. Support for each clade was assessed using SH-like aLRT [61].

Bayesian phylogenetic analysis was conducted using MrBayes v3.2.1 [62, 63]. We executed 2 independent runs, with 4 chains/run, and terminated the analysis when the deviation in average clade posterior probabilities between the 2 runs fell below 0.01. Tree topologies were sampled every 100 steps in the MCMC chains, with the first 25% of samples discarded as burnin. The amino-acid substitution model was integrated over using a mixed prior; the proportion of invariant sites was set to zero, and all other prior distributions were set to default values.

Protein-coding adaptation was inferred using the branch-sites test in PAML v4.7 [39, 64]. Each gene's codon sequence was mapped to the amino-acid alignment, and the likelihood of the ML tree was calculated under two evolutionary models, one in which an inferred proportion of codons are allowed to have a nonsynonymous/synonymous rate ratio ( $\omega$ ) > 1 on a single branch, indicative of positive selection on the protein sequence, and a null model that constrains  $\omega \leq 1$  for all sites and branches. We calculated the likelihood ratio of the positive-selection model to the null model and assessed significance using a Bonferroni-corrected chi-square test [39].

## Ancestral Sequence Reconstruction

Ancestral protein sequences were reconstructed using the June, 2012 release of Lazarus, which automates the integration of ancestral sequence reconstructions across multiple tree topologies [36]. We included the ML topology, the alternative topology (S8 Fig), and all topologies in the 95% credibility set inferred by Bayesian phylogenetic analysis. Ancestral protein sequences were inferred from each topology by marginal reconstruction using PAML v4.7 [64, 65], assuming the best-fit LG+F+G evolutionary model. The posterior probability of each ancestral-reconstructed residue was inferred by PAML and weighed by the empirical-Bayesian posterior probability of the tree on which it was reconstructed. Equivalent ancestral sequences were obtained using RAXML v8.0.0 [66]. Insertions and deletions (indels) in ancestral sequences were inferred by maximum parsimony using PHYLIP v3.695, assuming a simple presence-absence coding scheme. As with the sequence reconstructions, each indel reconstruction was weighted by the posterior probability of the tree on which it was reconstructed.

## Structural Modeling and Docking

Structural models were constructed using MODELLER v9.10 [67], with templates selected by sequence search of the PDB database [68]. For each sequence, we constructed 100 structural models and assessed the free energy of each model using MODELLER objective function and DOPE scores [69]. We converted each model's scores to units of deviation from the mean over the 100 models and averaged the deviation-converted MODELLER and DOPE scores to get a single overall score for each model; the model with the minimal overall score was selected as the most reliable structural model.

Hydrogen atoms were added to structural models, side-chain pKas calculated, structures optimized for favorable hydrogen bonding, and Amber force-field parameters assigned using PDB2PQR v1.7 and PROPKA v3.1 [70]. The distribution of electrostatic charge around each structure was estimated using APBS v1.4.1 [71] and projected onto the molecular surface for visualization using PyMOL.

We docked the structure of human RIG-I CARD1 (PDB ID: 4P4H) into IPS1 CARD (PDB ID: 2VGQ) using the ClusPro v2.0 [17] and Dot2 v2.0 [18] protein-protein docking protocols. For ClusPro, we used the default parameters. For Dot2, we generated and scored 54,000 orientations using a grid of 128x128x128 points with a step size of 1.0 Å and 6° rotation; other parameters had default values. We also docked structural models of human MDA5 CARD1 into IPS1 CARD using the same protocols.

## Supporting Information

**S1 Fig. We show the extant and ancestral protein sequences analyzed in this study, in FASTA format.** GenBank identifiers are provided for extant sequences. For proteins encoding multiple caspase activation and recruitment domains (CARDs), the first (CARD1) and second (CARD2) domains are highlighted different colors. Dual-CARD proteins were expressed as single constructs for kinetics analyses. Alternative human RIG-I and MDA5 CARD1 and CARD2 domains lacking the CARD1-CARD2 linker and having boundaries determined from structural data are indicated (S5 Fig).

(TIF)

**S2 Fig. The first caspase activation and recruitment domains (CARDs) of human RIG-like receptors (RLRs) RIG-I and MDA5 bind the CARD domain of IPS1.** We expressed and purified untagged RLR CARD constructs including the first CARD domain (CARD1), the second CARD domain (CARD2) and the N-terminal region expressing both CARD domains (CARD1+2) (Fig 1A). We measured binding of these constructs to flag-tagged IPS1 CARD using an in vitro kinetics assay (see Materials and Methods). For each kinetics experiment, we plot the shift in laser wavelength during RLR-IPS1 association (Y-axis) against the concentration of RLR CARD protein (X-axis), both at steady-state (black) and under initial conditions (blue). Bars indicate standard errors over 3 replicates. We fit one-site concentration-response curves to each set of results by nonlinear regression and estimate the ½-maximal steady-state binding concentration (Kd) and the ½-maximal initial binding rate (Km). Left-shifted curves indicate tighter binding.

(TIF)

**S3 Fig. Kinetics analysis of untagged RLR CARDs supports the conclusion that the first CARD domains of RIG-I and MDA5 bind tagged IPS1 CARD.** We expressed and purified untagged human RLR CARD1, CARD2 and CARD1+2 constructs (Fig 1A). We measured binding of these constructs to IPS1 CARD fused to a maltose binding protein using an in vitro kinetics assay (see Materials and Methods). We plot the shift in laser wavelength during

CARD-CARD association (Y-axis) against the concentration of RLR CARD (X-axis) at steady-state (black) and under initial conditions (blue). Bars indicate standard errors over 3 replicates. We fit one-site concentration-response curves by nonlinear regression and estimate the  $\frac{1}{2}$ -maximal steady-state concentration (Kd) and  $\frac{1}{2}$ -maximal initial binding rate (Km). Left-shifted curves indicate tighter binding. Inset, we plot the  $-\log$ -transformed steady-state dissociation constant (pKd) and the  $-\log$ -transformed initial binding rate (pKm) estimated over 3 replicates, with bars indicating standard errors. Human CASP9 CARD domain was used as a negative control.

(TIF)

**S4 Fig. Kinetics analysis of flag-tagged RLR CARDS supports the conclusion that the first CARD domains of RIG-I and MDA5 bind untagged IPS1 CARD.** We purified flag-tagged RLR CARD constructs including the first CARD domain (CARD1), the second CARD domain (CARD2) and both CARD domains in tandem (CARD1+2) (Fig 1A). We measured binding of these constructs to untagged IPS1 CARD in vitro (see [Materials and Methods](#)). We plot the shift in laser wavelength during association (Y-axis) against RLR CARD concentration (X-axis) at steady-state (black) and under initial conditions (blue). Bars indicate standard errors over 3 replicates. We fit one-site concentration-response curves by nonlinear regression and estimate the  $\frac{1}{2}$ -maximal steady-state concentration (Kd) and initial binding rate (Km). Inset, we plot the  $-\log$ -transformed steady-state dissociation constant (pKd) and the  $-\log$ -transformed initial binding rate (pKm). Bars indicate standard errors over 3 replicates. Human CASP9 CARD was used as a negative control.

(TIF)

**S5 Fig. Kinetics analysis of RLR CARD1 and CARD2 constructs matching structural domain boundaries supports the conclusion that CARD1 binds IPS1 CARD.** We purified human RIG-I and MDA5 CARD1 and CARD2 structural domains, excluding CARD1-CARD2 linkers (S1 Fig). We measured binding of these constructs to IPS1 CARD in vitro (see [Materials and Methods](#)). We plot the shift in laser wavelength during association (Y-axis) against RLR CARD concentration (X-axis) at steady-state (black) and under initial conditions (blue), with bars indicating standard errors over 3 replicates. We fit one-site concentration-response curves by nonlinear regression and estimate the  $\frac{1}{2}$ -maximal steady-state concentration (Kd) and initial binding rate (Km). Inset, we plot the  $-\log$ -transformed steady-state dissociation constant (pKd) and the  $-\log$ -transformed initial binding rate (pKm), with bars indicating standard errors over 3 replicates. Human CASP9 CARD was used as a negative control.

(TIF)

**S6 Fig. Mutations in human RIG-I CARD1 and IPS1 CARD disrupt RIG-I–IPS1 interactions in vitro.** We created mutant RIG-I CARD1 (Glu66,67—Arg) and IPS1 (Arg66,67—Glu) and measured the affinities with which mutant proteins bind their wild-type partners. For each interaction, we plot the shift in laser wavelength during association (Y-axis) against RIG-I CARD1 concentration (X-axis) at steady-state (black) and under initial conditions (blue), with bars indicating standard errors over 3 replicates. We fit one-site concentration-response curves by nonlinear regression and estimate the  $\frac{1}{2}$ -maximal steady-state concentration (Kd) and  $\frac{1}{2}$ -maximal initial binding rate (Km), with left-shifted curves indicating tighter binding. We show binding curves for wild-type RIG-I CARD1 and CASP9 interacting with wild-type IPS1 CARD as positive and negative controls, respectively.

(TIF)

**S7 Fig. Prediction of the human MDA5 CARD1 –IPS1 CARD interface by protein-protein docking was ambiguous.** We docked structural models of human MDA5 CARD1 into IPS1

CARD using ClusPro v2.0 (A) [17] and Dot2 v2.0 (B) [18]. Top panels show the best-scoring conformation from each analysis; bottom panels display the top 20 orientations of MDA5 CARD1 from each analysis.

(TIF)

**S8 Fig. Plausible evolutionary histories for RLR-IPS1 CARD domains.** We plot the maximum likelihood (A) and alternative (B) phylogenies inferred from our analysis of RLR and IPS1 CARDS. Given each phylogeny, RLR gene duplication events (green circles) and duplications of individual RLR CARDS (red stars) are shown. Branch lengths are not drawn to scale.

(TIF)

**S9 Fig. Removal of fast-evolving non-vertebrate sequences increases support for the RLR-IPS1 CARD consensus tree.** We reconstructed the RLR-IPS1 CARD phylogeny (Fig 3A) using only vertebrate and *Nematostella* CARD sequences. Support for key nodes is reported as maximum-likelihood SH-like aLRT scores (top) [61] and Bayesian posterior probabilities (bottom) [63]. RLR gene duplication events (green circles) and duplications of individual RLR CARDS (red stars) are indicated. Branch lengths are scaled to substitutions/site.

(TIF)

**S10 Fig. HMM-distance clustering supports the RLR-IPS1 CARD consensus phylogeny.**

We constructed hidden Markov models (HMMs) from separate alignments of each group of RLR and IPS1 CARDS (Fig 3) using HMMER v3.0 [24]. We then calculated the Kulback-Liebler distance between all pairs of HMMs [26] and reconstructed the neighbor joining (NJ) tree from these distances using QuickTree v1.1 [27]. We plot HMM-logo views [25] of each HMM as well as the resulting NJ tree. Branch lengths are not drawn to scale.

(TIF)

**S11 Fig. Gene-species tree reconciliation roots the RLR-IPS1 CARD phylogeny on *Nematostella*.** We used Notung v2.6 [28, 30] to identify the optimal root for the RLR-IPS1 CARD phylogeny by reconciling the consensus RLR-IPS1 CARD tree (Fig 3A) with our current understanding of the metazoan species tree [29]. Notung uses a duplication/loss model to root gene trees in the absence of outgroup information. We plot Notung rooting scores on the consensus RLR-IPS1 CARD phylogeny. Red branches indicate optimal rootings. Pink branches indicate possible alternative roots identified by Notung (none identified). RLR gene duplication events (green circles), duplications of individual RLR CARDS (red stars) and the origin of IPS1 CARD (orange star) are indicated. Numbers in parentheses on collapsed nodes indicate the number of individual species in each group. Branch lengths are scaled to substitutions/site, inferred by maximum likelihood.

(TIF)

**S12 Fig. *Branchiostoma floridae* RLR CARDS interact with *B. floridae* IPS1 CARD in vitro.**

We measured the kinetics of *B. floridae* RIG-I CARD1+2 and MDA5 CARD1+2 domains binding to *B. floridae* IPS1 CARD in vitro (see Materials and Methods) (S1 Fig). We plot the shift in laser wavelength during RLR-IPS1 association (Y-axis) against RLR concentration (X-axis) at steady-state (black) and under initial conditions (blue). Bars indicate standard errors over 3 replicates. We fit one-site concentration-response curves by nonlinear regression to estimate the  $\frac{1}{2}$ -maximal steady-state binding concentration (Kd) and the  $\frac{1}{2}$ -maximal initial binding rate (Km). Left-shifted curves indicate tighter binding. Human CASP9 was used as a negative control.

(TIF)

**S13 Fig. Genomic locations of RLR and IPS1 genes in pre-vertebrate deuterostomes suggest these genes were likely clustered in the ancestral deuterostome genome.** For each RLR (RIG-I, MDA5, LGP2) and IPS1 gene in the *Branchiostoma floridae* gene cluster (shown in center) (Fig 4A), we used BLAST to identify potential homologs in other pre-vertebrate deuterostome genome sequences. We identify the chromosome (for *Ciona intestinalis*) or genomic scaffold location of each BLAST hit. Colored lines indicate best-hit BLAST matches, whereas gray lines indicate alternative BLAST hits with  $e$ -values  $< 10^{-5}$ .  
(TIF)

**S14 Fig. Ancestral RLR-IPS1 CARD sequences were reconstructed with high confidence, despite ambiguity in the underlying phylogeny.** We reconstructed the ancestral sequences of key nodes on the RLR-IPS1 CARD phylogeny (Fig 3A) using maximum-likelihood methods that explicitly incorporate uncertainty about the tree topology (see [Materials and Methods](#)). For each sequence, we plot the frequency with which individual residues were reconstructed with posterior probability ranging from 0.0 to 1.0, binned every 0.05 (Fig 5). Inset into each graph, we simulated protein sequence data along the maximum-likelihood phylogeny using the best-fit evolutionary model and plot the proportion of ancestral residues incorrectly inferred at each node. Dark series indicate error rates when each residue is considered unique, whereas light series indicate error rates when residues with similar biochemical properties are treated as equivalent; bars indicate standard errors. In the colored inset at top, we plot the frequency of each amino-acid residue averaged across extant sequences of each type, with bars indicating standard errors.  
(TIF)

**S15 Fig. Simulations suggest low expected false-positive error rates for branch-sites tests of positive selection on the RLR CARD phylogeny.** We used PAML v4.7 to simulate 100 replicates of codon sequence data along the maximum-likelihood RLR CARD phylogeny using three simulation models. The Empirical model estimated all simulation parameters from our empirical sequence data assuming a 'sites' model, in which an inferred proportion of sites evolve with non-synonymous/synonymous rate ratio,  $\omega < 1$ , and the remaining sites evolve with  $\omega = 1$ . The Empirical Branch-Sites model is the same as the Empirical model, except all sites are released from selective constraint ( $\omega = 1$ ) on the specific branch being tested. The Neutral model allows all sites to evolve neutrally ( $\omega = 1$ ) on all branches. All other model parameters were estimated by maximum likelihood. We analyzed each replicate data set using the same procedure used to test our empirical data for protein-coding adaptation on specific branches (see [Materials and Methods](#)). We plot the mean and standard error in false-positive error rate over the 100 data sets simulated using each model. Dotted line indicates false-positive rate of 0.05.  
(TIF)

**S16 Fig. Ancestral RIG-like receptor (RLR) CARDs lost their capacity to bind IPS1 CARD after it diverged from RLR CARDs.** We measured the binding kinetics of ancestral RLR and IPS1 CARD domains in vitro (see [Materials and Methods](#)). We plot the shift in laser wavelength during association (Y-axis) against RLR CARD concentration (X-axis) at steady-state (black) and under initial conditions (blue), with bars indicating standard errors over 3 replicates. We fit one-site concentration-response curves by nonlinear regression and estimate the  $\frac{1}{2}$ -maximal steady-state concentration (Kd) and initial binding rate (Km). Ancestral RLR CARD binding to the ancestral IPS1 CARD progenitor before (A) and after (B) the deuterostome proliferation is shown (Fig 6). Human CASP9 CARD domain was used as a negative control.  
(TIF)

**S17 Fig. Ancestral RIG-I and MDA5 CARD domains retained their capacity to bind the MDA5 / IPS1 CARD progenitor after they diverged from their common ancestor in deuterostomes.** We measured the kinetics of ancestral RIG-I and MDA5 CARDS bound to the MDA5 CARD1 / IPS1 CARD progenitor (see [Materials and Methods](#)) ([Fig 7A](#)). We plot the shift in laser wavelength during association (Y-axis) against RLR CARD concentration (X-axis) at steady-state (black) and under initial conditions (blue), with bars indicating standard errors over 3 replicates. We fit one-site concentration-response curves by nonlinear regression and estimate the  $\frac{1}{2}$ -maximal steady-state concentration (Kd) and  $\frac{1}{2}$ -maximal initial binding rate (Km). Human CASP9 CARD was used as a negative control. (TIF)

**S18 Fig. After the proliferation of RLR and IPS1 CARDS in deuterostomes, RIG-I CARDS diversified into IPS1-binding CARD1 and CARD2, which does not bind IPS1.** We measured the kinetics of ancestral RIG-I and MDA5 CARDS bound to the ancestral IPS1 CARD after it differentiated from RLR CARDS (see [Materials and Methods](#)) ([Fig 7B](#)). We plot the shift in laser wavelength during association (Y-axis) against RLR CARD concentration (X-axis) at steady-state (black) and under initial conditions (blue), with bars indicating standard errors over 3 replicates. We fit one-site concentration-response curves by nonlinear regression and estimate the  $\frac{1}{2}$ -maximal steady-state concentration (Kd) and  $\frac{1}{2}$ -maximal initial binding rate (Km). Human CASP9 CARD was used as a negative control. (TIF)

**S1 Table. Gene conversion is rare among extant RIG-like receptor (RLR) caspase activation and recruitment domains (CARDS).** We used GENECONV v1.81a to identify regions of potential gene conversion among RLR CARDS [[32](#)]. We report results with significant support for gene conversion between sequence pairs, after correcting for multiple tests ( $p < 0.05$ ). **Sim P**, simulated  $p$  values based on 10,000 permutations; **BC KA P**, Bonferroni-corrected Karlin-Altschul  $p$  values; **Beg**, first nucleotide of the potential converted region; **End**, last nucleotide of the potential converted region; **Poly**, number of polymorphic sites in the region; **Len**, length of the converted region; **Diff**, number of nucleotide mismatches in the potential converted region; **Total Diff**, total number of nucleotide mismatches between the two sequences. (DOCX)

## Acknowledgments

This research was supported by the National Institute of Allergy and Infectious Diseases of the National Institutes of Health under award number R21AI101571, grant number FLA-MCS-005262 from the USDA National Institute of Food and Agriculture, and by the University of Florida. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health, the USDA or the University of Florida.

## Author Contributions

Conceived and designed the experiments: B. Korithoski B. Kolaczowski KM. Performed the experiments: B. Korithoski B. Kolaczowski OK KM RK CE. Analyzed the data: B. Korithoski B. Kolaczowski OK KM RK CE. Contributed reagents/materials/analysis tools: B. Korithoski. Wrote the paper: B. Korithoski B. Kolaczowski.

## References

1. Dean AM, Thornton JW. Mechanistic approaches to the study of evolution: the functional synthesis. *Nature reviews Genetics*. 2007; 8(9):675–88. doi: [10.1038/nrg2160](https://doi.org/10.1038/nrg2160) PMID: [17703238](https://pubmed.ncbi.nlm.nih.gov/17703238/); PubMed Central PMCID: PMC2488205.
2. Pichlmair A, Reis e Sousa C. Innate recognition of viruses. *Immunity*. 2007; 27(3):370–83. doi: [10.1016/j.immuni.2007.08.012](https://doi.org/10.1016/j.immuni.2007.08.012) PMID: [17892846](https://pubmed.ncbi.nlm.nih.gov/17892846/).
3. Takeuchi O, Akira S. Innate immunity to virus infection. *Immunological reviews*. 2009; 227(1):75–86. doi: [10.1111/j.1600-065X.2008.00737.x](https://doi.org/10.1111/j.1600-065X.2008.00737.x) PMID: [19120477](https://pubmed.ncbi.nlm.nih.gov/19120477/).
4. Loo YM, Fornek J, Crochet N, Bajwa G, Perwitasari O, Martinez-Sobrido L, et al. Distinct RIG-I and MDA5 signaling by RNA viruses in innate immunity. *Journal of virology*. 2008; 82(1):335–45. doi: [10.1128/JVI.01080-07](https://doi.org/10.1128/JVI.01080-07) PMID: [17942531](https://pubmed.ncbi.nlm.nih.gov/17942531/); PubMed Central PMCID: PMC2224404.
5. Kawai T, Takahashi K, Sato S, Coban C, Kumar H, Kato H, et al. IPS-1, an adaptor triggering RIG-I and Mda5-mediated type I interferon induction. *Nature immunology*. 2005; 6(10):981–8. doi: [10.1038/ni1243](https://doi.org/10.1038/ni1243) PMID: [16127453](https://pubmed.ncbi.nlm.nih.gov/16127453/).
6. Xu LG, Wang YY, Han KJ, Li LY, Zhai Z, Shu HB. VISA is an adapter protein required for virus-triggered IFN-beta signaling. *Molecular cell*. 2005; 19(6):727–40. doi: [10.1016/j.molcel.2005.08.014](https://doi.org/10.1016/j.molcel.2005.08.014) PMID: [16153868](https://pubmed.ncbi.nlm.nih.gov/16153868/).
7. Rothenfusser S, Goutagny N, DiPerna G, Gong M, Monks BG, Schoenemeyer A, et al. The RNA helicase Lgp2 inhibits TLR-independent sensing of viral replication by retinoic acid-inducible gene-I. *Journal of immunology*. 2005; 175(8):5260–8. PMID: [16210631](https://pubmed.ncbi.nlm.nih.gov/16210631/).
8. Bruns AM, Leser GP, Lamb RA, Horvath CM. The innate immune sensor LGP2 activates antiviral signaling by regulating MDA5-RNA interaction and filament assembly. *Molecular cell*. 2014; 55(5):771–81. doi: [10.1016/j.molcel.2014.07.003](https://doi.org/10.1016/j.molcel.2014.07.003) PMID: [25127512](https://pubmed.ncbi.nlm.nih.gov/25127512/); PubMed Central PMCID: PMC4156907.
9. Satoh T, Kato H, Kumagai Y, Yoneyama M, Sato S, Matsushita K, et al. LGP2 is a positive regulator of RIG-I- and MDA5-mediated antiviral responses. *Proceedings of the National Academy of Sciences of the United States of America*. 2010; 107(4):1512–7. doi: [10.1073/pnas.0912986107](https://doi.org/10.1073/pnas.0912986107) PMID: [20080593](https://pubmed.ncbi.nlm.nih.gov/20080593/); PubMed Central PMCID: PMC2824407.
10. Mukherjee K, Korithoski B, Kolaczowski B. Ancient origins of vertebrate-specific innate antiviral immunity. *Molecular biology and evolution*. 2014; 31(1):140–53. doi: [10.1093/molbev/mst184](https://doi.org/10.1093/molbev/mst184) PMID: [24109602](https://pubmed.ncbi.nlm.nih.gov/24109602/); PubMed Central PMCID: PMC3879448.
11. Sodergren E, Weinstock GM, Davidson EH, Cameron RA, Gibbs RA, Angerer RC, et al. The genome of the sea urchin *Strongylocentrotus purpuratus*. *Science*. 2006; 314(5801):941–52. Epub 2006/11/11. doi: [10.1126/science.1133609](https://doi.org/10.1126/science.1133609) PMID: [17095691](https://pubmed.ncbi.nlm.nih.gov/17095691/); PubMed Central PMCID: PMCpmc3159423.
12. Potter JA, Randall RE, Taylor GL. Crystal structure of human IPS-1/MAVS/VISA/Cardif caspase activation recruitment domain. *BMC structural biology*. 2008; 8:11. doi: [10.1186/1472-6807-8-11](https://doi.org/10.1186/1472-6807-8-11) PMID: [18307765](https://pubmed.ncbi.nlm.nih.gov/18307765/); PubMed Central PMCID: PMC2291057.
13. Wu B, Peisley A, Tetrault D, Li Z, Egelman EH, Magor KE, et al. Molecular imprinting as a signal-activation mechanism of the viral RNA sensor RIG-I. *Molecular cell*. 2014; 55(4):511–23. doi: [10.1016/j.molcel.2014.06.010](https://doi.org/10.1016/j.molcel.2014.06.010) PMID: [25018021](https://pubmed.ncbi.nlm.nih.gov/25018021/); PubMed Central PMCID: PMC4142144.
14. Jiang X, Kinch LN, Brautigam CA, Chen X, Du F, Grishin NV, et al. Ubiquitin-induced oligomerization of the RNA sensors RIG-I and MDA5 activates antiviral innate immune response. *Immunity*. 2012; 36(6):959–73. doi: [10.1016/j.immuni.2012.03.022](https://doi.org/10.1016/j.immuni.2012.03.022) PMID: [22705106](https://pubmed.ncbi.nlm.nih.gov/22705106/); PubMed Central PMCID: PMC3412146.
15. Gack MU, Shin YC, Joo CH, Urano T, Liang C, Sun L, et al. TRIM25 RING-finger E3 ubiquitin ligase is essential for RIG-I-mediated antiviral activity. *Nature*. 2007; 446(7138):916–20. doi: [10.1038/nature05732](https://doi.org/10.1038/nature05732) PMID: [17392790](https://pubmed.ncbi.nlm.nih.gov/17392790/).
16. Peisley A, Wu B, Xu H, Chen ZJ, Hur S. Structural basis for ubiquitin-mediated antiviral signal activation by RIG-I. *Nature*. 2014; 509(7498):110–4. Epub 2014/03/05. doi: [10.1038/nature13140](https://doi.org/10.1038/nature13140) PMID: [24590070](https://pubmed.ncbi.nlm.nih.gov/24590070/).
17. Comeau SR, Gatchell DW, Vajda S, Camacho CJ. ClusPro: an automated docking and discrimination method for the prediction of protein complexes. *Bioinformatics*. 2004; 20(1):45–50. PMID: [14693807](https://pubmed.ncbi.nlm.nih.gov/14693807/).
18. Roberts VA, Thompson EE, Pique ME, Perez MS, Ten Eyck LF. DOT2: Macromolecular docking with improved biophysical models. *Journal of computational chemistry*. 2013; 34(20):1743–58. doi: [10.1002/jcc.23304](https://doi.org/10.1002/jcc.23304) PMID: [23695987](https://pubmed.ncbi.nlm.nih.gov/23695987/).
19. Qin H, Srinivasula SM, Wu G, Fernandes-Alnemri T, Alnemri ES, Shi Y. Structural basis of procaspase-9 recruitment by the apoptotic protease-activating factor 1. *Nature*. 1999; 399(6736):549–57. doi: [10.1038/21124](https://doi.org/10.1038/21124) PMID: [10376594](https://pubmed.ncbi.nlm.nih.gov/10376594/).
20. Lensink MF, Wodak SJ. Docking, scoring, and affinity prediction in CAPRI. *Proteins*. 2013; 81(12):2082–95. doi: [10.1002/prot.24428](https://doi.org/10.1002/prot.24428) PMID: [24115211](https://pubmed.ncbi.nlm.nih.gov/24115211/).

21. Kozakov D, Beglov D, Bohnuud T, Mottarella SE, Xia B, Hall DR, et al. How good is automated protein docking? *Proteins*. 2013; 81(12):2159–66. 10.1002/prot.24403. PMID: [23996272](#); PubMed Central PMCID: PMC3934018. doi: [10.1002/prot.24403](#)
22. Atkinson HJ, Morris JH, Ferrin TE, Babbitt PC. Using sequence similarity networks for visualization of relationships across diverse protein superfamilies. *PLoS One*. 2009; 4(2):e4345. Epub 2009/02/05. doi: [10.1371/journal.pone.0004345](#) PMID: [19190775](#); PubMed Central PMCID: PMC2631154.
23. Smoot ME, Ono K, Ruscheinski J, Wang PL, Ideker T. Cytoscape 2.8: new features for data integration and network visualization. *Bioinformatics*. 2011; 27(3):431–2. Epub 2010/12/15. doi: [10.1093/bioinformatics/btq675](#) PMID: [21149340](#); PubMed Central PMCID: PMC3031041.
24. Finn RD, Clements J, Eddy SR. HMMER web server: interactive sequence similarity searching. *Nucleic Acids Res*. 2011; 39(Web Server issue):W29–37. Epub 2011/05/20. doi: [10.1093/nar/gkr367](#) PMID: [21593126](#); PubMed Central PMCID: PMC3125773.
25. Schuster-Bockler B, Schultz J, Rahmann S. HMM Logos for visualization of protein families. *BMC Bioinformatics*. 2004; 5:7. Epub 2004/01/23. doi: [10.1186/1471-2105-5-7](#) PMID: [14736340](#); PubMed Central PMCID: PMC341448.
26. Falkhausen M, Reiniger H, Wolf D. Calculation of Distance Measures Between Hidden Markov Models. In *Proc Eurospeech*. 1995:1487–90.
27. Howe K, Bateman A, Durbin R. QuickTree: building huge Neighbour-Joining trees of protein sequences. *Bioinformatics*. 2002; 18(11):1546–7. Epub 2002/11/09. PMID: [12424131](#).
28. Chen K, Durand D, Farach-Colton M. NOTUNG: a program for dating gene duplications and optimizing gene family trees. *J Comput Biol*. 2000; 7(3–4):429–47. Epub 2000/12/07. doi: [10.1089/106652700750050871](#) PMID: [11108472](#).
29. Dunn CW, Hejnal A, Matus DQ, Pang K, Browne WE, Smith SA, et al. Broad phylogenomic sampling improves resolution of the animal tree of life. *Nature*. 2008; 452(7188):745–9. Epub 2008/03/07. doi: [10.1038/nature06614](#) PMID: [18322464](#).
30. Vernot B, Stolzer M, Goldman A, Durand D. Reconciliation with non-binary species trees. *J Comput Biol*. 2008; 15(8):981–1006. Epub 2008/09/24. doi: [10.1089/cmb.2008.0092](#) PMID: [18808330](#).
31. Mansai SP, Innan H. The power of the methods for detecting interlocus gene conversion. *Genetics*. 2010; 184(2):517–27. doi: [10.1534/genetics.109.111161](#) PMID: [19948889](#); PubMed Central PMCID: PMC2828729.
32. Sawyer S. Statistical tests for detecting gene conversion. *Molecular biology and evolution*. 1989; 6(5):526–38. PMID: [2677599](#).
33. Putnam NH, Butts T, Ferrier DE, Furlong RF, Hellsten U, Kawashima T, et al. The amphioxus genome and the evolution of the chordate karyotype. *Nature*. 2008; 453(7198):1064–71. doi: [10.1038/nature06967](#) PMID: [18563158](#).
34. Drosophila 12 Genomes C, Clark AG, Eisen MB, Smith DR, Bergman CM, Oliver B, et al. Evolution of genes and genomes on the Drosophila phylogeny. *Nature*. 2007; 450(7167):203–18. doi: [10.1038/nature06341](#) PMID: [17994087](#).
35. Gao X, Lynch M. Ubiquitous internal gene duplication and intron creation in eukaryotes. *Proceedings of the National Academy of Sciences of the United States of America*. 2009; 106(49):20818–23. doi: [10.1073/pnas.0911093106](#) PMID: [19926850](#); PubMed Central PMCID: PMC2791625.
36. Hanson-Smith V, Kolaczkowski B, Thornton JW. Robustness of ancestral sequence reconstruction to phylogenetic uncertainty. *Molecular biology and evolution*. 2010; 27(9):1988–99. doi: [10.1093/molbev/msq081](#) PMID: [20368266](#); PubMed Central PMCID: PMC2922618.
37. Susko E, Roger AJ. Problems with estimation of ancestral frequencies under stationary models. *Syst Biol*. 2013; 62(2):330–8. doi: [10.1093/sysbio/sys075](#) PMID: [22949482](#).
38. Yang Z, dos Reis M. Statistical properties of the branch-site test of positive selection. *Molecular biology and evolution*. 2011; 28(3):1217–28. doi: [10.1093/molbev/msq303](#) PMID: [21087944](#).
39. Zhang J, Nielsen R, Yang Z. Evaluation of an improved branch-site likelihood method for detecting positive selection at the molecular level. *Molecular biology and evolution*. 2005; 22(12):2472–9. doi: [10.1093/molbev/msi237](#) PMID: [16107592](#).
40. Lu A, Guindon S. Performance of standard and stochastic branch-site models for detecting positive selection among coding sequences. *Molecular biology and evolution*. 2014; 31(2):484–95. doi: [10.1093/molbev/mst198](#) PMID: [24132121](#).
41. Gharib WH, Robinson-Rechavi M. The branch-site test of positive selection is surprisingly robust but lacks power under synonymous substitution saturation and variation in GC. *Molecular biology and evolution*. 2013; 30(7):1675–86. doi: [10.1093/molbev/mst062](#) PMID: [23558341](#); PubMed Central PMCID: PMC3684852.



42. Suzuki Y. False-positive results obtained from the branch-site test of positive selection. *Genes & genetic systems*. 2008; 83(4):331–8. PMID: [18931458](#).
43. Nozawa M, Suzuki Y, Nei M. Reliabilities of identifying positive selection by the branch-site and the site-prediction methods. *Proceedings of the National Academy of Sciences of the United States of America*. 2009; 106(16):6700–5. doi: [10.1073/pnas.0901855106](#) PMID: [19339501](#); PubMed Central PMCID: PMC2672471.
44. Sarkar D, Desalle R, Fisher PB. Evolution of MDA-5/RIG-I-dependent innate immunity: independent evolution by domain grafting. *Proceedings of the National Academy of Sciences of the United States of America*. 2008; 105(44):17040–5. doi: [10.1073/pnas.0804956105](#) PMID: [18971330](#); PubMed Central PMCID: PMC2579374.
45. Zou J, Chang M, Nie P, Secombes CJ. Origin and evolution of the RIG-I like RNA helicase gene family. *BMC evolutionary biology*. 2009; 9:85. doi: [10.1186/1471-2148-9-85](#) PMID: [19400936](#); PubMed Central PMCID: PMC2686710.
46. Zor T, Selinger Z. Linearization of the Bradford protein assay increases its sensitivity: theoretical and experimental studies. *Analytical biochemistry*. 1996; 236(2):302–8. Epub 1996/05/01. doi: [10.1006/abio.1996.0171](#) PMID: [8660509](#).
47. Abdiche Y, Malashock D, Pinkerton A, Pons J. Determining kinetics and affinities of protein interactions using a parallel real-time label-free biosensor, the Octet. *Analytical biochemistry*. 2008; 377(2):209–17. Epub 2008/04/15. doi: [10.1016/j.ab.2008.03.035](#) PMID: [18405656](#).
48. Pruitt KD, Tatusova T, Maglott DR. NCBI reference sequences (RefSeq): a curated non-redundant sequence database of genomes, transcripts and proteins. *Nucleic Acids Res*. 2007; 35(Database issue):D61–5. doi: [10.1093/nar/gkl842](#) PMID: [17130148](#); PubMed Central PMCID: PMC1716718.
49. Cunningham F, Amode MR, Barrell D, Beal K, Billis K, Brent S, et al. Ensembl 2015. *Nucleic Acids Res*. 2014. doi: [10.1093/nar/gku1010](#) PMID: [25352552](#).
50. Camacho C, Coulouris G, Avagyan V, Ma N, Papadopoulos J, Bealer K, et al. BLAST+: architecture and applications. *BMC Bioinformatics*. 2009; 10:421. Epub 2009/12/17. 1471-2105-10-421 [pii] doi: [10.1186/1471-2105-10-421](#) PMID: [20003500](#); PubMed Central PMCID: PMC2803857.
51. Salamov AA, Solovyev VV. Ab initio gene finding in Drosophila genomic DNA. *Genome research*. 2000; 10(4):516–22. PMID: [10779491](#); PubMed Central PMCID: PMC310882.
52. Finn RD, Bateman A, Clements J, Coggill P, Eberhardt RY, Eddy SR, et al. Pfam: the protein families database. *Nucleic Acids Res*. 2014; 42(Database issue):D222–30. doi: [10.1093/nar/gkt1223](#) PMID: [24288371](#); PubMed Central PMCID: PMC3965110.
53. Marchler-Bauer A, Zheng C, Chitsaz F, Derbyshire MK, Geer LY, Geer RC, et al. CDD: conserved domains and protein three-dimensional structure. *Nucleic Acids Res*. 2013; 41(Database issue):D348–52. doi: [10.1093/nar/gks1243](#) PMID: [23197659](#); PubMed Central PMCID: PMC3531192.
54. Letunic I, Doerks T, Bork P. SMART: recent updates, new developments and status in 2015. *Nucleic Acids Res*. 2014. doi: [10.1093/nar/gku949](#) PMID: [25300481](#).
55. Katoh K, Standley DM. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Molecular biology and evolution*. 2013; 30(4):772–80. doi: [10.1093/molbev/mst010](#) PMID: [23329690](#); PubMed Central PMCID: PMC3603318.
56. Edgar RC. MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res*. 2004; 32(5):1792–7. doi: [10.1093/nar/gkh340](#) PMID: [15034147](#); PubMed Central PMCID: PMC390337.
57. Shindyalov IN, Bourne PE. Protein structure alignment by incremental combinatorial extension (CE) of the optimal path. *Protein engineering*. 1998; 11(9):739–47. PMID: [9796821](#).
58. Guindon S, Dufayard JF, Lefort V, Anisimova M, Hordijk W, Gascuel O. New algorithms and methods to estimate maximum-likelihood phylogenies: assessing the performance of PhyML 3.0. *Syst Biol*. 2010; 59(3):307–21. doi: [10.1093/sysbio/syq010](#) PMID: [20525638](#).
59. Darriba D, Taboada GL, Doallo R, Posada D. ProtTest 3: fast selection of best-fit models of protein evolution. *Bioinformatics*. 2011; 27(8):1164–5. doi: [10.1093/bioinformatics/btr088](#) PMID: [21335321](#).
60. Le SQ, Gascuel O. An improved general amino acid replacement matrix. *Molecular biology and evolution*. 2008; 25(7):1307–20. doi: [10.1093/molbev/msn067](#) PMID: [18367465](#).
61. Anisimova M, Gascuel O. Approximate likelihood-ratio test for branches: A fast, accurate, and powerful alternative. *Syst Biol*. 2006; 55(4):539–52. Epub 2006/06/21. T808388N86673K61 [pii] doi: [10.1080/10635150600755453](#) PMID: [16785212](#).
62. Ronquist F, Teslenko M, van der Mark P, Ayres DL, Darling A, Höhna S, et al. MrBayes 3.2: efficient Bayesian phylogenetic inference and model choice across a large model space. *Syst Biol*. 2012; 61(3):539–42. doi: [10.1093/sysbio/sys029](#) PMID: [22357727](#); PubMed Central PMCID: PMC3329765.

63. Huelsenbeck JP, Ronquist F, Nielsen R, Bollback JP. Bayesian inference of phylogeny and its impact on evolutionary biology. *Science*. 2001; 294(5550):2310–4. Epub 2001/12/18. doi: [10.1126/science.1065889](https://doi.org/10.1126/science.1065889) 294/5550/2310 [pii]. PMID: [11743192](https://pubmed.ncbi.nlm.nih.gov/11743192/).
64. Yang Z. PAML 4: phylogenetic analysis by maximum likelihood. *Molecular biology and evolution*. 2007; 24(8):1586–91. doi: [10.1093/molbev/msm088](https://doi.org/10.1093/molbev/msm088) PMID: [17483113](https://pubmed.ncbi.nlm.nih.gov/17483113/).
65. Yang Z, Kumar S, Nei M. A new method of inference of ancestral nucleotide and amino acid sequences. *Genetics*. 1995; 141(4):1641–50. PMID: [8601501](https://pubmed.ncbi.nlm.nih.gov/8601501/); PubMed Central PMCID: PMC1206894.
66. Stamatakis A. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics*. 2014; 30(9):1312–3. doi: [10.1093/bioinformatics/btu033](https://doi.org/10.1093/bioinformatics/btu033) PMID: [24451623](https://pubmed.ncbi.nlm.nih.gov/24451623/); PubMed Central PMCID: PMC3998144.
67. Eswar N, Eramian D, Webb B, Shen MY, Sali A. Protein structure modeling with MODELLER. *Methods Mol Biol*. 2008; 426:145–59. Epub 2008/06/11. doi: [10.1007/978-1-60327-058-8\\_8](https://doi.org/10.1007/978-1-60327-058-8_8) PMID: [18542861](https://pubmed.ncbi.nlm.nih.gov/18542861/).
68. Rose PW, Bi C, Bluhm WF, Christie CH, Dimitropoulos D, Dutta S, et al. The RCSB Protein Data Bank: new resources for research and education. *Nucleic Acids Res*. 2013; 41(Database issue):D475–82. doi: [10.1093/nar/gks1200](https://doi.org/10.1093/nar/gks1200) PMID: [23193259](https://pubmed.ncbi.nlm.nih.gov/23193259/); PubMed Central PMCID: PMC3531086.
69. Shen MY, Sali A. Statistical potential for assessment and prediction of protein structures. *Protein science: a publication of the Protein Society*. 2006; 15(11):2507–24. doi: [10.1110/ps.062416606](https://doi.org/10.1110/ps.062416606) PMID: [17075131](https://pubmed.ncbi.nlm.nih.gov/17075131/); PubMed Central PMCID: PMC2242414.
70. Dolinsky TJ, Czodrowski P, Li H, Nielsen JE, Jensen JH, Klebe G, et al. PDB2PQR: expanding and upgrading automated preparation of biomolecular structures for molecular simulations. *Nucleic Acids Res*. 2007; 35(Web Server issue):W522–5. doi: [10.1093/nar/gkm276](https://doi.org/10.1093/nar/gkm276) PMID: [17488841](https://pubmed.ncbi.nlm.nih.gov/17488841/); PubMed Central PMCID: PMC1933214.
71. Baker NA, Sept D, Joseph S, Holst MJ, McCammon JA. Electrostatics of nanosystems: application to microtubules and the ribosome. *Proceedings of the National Academy of Sciences of the United States of America*. 2001; 98(18):10037–41. doi: [10.1073/pnas.181342398](https://doi.org/10.1073/pnas.181342398) PMID: [11517324](https://pubmed.ncbi.nlm.nih.gov/11517324/); PubMed Central PMCID: PMC56910.