



**UNIVERSIDADE FEDERAL DE UBERLÂNDIA**  
**INSTITUTO DE GENÉTICA E BIOQUÍMICA**  
**PROGRAMA DE PÓS-GRADUAÇÃO EM GENÉTICA E BIOQUÍMICA**



**Aspectos evolutivos das proteínas MyD88 e Tollip da via NF- $\kappa$ B**

**Aluno: Denis Prudencio Luiz**

**Orientador: Prof. Dr. Malcon Antônio Manfredi Brandeburgo**

**UBERLÂNDIA - MG**

**2015**



**UNIVERSIDADE FEDERAL DE UBERLÂNDIA**  
**INSTITUTO DE GENÉTICA E BIOQUÍMICA**  
**PROGRAMA DE PÓS-GRADUAÇÃO EM GENÉTICA E BIOQUÍMICA**



**Aspectos evolutivos das proteínas MyD88 e Tollip da via NF- $\kappa$ B**

**Aluno: Denis Prudencio Luiz**

**Orientador: Prof. Dr. Malcon Antônio Manfredi Brandeburgo**

**Tese apresentada à Universidade  
Federal de Uberlândia como parte  
dos requisitos para obtenção do  
Título de Doutor em Genética e  
Bioquímica (Área Genética)**

**UBERLÂNDIA - MG**

**2015**

Dados Internacionais de Catalogação na Publicação (CIP)  
Sistema de Bibliotecas da UFU, MG, Brasil.

---

L953a      Luiz, Denis Prudencio, 1985  
2015      Aspectos evolutivos das proteínas MyD88 e Tollip da via NF-KB /  
            Denis Prudencio Luiz. - 2015.  
            86 p.

            Orientador: Malcon Antônio Manfredi Brandeburgo.  
            Tese (doutorado) - Universidade Federal de Uberlândia, Programa  
de Pós-Graduação em Genética e Bioquímica.  
            Inclui bibliografia.

            1. Genética - Teses. 2. Bioinformática - Teses. 3. Sistema imune -  
Teses. I. Brandeburgo, Malcon Antônio Manfredi. II. Universidade  
Federal de Uberlândia. Programa de Pós-Graduação em Genética e  
Bioquímica. III. Título.

CDU: 577.1

---



**UNIVERSIDADE FEDERAL DE UBERLÂNDIA**  
**INSTITUTO DE GENÉTICA E BIOQUÍMICA**  
**PROGRAMA DE PÓS-GRADUAÇÃO EM GENÉTICA E BIOQUÍMICA**



**Aspectos evolutivos das proteínas MyD88 e Tollip da via NF- $\kappa$ B**

**ALUNO:** \_\_\_\_\_

**COMISSÃO EXAMINADORA**

**Presidente:** \_\_\_\_\_ (Orientador)

**Examinadores:**

\_\_\_\_\_  
\_\_\_\_\_  
\_\_\_\_\_  
\_\_\_\_\_

**Data da Defesa:** \_\_\_\_ / \_\_\_\_ / \_\_\_\_

**As sugestões da Comissão Examinadora e as Normas PGGB para o formato da Tese foram contempladas**

\_\_\_\_\_  
**(NOME DO ORIENTADOR)**

Dedico à minha família que através de muito trabalho possibilitou que eu pudesse trilhar o caminho do conhecimento.

Aos meus professores que me deram a oportunidade de fazer parte do mundo científico.

Aos meus amigos que permaneceram ao meu lado durante todo esse processo de aprendizagem.

E a todas as pessoas, que assim como eu, apreciam o conhecimento científico e acredita que através da ciência podemos nos conhecer e conhecer o mundo à nossa volta.

## **Agradecimentos**

Agradeço primeiramente aos meus pais, João de Deus e Clarice, e ao meu irmão Lelis sem os quais eu não conseguiria chegar onde cheguei.

Ao Prof. Dr. Malcon Antônio Manfredi Brandeburgo que foi meu orientador na monografia e no doutorado e abriu as portas do laboratório me dando a oportunidade de trilhar esse caminho.

Ao Prof. Dr. Warwick Estevam Kerr que foi meu orientador na iniciação científica.

Ao Prof. Dr. Carlos Ueira Vieira, que me orientou no mestrado e me inseriu na área da genética ligada ao sistema imune.

À Prof. Dra. Ana Maria Bonetti, que orientou a mim e aos meus amigos de laboratório durante todos nossos trabalhos, sendo uma segunda mãe para todos nós.

Agradeço aos meus amigos de laboratório (LABGEN), professores e alunos dos outros laboratórios, (Nanobiotecnologia) e (Labitox). Em especial à Juliana Franco, Célio Dias, Natália Melquie, Isabel Marques, Naiara Borges, Ana Carolina Dias, Romualdo Morandi, Edimar Olegário, Boscolli Pereira, Thais Ferreira, Patrícia Tieme, Emília Vaz, Galber Araujo, Zuleide Fátima, Gustavo Santos, Manuella Silverio, Annelise Serra, Luiza Diniz, Fernando Almeida, Roberta Barreto, Renato Mendonça, Vinícius Borges, Cássio Resende, Luana Scalia, Rafaela Cabral, Mariana Gonçalves e José Geraldo, pela ajuda científica e convívio diário que construiu o melhor ambiente de trabalho que já convivi.

Agradeço à equipe técnica e administrativa que sempre ajudou no suporte aos nossos trabalhos, Gerson Fraissat, Michelle Cristina, Tamiris Rodrigues, Ana Paula de Oliveira, Madson Caetano, Ismair Teodoro e Janaina Mota.

Agradeço aos meus amigos do dia a dia e parentes, Bruna Prudencio, Brenda Prudencio, Matheus Luiz, Bruno Luiz, Guilherme Diniz, Hugo César, Gabriel Rimoldi, Luciana Felipa, Guilherme Cunha, Fernando Fantim, Rebert Santos, Fernando Rabelo, João Paulo Aguiar, Marlon Buzart, Aline Duarte, Carla Duarte, Jefferson Loureiro, Juliana Souza, Maycon Mendes e Rhayenne Mendes.

## RESUMO

O sistema imune inato atua como a primeira linha de defesa de todos os animais. Ele possui células e moléculas que atuam no combate de diversos tipos de patógenos. As células de defesas possuem receptores de reconhecimento, as quais, desencadeiam reações para a produção de moléculas que irão combater os patógenos. A via NF- $\kappa$ B é uma via conservada nos animais, ela faz parte deste sistema e possui duas moléculas importantes para seu funcionamento, MyD88 e Tollip. O papel da proteína MyD88 é de acoplamento juntamente com outras proteínas, transmitindo o sinal na parte interna das células, enquanto Tollip funciona como um regulador negativo no final da mesma via. Essas duas proteínas tiveram suas sequências analisadas em diferentes espécies de animais, seus aspectos evolucionários foram estudados, dentre os domínios da proteína MyD88, e entre as sequências completa e as partes conservadas da Tollip. Utilizando ferramentas da bioinformática podemos ver que Tollip possui sucessivas modificações para otimizar sua estabilização acumulando resíduos alifáticos. Foram encontradas diferenças na estabilidade desta proteína entre primatas e artrópodes. Além de Tollip apresentar a necessidade de íon  $\text{Ca}^{2+}$  para manutenção de sua cavidade, foi possível identificar um ligando 768 provavelmente relacionado à inibição de Tollip. Na proteína MyD88 identificamos uma maior identidade nas sequências de primatas em relação às outras espécies analisadas, e foi reforçado a alta similaridade de alguns aminoácidos da interface com a funcionalidade da proteína. O domínio TIR demonstrou possuir maior identidade entre os domínios da proteína MydD88 e uma maior ancestralidade da sua sequência em relação à região DM, podendo sugerir uma relação com o maior número de parceiros de interação.

Palavras chave: Imune, MyD88, Tollip, Bioinformática, Domínios.

## ABSTRACT

The innate immune system plays as the first line of defense of all animals. It possesses cells and molecules that act in fighting many types of pathogens. The defense cells carry recognition receptors, which, trigger reactions for the production of molecules that will combat the pathogens. The NF- $\kappa$ B pathway is a pathway conserved in animals, it is part of this system and possesses two important molecules for its operation, MyD88 and Tollip. The role of MyD88 protein is coupling together with other proteins and transmit the signal inside the cell, while Tollip functions as a negative regulator at the end of the same pathway. These two protein sequences were analyzed in different animal species, their evolutionary aspects were studied among the domains of the MyD88 protein, and between the complete sequences and conserved parts of the Tollip. Using bioinformatics tools, we can see that Tollip has successive modifications to optimize its stabilization accumulating aliphatic residues. Differences were found between the stability of primates and arthropods proteins. In addition to the Tollip showing the need for ion  $\text{Ca}^{2+}$  for the maintenance of the cavity, it was possible to identify a ligand 768 probably related to inhibition of Tollip. In MyD88 protein we identified a greater identity in sequences of primates in relation to other species analyzed, also it was enhanced the high similarity of some amino acids interface with the functionality of the protein. The TIR domain has demonstrated greater identity between the fields of protein and a greater MydD88 ancestry of its sequence regarding the MD region, which may suggest a relationship with the largest number of interaction partners.

Keywords: Immune, MyD88, Tollip, Bioinformatics, Domains



## ÍNDICE DE ILUSTRAÇÕES

### CAPÍTULO 1

**FIGURA 1:** Diferença entre genes. Genes que sofrem duplicações são parálogos, genes que são encontrados em diferentes espécies e tem a mesma origem (homólogos) e a mesma função são denominados ortólogos. O conjunto dos ortólogos e parálogos que possuem a mesma origem são homólogos..... 4

**FIGURA 2:** Diferentes tipos de alinhamento..... 5

**FIGURA 3:** Matriz PAM com valores menores são mais indicadas para alinhar sequências menos divergentes e valores maiores para sequências mais divergentes, ao contrário da Matriz BLOSUM..... 6

**FIGURA 4:** Diagrama mostrando uma comparação entre a via NFkB em mamíferos e Dif/Relish em Drosophila, ativadas pelos receptores TLR e Toll respectivamente. Os homólogos entre as duas vias NFkB e Dif/Relish são apresentadas respectivamente, MyD88 e dMyD88, IRAK e Pelle, TRAF6 e dTRAF6, IKK e Cactus kinase, Ikb e Cactus..... 8

**FIGURA 5:** Diagrama dos receptores da família TLR e IL-R na via de sinalização de NF-κB. A via compartilha as mesmas proteínas Tollip, Myd88 e IRAK. MyD88 inicia a sinalização da via com a ligação entre os receptores e IRAK. Tollip inibe a sinalização da via se ligando ao IRAK..... 11

### CAPÍTULO 2

**FIGURA 1:** Molecular Phylogenetic analysis by Maximum Likelihood method. The evolutionary history of ToL (A), complete Tollip protein (B) and G-block cured Tollip protein (C) are shown. Trees B and C are drawn to scale, with branch lengths measured in the number of substitutions per site. The analysis involved 36

amino acid sequences. All positions containing gaps and missing data were eliminated. Group sorting was made in roman numerals (I - vertebrata, II - echinodermata, III - arthropod, IV - cnidaria, V - porifera) and the subgroups were coded in arabic numerals (1 - primates, 2 - carnivora, 3 - rodentia, 4 - bovidae, 5 - equidae 6 - marsupialia 7 - monotremata, 8 - birds, 9 - reptilia, 10 - amphibia, 11 - actinopterygii, 12 - ascidiacea, 13 - echinoidea, 14 - hexapoda, 15 - arachnida, 16 - hydrozoa, 17 - demospongiae)..... 39

**FIGURA 2:** Variability of Tollip protein.

The variability residue per residue measured with Shannon Index (A). Protein conserved residues are disposed at (B), with variable positions as ".", these positions reflect some essential molecular arrangement to Tollip function. All analyses were made with Tollip G-block cured, to avoid the gaps and non-informative parsimony site..... 40

**FIGURA 3:** Human (hsa) Tollip tertiary structure.

This structure was modeled at I-Tasser server, using default sets. It could be seen the aliphatic residues distribution along all sequence (A and B). The ligands are shown arranged at lateral chain of the right residues, Ca<sup>2+</sup> (C) and 768 (D)..... 41

**FIGURA 4:** Splits graph of complete Tollip protein alignment.

The parsimony uninformative sites, gaps and constant sites were excluded. There were 1000 pseudoreplicates performed as bootstraps to support the derivations, it was used as ProteinML distance correction the model mtREV24. Green operational taxonomic units represents arthropods groups, blue taxa represent primates and black represent the other groups..... 42

**FIGURA 5:** Recombinational events involved with Tollip evolution.

Each sequence are represented by a color and the recombination is evidenced by donor. All analyses were evaluated with RDP and the most significant P value to support the findings are shown at Table 3..... 43

**FIGURA 6:** Positive selections operating in each codon of Tollip, evidenced by Selecton algorithm. There were used two models which were evaluated separately, M8 and M8a, where the last is referenced as null model..... 44

### CAPÍTULO 3

**FIGURA 1:** Disposição das porções DD, INT e TIR na molécula MyD88..... 55

**FIGURA 2:** Heat Map da sequência completa (A), domínio egião DD (B), domínio INT (C) e domínio TIR (D)..... 60

**FIGURA 3:** BoxPlot da distribuição de identidade entre a sequência e regiões. O grafico mostra a maior identidade entre os dominios TIR, entre (60 e 100%), seguido da sequência completa (40 a 100%), domínio DD (45% a 100%) e a região INT 40% e 100% mais alguns outliers em preto..... 61

**FIGURA 4:** Análise Filogenética Molecular pelo método de Máxima Verossimilhança. Histórico evolutivo de ToL (A), Região DD da proteína MyD88 (B) e Região TIR da proteína TIR (C) são mostrados. Árvores B e C são construídas em escala, com medidas de comprimento de ramos em número de substituição por sítios. As análises envolveram 46 sequências de aminoácidos. Todas posições que continham intervalos e dados faltante foram eliminados. Os grupos foram classificados em números romanos (I - marsupialia, II - eutheria, III - reptilia, IV - amphibia, V - coelacanthimorpha, VI - neopterygii, VII - chondrichthyes) e os subgrupos foram classificados em números arábicos (1 - dasyuromorphia, 2 - logomorpha, 3 - rodentia, 4 - primatas, 5 - scandentia, 6 - chiroptera, 7 - carnivora, 8 - suidae, 9 - ruminantia, 10 - cetacea, 11 - perissodactyla, 12 - testudines, 13 - neornithes, 14 - crocodylomorpha, 15 - amphibia, 16 - coelacanthimorpha, 17 - neopterygii, 18 - chondrichthyes)..... 62

## ÍNDICE DE TABELAS

### CAPÍTULO 2

<b>TABELA 1:</b> Tollip downloaded reference data from KEGG and principal protein features.....	35
---	----

<b>TABELA 2:</b> Correlation between main assessed protein characteristics of Tollip.....	36
---	----

<b>TABELA 3:</b> Potential recombinant events identified in Tollip with RDP.....	37
--	----

<b>TABELA 4:</b> Results from a test of molecular clocks using the Maximum Likelihood method.....	38
---	----

### CAPÍTULO 3

<b>TABELA 1:</b> Sigla, nome das espécies e número no banco de dados das espécies analisadas.....	58
---	----

<b>TABELA 2:</b> Valores de bootstraps dos grupos formados na árvore B (DD) e C (TIR). <sup>a</sup> mmu, rno, cge. <sup>b</sup> mmu, rno. <sup>c</sup> apla, gga, mgp, tgu, phi. <sup>d</sup> apla, gga, mgp. <sup>e</sup> tgu, phi.....	63
--	----

<b>TABELA 3:</b> Resíduos alinhados através da ferramenta T-coffee Expresso. Sítios com grande importância no funcionamento dos domínios DD e TIR de MyD88. Os resíduos em negro foram conservado em todas sequências ao longo da evolução, em vermelho estão os resíduos que sofreram substituições no processo evolutivo.....	65
---	----

## LISTA DE ABREVIATURAS E SIGLAS

<b>AD</b>	<i>Atopic Dermatitis</i>
<b>AI</b>	<i>Aliphatic Index</i> , Índice Alifático
<b>AIC</b>	<i>Akaike Information Criterion</i> , Critério de Informação de Akaike
<b>Ala</b>	Alanina
<b>Arg</b>	Argenina
<b>Asp</b>	Asparagina
<b>BIC</b>	<i>Bayesian information criterion</i> , Critério de Informação Bayesiano
<b>BLOSUM</b>	<i>Blocks of Amino Acid Substitution Matrix</i> , Matriz de Substituição de Aminoácidos por Blocos
<b>BS Score</b>	<i>Binding Score</i> , Pontuação de Ligação
<b>C2</b>	<i>Core Domain 2</i> , Domínio do Núcleo 2
<b>Ca<sup>2+</sup></b>	Íon Cálcio 2+
<b>CUE</b>	<i>Coupling of Ubiquitin to Endoplasmic Reticulum Degradation</i> , Acoplamento de Ubiquitina associado à Degradação de proteína do Retículo Endoplasmático
<b>DI; INT</b>	Domínio Intermediário
<b>Dif/Relish</b>	<i>Dorsal-related immunity factor and Relish</i>
<b>DM</b>	Domínio de Morte
<b>DNA</b>	Ácido Desoxirribonucleico
<b>F81</b>	Modelo Felsenstein 1981
<b>Gln</b>	Glutamina
<b>Glu</b>	Glutamato
<b>Gly</b>	Glicina
<b>GRAVY</b>	<i>Grand Average of Hydropathicity</i> , Índice de Hidropaticidade Média
<b>I<math>\kappa</math>B</b>	<i>Inhibitor of kappa B</i> , Inibidor de kappa B
<b>IKK</b>	<i>I<math>\kappa</math>B kinase</i> , I $\kappa$ B quinase
<b>IL-1RAcP</b>	Proteína Associada ao IL-1R
<b>IL-R; 1; 18; 1<math>\beta</math></b>	Receptor de Interleucina; 1; 18 e 1 $\beta$

<b>In/del</b>	Inserção/deleção
<b>ININ</b>	<i>Instability Index</i> , Índice de Instabilidade
<b>IRAK1; 2; 4</b>	L-1 Receptor Associado a Quinase; 1; 2 e 4
<b>JC</b>	Matriz Jukes e Cantor
<b>JCK</b>	Quinase c-Jun N-Terminal
<b>JTT</b>	Matriz Jones-Taylor-Thornton
<b>K80</b>	Matriz Kimura 1980
<b>KEGG</b>	<i>Kyoto Encyclopedia of Genes and Genomes</i> , Enciclopédia Kyoto de Genes e Genoma
<b>lnL</b>	<i>log-likelihood</i> , log de verossimilhança
<b>LPS</b>	Lipopolissacarídeos
<b>Lys</b>	Lisina
<b>M8; M8a</b>	<i>Positive selection beta model</i> , Modelo beta de seleção positiva
<b>M8a</b>	<i>Nearly-neutral beta model</i> , Modelo beta quase neutro
<b>MAPK</b>	Proteína Quinase Ativada por Mitógeno
<b>mRNA</b>	RNA mensageiro
<b>mtREV24</b>	General Reversible Mitochondrial 24
<b>MW</b>	<i>Molecular Weight</i> , Peso Molecular
<b>Myb1</b>	Myeloblastosis 1, Mieloblastose 1
<b>MyD88</b>	Gene 88 da resposta primária de Diferenciação Mieloide
<b>MyD88s</b>	Splice variante de MyD88
<b>NCBI</b>	<i>National Center for Biotechnology Information</i> , Centro Nacional para Informação Biotecnológica
<b>NF-κB</b>	<i>Factor Nuclear kappa B</i> , Fator Nuclear kappa B
<b>PAM</b>	Porcentagem de Mutação Aceita
<b>PAMPs</b>	Padrões Moleculares Associados à Patógenos
<b>pI</b>	Isoelectric point, ponto Isoelétrico
<b>PI3</b>	Fosfatidilinositol-3
<b>PIP3</b>	Fosfatidilinositol-tri-fosfato
<b>Pro</b>	Prolina
<b>PtdIns (3,4,5)P</b>	Fosfatidilinositol 3,4,5-trifosfato
<b>RMSD</b>	<i>Root-Mean-Square Deviation</i> , Desvio Da Raiz Quadrada

	Média
<b>RNA</b>	Ácido Ribonucleico
<b>Ser</b>	Serina
<b>SNP</b>	<i>Single-Nucleotide Polymorphism</i> , Polimorfismo de Nucleotídeo Único
<b>SSCP</b>	<i>Single-Strand Conformation Polymorphism</i> , Polimorfismo Conformacional de Fita Simples
<b>TB</b>	<i>Tuberculosis</i> , Tuberculose
<b>TDB</b>	Domínio Ligado a Tom1
<b>TIR</b>	Receptor Toll/IL-1
<b>TLR; 2; 3; 4</b>	<i>Toll Like Receptor</i> , Receptor Tipo Toll; 2; 3 e 4
<b>TM-Score</b>	<i>Template Modeling Score</i> , Modelo de Pontuação de Modelagem
<b>ToL</b>	<i>Tree of Life</i> , Árvore da Vida
<b>Tollip</b>	<i>Toll-interacting protein</i> , Proteína de Interação Toll
<b>Tom1</b>	<i>Target of Myb protein 1</i> , Proteína alvo de Myb 1
<b>TRAF6; dTRAF6</b>	<i>TNF receptor associated factor 6</i> , Receptor TNF associado ao fator 6; <i>drosophila</i> TRAF6
<b>Uniprot</b>	<i>Universal Protein Resource</i> , Banco de Dados Universal de Proteínas
<b>UTR</b>	<i>Untranslated Region</i> , Região não Traduzível

# ÍNDICE

APRESENTAÇÃO .....	1
CAPÍTULO 1 .....	3
FUNDAMENTAÇÃO TEÓRICA.....	3
BIOINFORMÁTICA E FILOGENIA .....	3
EVOLUÇÃO E SISTEMA IMUNE .....	7
MyD88: molécula de acomplamento da via TLR/NF-kB .....	9
TOLLIP: molécula de regulação negativa da via TLR/NF-kB.....	12
REFERÊNCIAS .....	14
CAPÍTULO 2 .....	21
Tollip or Not Tollip: What Are the Evolving Questions behind It? .....	21
ABSTRACT .....	21
INTRODUCTION .....	22
MATERIAL AND METHODS .....	24
RESULTS AND DISCUSSION .....	26
CONCLUSIONS .....	33
ACKNOWLEDGMENTS .....	34
REFERENCES .....	45
CAPÍTULO 3 .....	54
Aspectos evolutivos da proteína MyD88 e seus domínios .....	54
RESUMO .....	54
INTRODUÇÃO.....	55
MATERIAIS E MÉTODOS .....	57
RESULTADO E DISCUSSÃO .....	59
CONCLUSÃO .....	67
REFERÊNCIAS .....	68



## APRESENTAÇÃO

---

A ciência avança a cada dia em direção à luz do conhecimento. A tecnologia ligada aos conhecimentos biológicos possibilitou responder antigas questões, e o entendimento das relações dos organismos vivos foi uma delas. Com a formulação da teoria da evolução surgiram mais perguntas a serem respondidas pela ciência. Por meio de estudos conseguimos relacionar aspectos naturais à padrões matemáticos, possibilitando, organizar dados, compartilhar informações, responder dúvidas e lançar questionamentos relacionados à nossa existência.

Um aspecto importante à sobrevivência das espécies está ligada a uma boa adaptação do organismo ao meio que vive, uma característica que permite a continuidade da vida nos organismos é a presença de um sistema de defesa. O sistema imunológico está presente nos seres vivos, o que possibilita a interação entre eles e o meio que estão inseridos com o principal propósito de identificar o que é prejudicial ou não a esse organismo. O sistema imune inato é presente tanto nos vertebrados quanto em invertebrados. Este sistema possui características comuns em espécies diferentes, como genes ortólogos, indicando uma ancestralidade comum deste sistema no processo evolutivo.

A via NF- $\kappa$ B ativada pelos receptores das famílias TLR e IL-R se mostrou de grande importância nesse sistema. Dentre as proteínas dessa via existem duas moléculas de suma importância para o seu funcionamento. Uma denominada MyD88 que se liga aos receptores e permite que o sinal seja passado para dentro da célula e inicie o processo de resposta imune. A outra denominada Tollip que inibe a mesma sinalização da via, impedindo que a célula continue respondendo ao estímulo, evitando o desenvolvimento de uma doença autoimune.

Essas duas proteínas já foram encontradas em diversas espécies com a ajuda de sequenciamento genômico e suas informações estão disponíveis em diversos bancos de dados. Com o avanço dos diversos programas de análise na Bioinformática, mas dados estão disponíveis, contudo, a velocidade da análise dos dados não segue o mesmo padrão.

Esse trabalho teve como objetivo principal analisar esses dados disponíveis, das duas proteínas importantes ao sistema imune, utilizando programas que são acessíveis à comunidade científica (programas gratuitos), trabalhando esses dados e identificando aspectos ainda não estudados.

## **CAPÍTULO 1**

---

### **FUNDAMENTAÇÃO TEÓRICA**

#### **BIOINFORMÁTICA E FILOGENIA**

Depois das descobertas de Darwin sobre a evolução das espécies surgiram dúvidas, sendo que uma delas seria entender o processo de especiação. As informações obtidas de fósseis e de espécies já conhecidas permitiram a construção de padrões entre várias metodologias, com o objetivo de comparar a homologia entre os organismos (HALL, 2001).

Na década de 60 e 70 houve uma grande disputa entre os pesquisadores na área da cladística e fenética sobre os processos evolutivos, os quais tinham os propósitos de estipular o melhor modelo de entendimento e explicar como essa evolução acontecia entre as espécies, ambos apresentavam metodologias que ajudavam na compreensão sobre o processo de especiação, mas as explicações não eram totalmente satisfatórias (FELSENSTEIN; FELSENSTEIN, 2004).

Margoliash (1963), em suas pesquisas analisou uma única proteína presente em todos organismos já estudados, a qual foi utilizada para a reconstrução de uma filogenia completa, o citocromo C, proteína envolvida na transferência de elétrons. As similaridades e diferenças foram analisadas por métodos estatísticos e a homologia também foi comprovada usando esses critérios. Utilizando desses dados, Fitch foi um dos primeiros a trabalhar com programas de computador nessa área e o mais influente no desenvolvimento desses métodos. Construiu uma filogenia molecular e estipulou a mínima distância de mutações, estas relacionadas à similaridade entre dois citocromos, desse modo fazendo uma comparação par a par dos aminoácidos homólogos (FITCH; MARGOLASH, 1967).

Dentre todos os dados moleculares que poderiam servir como caracteres evolutivos, as sequências de DNA, RNA e proteínas foram as mais validadas durante os estudos. No final da década de 80 e início de 90 ocorreram várias tentativas para avaliar inferências filogenéticas com hibridização de DNA

(TEMPLETON, 1985; 1986; FELSENSTEIN, 1987; MARKS; SCHMID; SARICH, 1988; SIBLEY; COMSTOCK; AHLQUIST, 1990; DICKERMAN, 1991).

Na reconstrução filogenética são utilizadas sequências ortólogas, classificadas por apresentarem a mesma função em espécies diferentes, e por possuir a mesma origem evolutiva são denominadas como homólogas (Figura 1). Para se realizar estes estudos de evolução de sequências sabe-se hoje a importância em se fazer um alinhamento entre elas. (FITCH, 1970; 1971; 2000).

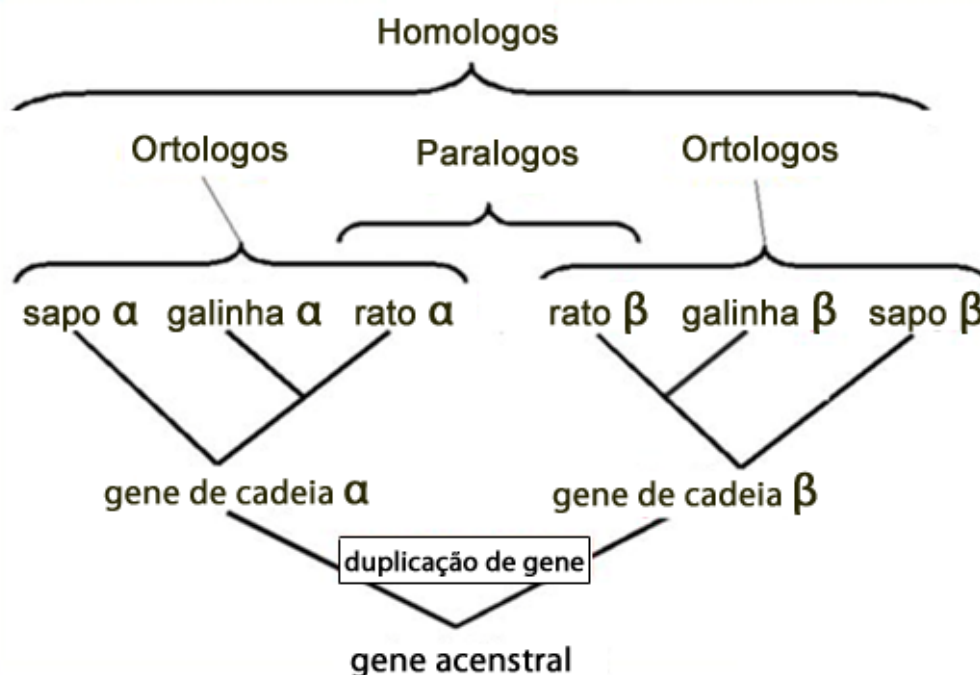


Figura 1: Diferença entre genes. Genes que sofrem duplicações são parálogos, genes que são encontrados em diferentes espécies e tem a mesma origem (homólogos) e a mesma função são denominados ortólogos. O conjunto dos ortólogos e parálogos que possuem a mesma origem são homólogos.

Fonte: Modificado de <http://www.virology.ws/2010/07/27/viral-bioinformatics-introduction-homology/>

O alinhamento permite o pesquisador identificar os sítios de nucleotídeos ou aminoácidos idênticos, não idênticos e faltantes. Ele foi introduzido de forma computacional no início da década de 80, mas não era utilizado em todas as análises até na década de 90, quando foram estipulados diferentes tipos de alinhamento. O primeiro, o alinhamento Global, visa todo o comprimento da sequência, possibilitando assim uma busca pela ancestralidade comum. O segundo o alinhamento Local, o qual contempla apenas uma parte das

sequências, muito utilizado na busca de sequências em banco de dados. E posteriormente um terceiro denominado alinhamento Semi-global, uma mistura dos dois métodos anteriores, nos quais as extremidades das sequências não são avaliadas no alinhamento (THORNE; KISHINO; FELSENSTEIN, 1991; HENIKOFF; HENIKOFF, 1992; BRUDNO *et al.*, 2003), estes 3 alinhamentos podem ser visto na figura 2.

1	2	3	4	5	6	7	8	9	10	11
C	G	T	C	C	G	A	A	G	T	G
			.							
★	★	T	A	C	G	A	A	★	★	★

(a) Alinhamento Global

3	4	5	6	7	8
T	C	C	G	A	A
	.				
T	A	C	G	A	A

(b) Alinhamento Local

1	2	3	4	5	6	7	8
C	G	T	C	C	G	A	A
			.				
★	★	T	A	C	G	A	A

(c) Alinhamento Semi-Global

Figura 2: Diferentes tipos de alinhamento.

Fonte: Modificado de [https://www.researchgate.net/figure/260376769\\_fig2\\_Global-local-and-semi-global-alignment-The-global-local-and-semi-global-alignments](https://www.researchgate.net/figure/260376769_fig2_Global-local-and-semi-global-alignment-The-global-local-and-semi-global-alignments).

Na realização de um alinhamento há uma sobreposição entre sítios (aminoácidos ou nucleotídeos) das sequências, o que permite a identificação de caracteres similares, diferentes ou gaps (intervalos) quando há a necessidade de acrescentá-los para uma melhor análise, para todos esses casos é definido uma pontuação em cada situação no cálculo do método. Esses cálculos são criados de acordo com matrizes de substituição, as quais atribuirão valores para a seleção na escolha do melhor alinhamento (THORNE; KISHINO; FELSENSTEIN, 1991).

Ao longo do tempo essas matrizes de substituição foram analisadas, validadas e reformuladas para uma melhor otimização do processo de alinhamento. No caso de sequências de aminoácidos foram levados em

consideração diferentes características como: frequência de substituições em proteína homóloga, sua propriedade hidrofílica, hidrofóbica (ALTSCHUL, 1991; MCLACHLAN, 1972), frequência ao se manter o mesmo aminoácido no processo evolutivo (formulando assim uma Porcentagem de Mutação Aceita, PAM) (DAYHOFF; SCHWARTZ, 1978) e para pequenos blocos de sequências uma matriz mais específica chamada BLOSUM (HENIKOFF & HENIKOFF, 1992) assim como mostrado na figura 3.

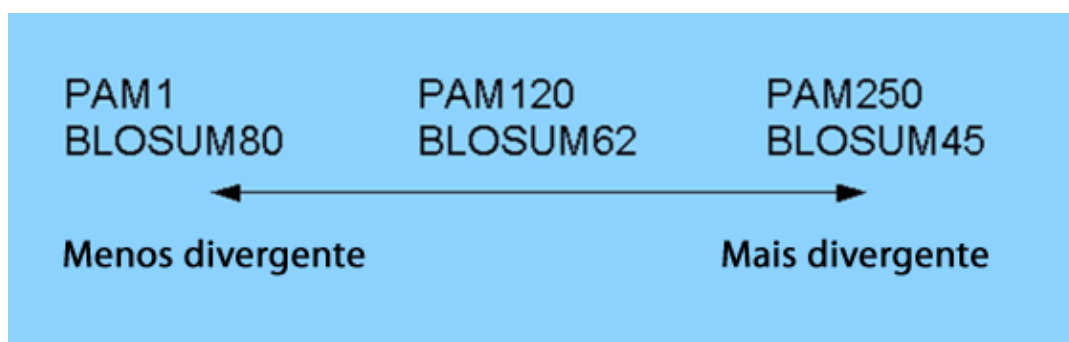


Figura 3: Matriz PAM com valores menores são mais indicadas para alinhar sequências menos divergentes e valores maiores para sequências mais divergentes, ao contrário da Matriz BLOSUM. Fonte: Modificado de [http://www.clcsupport.com/clcgenomicsworkbench/650/Use\\_scoring\\_matrices.html](http://www.clcsupport.com/clcgenomicsworkbench/650/Use_scoring_matrices.html).

Jukes e Cantor, 1969, foram os primeiros a estipular um modelo de substituição para nucleotídeos (JC), o qual determinava a mesma frequência para a substituição de um nucleotídeo por outro. Depois houve uma mudança dessa análise ao considerar o aspecto de transição, mudanças de bases com a mesma natureza química entre purinas, Adenina e Guanina, e entre pirimidinas, Citosina e Timina. As transversões como menos comuns, que é a troca de bases purinas por pirimidinas e vice-versa (KIMURA, 1980; COLLINS, JUKES, 1994).

Kimura, 1980, apresentou o modelo de dois parâmetros baseados nessas substituições (K80), o qual estipula valores diferentes para as duas situações. O mesmo foi feito por Felsenstein, 1981, criando o modelo (F81) baseado no JC, o qual leva em consideração as propriedades físicoquímicas e de seleção natural dos nucleótidos

## EVOLUÇÃO E SISTEMA IMUNE

Ao analisarmos uma ameba, percebemos que ela possui um sistema de distinção entre um microorganismo que pode servir como alimento, e uma outra ameba, assim podemos inferir que ela possua receptores específicos para essa distinção. E esse papel lembra o mesmo desempenhado pelos macrófagos em todos vertebrados e muitos invertebrados, os quais possuem células fagocitárias que são transportadas pelo sistema vascular e tecidual (JANEWAY *et al.*, 2001; ESCOLL *et al.*, 2013).

As funções dos genes responsáveis pela ativação das via do NF- $\kappa$ B (Fator de Transcrição Nuclear kappa B) ativadas pelo receptor Toll, foram evidenciadas por análises genômicas, supondo uma evolução dos processos de defesa em ancestrais comuns que pode ter ocorrido na divergência entre animais e plantas, as sequências de DNA referente às essas vias são encontradas tanto em vertebrados, invertebrados e plantas.

Após a identificação do receptor Toll em *Drosophila*, essas pesquisas se estenderam à ratos, encontrando o mesmo tipo de receptor neles, o Toll Like Receptor (TLR). Ratos deficientes para a produção desse receptor apresentaram dificuldade na identificação e na resposta aos Lipopolissacarídeos (LPS) bacterianos, que fazem parte dos Padrões Moleculares Associados à Patógenos (PAMPs). Essa deficiência estava ligada à porção do domínio TIR (Receptor Toll/IL-1) do receptor TLR4. O resultado da ineficiência dessa via levou a falta de resposta imune à bactérias gram-negativas.

Estudos em moscas da fruta mostraram uma especificidade de diferentes mutantes do Toll em relação a determinados tipos de LPS. A comparação entre as proteínas relacionadas à essa via NF- $\kappa$ B, ativada pelo TLR, entre mamíferos e *Drosophila*, está representada na figura 4 (ESCOLL *et al.*, 2013).

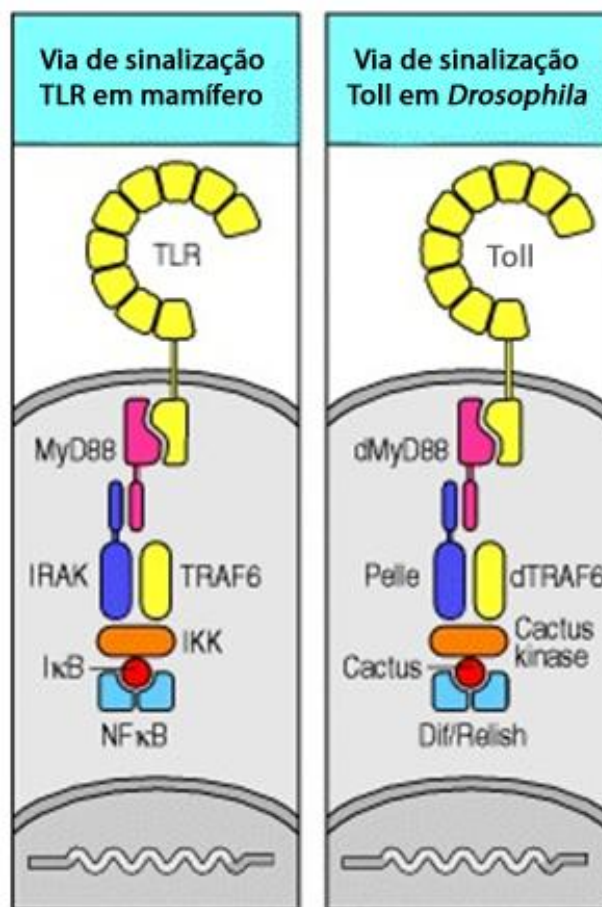


Figura 4: Diagrama mostrando uma comparação entre a via NFκB em mamíferos e Dif/Relish em *Drosophila*, ativadas pelos receptores TLR e Toll respectivamente. Os homólogos entre as duas vias NFκB e Dif/Relish são apresentadas respectivamente, MyD88 e dMyD88, IRAK e Pelle, TRAF6 e dTRAF6, IKK e Cactus kinase, IκB e Cactus.

Fonte: Modifica de ESCOLL *et al.*, 2013.



## **MyD88: molécula de acoplamento da via TLR/NF- $\kappa$ B**

MyD88 foi primeiramente identificado como gene 88 da resposta primária de Diferenciação Mieloide (LORD; HOFFMAN-LIEBERMANN; LIEBERMANN, 1990). Outro estudo mostrou homologia entre essa proteína e domínios citoplasmáticos dos receptores Toll em *Drosophila* e IL-1R (Receptor de Interleucina 1) em mamíferos (HULTMARK, 1994) e posteriormente foi classificada como uma proteína adaptadora da via de sinalização do IL-1R (Receptor de Interleucina 1) (MUZIO *et al.*, 1997; WESCHE *et al.*, 1997).

MyD88 possui na porção C-terminal um domínio TIR homólogo à porção citoplasmática TIR dos receptores Toll e IL-1R. Na porção N-terminal da proteína se encontra o Domínio de Morte (DM) homóloga com o domínio de mesmo nome presente nas proteínas Pelle em *Drosophila* e IRAK1 e 2 (IL-1 Receptor Associado a Quinase) em mamíferos (MUZIO *et al.*, 1997; WESCHE *et al.*, 1997), figura 3. O papel dessa proteína IRAK em *Drosophila* é desempenhado pela proteína Pelle na via do receptor Toll (GROSSHANS *et al.*, 1994; LETSOU *et al.*, 1991; NORRIS; MANLEY, 1996).

Posteriormente, outro estudo analisou ratos com deleção do gene MyD88, este trabalho demonstrou a necessidade dessa proteína na proliferação de células T, produção de citocinas, produção de proteínas de fase aguda na resposta de IL-1 $\beta$  *in vivo* e indução das vias NF- $\kappa$ B e MAPK (Proteína Quinase Ativada por Mitógeno) em resposta ao estímulo no receptor IL-18R (ADACHI *et al.*, 1998).

Outro estudo, utilizando ratos deficientes para MyD88 confirmaram sua importância como um adaptador crítico na sinalização induzida pelos receptores das superfamílias IL-1R, IL-18R e Toll (KAWAI *et al.*, 1999). Ratos sem MyD88 mostraram maior suscetibilidade à infecção por patógenos ou parasitas (TAKEUCHI; HOSHINO; AKIRA, 2000; SCANGA *et al.*, 2002). Com o avanço de pesquisas ligadas à genética de insetos e bioinformática possibilitaram o entendimento na área de imunologia de mamífero e a descoberta de outras proteínas da família IL-R (MITCHAM *et al.*, 1996; O'NEILL, 2008).

Em alguns estudos foram observados que a super-expressão do DM pode levar a uma ativação espontânea de NF- $\kappa$ B e JCK (Quinase c-Jun N-Terminal)

enquanto TIR foi considerado como um dominante negativo do processo (MUZIO; FENG; DIXIT, 1997; MEDZHITOV *et al.* 1998, BURNS *et al.*, 2000). Entre esses dois domínios existe uma sequência determinada como INT ou DI (Domínio Intermediário) importante para a ativação de IRAK4 (AVBELJ; HORVAT; JERALA, 2011). Em outro estudo foi analisado o MyD88s, um splice variante que não possui o DI, o qual atua como um dominante negativo na via (JANSSENS *et al.*, 2003, BURNS *et al.*, 2003).

Quando analisado em humanos, as porções DM e DI formam um dímero em solução, e a porção TIR continua como um monômero livre que interage com outros domínios TIRs. Tanto em MyD88 quanto nas outras proteínas que também possuem o domínio TIR foi observado 5 cadeias centrais de folha- $\beta$ , cercado por 4  $\alpha$ -hélice, com uma maior semelhança com a mesma porção no TLR2. O domínio apresenta três boxes importantes, 1 (F/Y)DA, 2 (RDXXP), and 3 (FW). O box2 é um loop, composto por 10 resíduos de aminoácidos altamente conservados dentre as proteínas que possuem o mesmo domínio, incluindo uma Arg, um Asp, dois resíduos hidrofóbicos, seguidos de uma Pro e Gly. Esta estrutura impede que a porção TIR forme dímero com ele mesmo de forma independente. Essa suposta fraqueza de interação permite mudanças conformacionais aumentando o reconhecimento de ligantes de TLR trazendo domínios TIR mais próximos, permitindo o início da sinalização pelo MyD88 (Figura 5) (BURNS *et al.*, 2003; KHAN *et al.*, 2004; OHNISHI *et al.*, 2009; DEGUINE; BARTON, 2014).

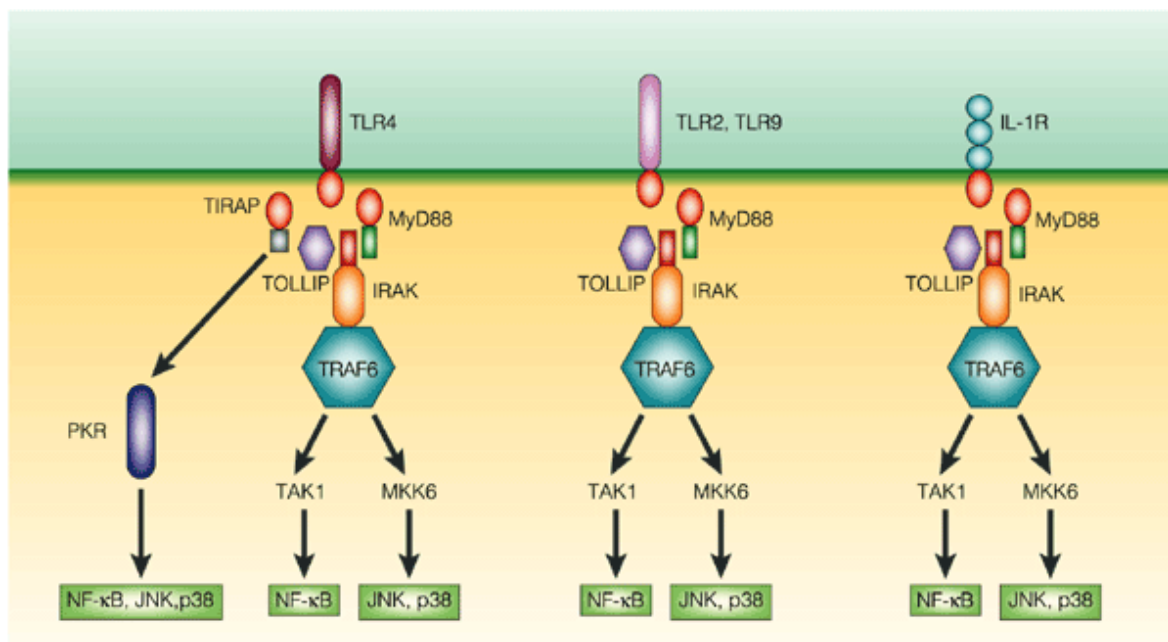


Figura 5: Diagrama dos receptores da família TLR e IL-R na via de sinalização de NF-κB. A via compartilha as mesmas proteínas Tollip, MyD88 e IRAK. MyD88 inicia a sinalização da via com a ligação entre os receptores e IRAK. Tollip inibe a sinalização da via se ligando ao IRAK.  
Fonte: MEDZHITOV, 2001.

## **TOLLIP: molécula de regulação negativa da via TLR/NF-κB**

A proteína Tollip (Proteína de Interação Toll), composta de 274 aminoácidos, foi identificada através de um experimento em levedura utilizando o homólogo IL-1RAcP (Proteína Associada ao IL-1R) como uma molécula de recrutamento. Por meio de análise de conservação foi identificado o domínio C2 (domínio do Núcleo 2) entre os aminoácidos 54-186 da Tollip, com o estímulo de receptores da família IL-R foi possível avaliar que a Tollip participava da via que era ativada por esses receptores. Neste mesmo estudo a proteína Tollip mostrou interagir com a proteína IRAK, sendo a autofosforilação de IRAK bloqueada pelo complexo Tollip-IRAK, dando a Tollip a característica inibitória (BURNS *et al.*, 2000).

Além da via estimulada pelos receptores da família IL-R, Tollip também regula negativamente a via de sinalização ativada por receptores da família TLR (Figura 5), demonstrando uma correlação de Tollip com a ativação de NF-κB e indícios de uma fosforilação do próprio Tollip (ZHANG; GHOSH, 2002). Esta fosforilação está relacionado a um segundo domínio CUE (acoplamento de ubiquitina conjugada à degradação do retículo endoplasmático) localizado na porção C-terminal nos aminoácidos 230–270 (LI *et al.*, 1999; LI *et al.*, 2001). Tollip possui também um terceiro domínio na posição N-terminal, denominado TDB (Domínio Ligado a Tom1) (KATOH *et al.*, 2004).

Estudos realizados com ratos sem Tollip indicaram que a proteína tem um papel importante na sinalização próinflamatória (DIDIERLAURENT *et al.*, 2006). O domínio C2 em proteínas que apresentam o mesmo domínio tem a função de ligação do cálcio, enquanto este domínio na Tollip, se liga ao Fosfatidilinositol-trifosfato (PIP3) e ao Fosfatidilinositol 3,4,5-trifosfato [PtdIns (3,4,5)P] que são proteínas encontradas na membrana plasmática das células (LI *et al.*, 2004). Essas duas proteínas, PIP3 e PtdIns (3,4,5)P, são produtos da PI3 quinase que estão ligadas a eventos de sinalização múltipa e de tráfego (MARTIN, 1998).

O domínio CUE interage com receptores da via de sinalização enquanto o domínio TBD, através de sua super-expressão mostrou diminuir a ativação de fatores de transcrição de NF-κB (YAMAKAMI; YOSHIMORI; YOKOSAWA, 2003; BRISSONI *et al.*, 2006).

Existem isoformas da proteína Tollip que já foram estudadas demonstrando que a diversificação da sinalização imunológica é importante para o fitness evolutivo das espécies, podendo contribuir para diversificação da sinalização de macrófagos e de respostas à infecções de fase aguda (LO *et al.*, 2009).

## REFERÊNCIAS

ADACHI, O.; KAWAI, T.; TAKEDA, K.; MATSUMOTO, M.; TSUTSUI, H.; SAKAGAMI, M.; NAKANISHI K.; AKIRA, S. Targeted disruption of the MyD88 gene results in loss of IL-1-and IL-18-mediated function. **Immunity**, v. 9, n. 1, p. 143-150, 1998.

ALTSCHUL, S. F. Amino acid substitution matrices from an information theoretic perspective. **Journal of molecular biology**, v. 219, n. 3, p. 555-565, 1991.

AVBELJ, M.; HORVAT, S.; JERALA, R. The role of intermediary domain of MyD88 in cell activation and therapeutic inhibition of TLRs. **The Journal of Immunology**, v. 187, n.5, p. 2394-2404, 2011.

BRISSONI, B.; AGOSTINI, L.; KROPF, M.; MARTINON, F.; SWOBODA, V.; LIPPENS, S.; EVERETT, H.; AEBI, N.; JANSSENS, S.; MEYLAN, E.; FELBERBAUM-CORTI, M.; HIRLING, H.; GRUENBERG, J.; TSCHOPP, J.; BURNS, K. Intracellular trafficking of interleukin-1 receptor I requires Tollip. **Current Biology**, v. 16, n. 22, p. 2265-2270, 2006.

BRUDNO, M.; MALDE, S.; POLIAKOV, A.; DO, C. B.; COURONNE, O.; DUBCHAK, I.; BATZOGLOU, S. Glocal alignment: finding rearrangements during alignment. **Bioinformatics**, v. 19, n. 1, p. i54-i62, 2003.

BURNS, K.; CLATWORTHY, J.; MARTIN, L.; MARTINON, F.; PLUMPTON, C.; MASCHERA, B.; LEWIS, A.; RAY, K.; TSCHOPP, J.; VOLPE, F. Tollip, a new component of the IL-1RI pathway, links IRAK to the IL-1 receptor. **Nature cell biology**, v. 2, n. 6, p. 346-351, 2000.

BURNS, K.; JANSSENS, S.; BRISSONI, B.; OLIVOS, N.; BEYAERT, R.; TSCHOPP, J. Inhibition of interleukin 1 receptor/Toll-like receptor signaling through the alternatively spliced, short form of MyD88 is due to its failure to recruit IRAK-4. **The Journal of experimental medicine**, v. 197, n. 2, p. 263-268, 2003.

DAYHOFF, M. O.; SCHWARTZ, R. M. A model of evolutionary change in proteins. In: **Atlas of protein sequence and structure**, p. 345–358, 1978.

DEGUINE, J.; BARTON, G. M. MyD88: a central player in innate immune signaling. **F1000prime reports**, v. 6; p. 1-7, 2014.

DICKERMAN, A. W. Among-run artifacts in DNA hybridization. **Systematic Biology**, v. 40, n. 4, p. 494-499, 1991.

DIDIERLAURENT, A.; BRISSONI, B.; VELIN, D.; AEBI, N.; TARDIVEL, A.; KÄSLIN, E.; SIRARD, J. C.; ANGELOV, G.; TSCHOPP, J; BURNS, K. Tollip regulates proinflammatory responses to interleukin-1 and lipopolysaccharide. **Molecular and Cellular biology**, v. 26, n. 3, p. 735-742, 2006.

ESCOLL, P.; ROLANDO, M.; GOMEZ-VALERO, L.; BUCHRIESER, C. From amoeba to macrophages: exploring the molecular mechanisms of *Legionella pneumophila* infection in both hosts. **Current Topics in Microbiology and Immunology**. v. 376, p. 1-34, 2013.

FELSENSTEIN, J. Estimation of hominoid phylogeny from a DNA hybridization data set. **Journal of molecular evolution**, v. 26, n. 1-2, p. 123-131, 1987.

FELSENSTEIN, J.; FELENSTEIN, J.. **Inferring phylogenies**. 1. ed. Seattle, Universty Washington: Sinauer Press, 2004. 580 p.

FITCH, W. M.; MARGOLIASH, E. Construction of phylogenetic trees. A method based on mutation distances as estimated from cytochrome c sequences is of general applicability. **Science**, v. 155, n. 3760, p. 279-284, 1967.

FITCH, W. M. Distinguishing homologous from analogous proteins. **Systematic Biology**, v. 19, n. 2, p. 99-113, 1970.

FITCH, W. M. Toward defining the course of evolution: minimum change for a specific tree topology. **Systematic zoology**, p. 406-416, 1971.

FITCH, W. M. Homology: a personal view on some of the problems. **Trends in genetics**, v. 16, n. 5, p. 227-231, 2000.

GROSSHANS, J.; BERGMANN, A.; HAFFTER, P.; NÜSSLEIN-VOLHARD, C. Activation of the kinase Pelle by Tube in the dorsoventral signal transduction pathway of Drosophila embryo. **Nature**, v. 372, n. 6506, p. 563-566, 1994.

HALL, B. K. **Homology: The hierarchial basis of comparative biology**. 1. ed. Academic Press, 2001. 504 p.

HENIKOFF, S.; HENIKOFF, J. G. Amino acid substitution matrices from protein blocks. **Proceedings of the National Academy of Sciences**, v. 89, n. 22, p. 10915-10919, 1992.

HULTMARK, D. Macrophage differentiation marker MyD88 is a member of the Toll/IL-1 receptor family. **Biochemical and biophysical research communications**, v. 199, n. 1, p. 144-146, 1994.

JANEWAY, C. A.; TRAVERS, P.; WALPORT, M. J.; SHLOMCHIK, M. J. **Immunobiology: the immune system in health and disease**. 5. ed. New York: Garland Science, 2001, 752p.

JANSSENS, S.; BURNS, K.; VERCAMMEN, E.; TSCHOPP, J.; BEYAERT, R. MyD88 S, a splice variant of MyD88, differentially modulates NF- $\kappa$ B-and AP-1-dependent gene expression. **FEBS letters**, 548, n. 1, p. 103-107, 2003.

KATOH, Y.; SHIBA, Y.; MITSUHASHI, H.; YANAGIDA, Y.; TAKATSU, H.; NAKAYAMA, K. Tollip and Tom1 form a complex and recruit ubiquitin-conjugated



proteins onto early endosomes. **Journal of Biological Chemistry**, v. 279, n. 23, p. 24435-24443, 2004.

KAWAI, T.; ADACHI, O.; OGAWA, T.; TAKEDA, K.; AKIRA, S. Unresponsiveness of MyD88-deficient mice to endotoxin. **Immunity**, v. 11, n. 1, p. 115-122, 1999.

KHAN, J. A.; BRINT, E. K.; O'NEILL, L. A.; TONG, L. Crystal structure of the Toll/interleukin-1 receptor domain of human IL-1RAPL. **Journal of Biological Chemistry**, v. 279, n. 30, p. 31664-31670, 2004.

LETSOU, A.; ALEXANDER, S.; ORTH, K.; WASSERMAN, S. A. Genetic and molecular characterization of tube, a Drosophila gene maternally required for embryonic dorsoventral polarity. **Proceedings of the National Academy of Sciences**, v. 88, n. 3, p. 810-814, 1991.

LI, X.; COMMANE, M.; BURNS, C.; VITHALANI, K.; CAO, Z.; STARK, G. R. Mutant cells that do not respond to interleukin-1 (IL-1) reveal a novel role for IL-1 receptor-associated kinase. **Molecular and cellular biology**, v. 19 n. 7, p. 4643-4652, 1999.

LI, X.; COMMANE, M.; JIANG, Z.; STARK, G. R. IL-1-induced NF $\kappa$ B and c-Jun N-terminal kinase (JNK) activation diverge at IL-1 receptor-associated kinase (IRAK). **Proceedings of the National Academy of Sciences**, v. 98 n. 8, p. 4461-4465, 2001.

LI, T., HU, J.; LI, L. Characterization of Tollip protein upon Lipopolysaccharide challenge. **Molecular immunology**, v. 41, n. 1, p. 85-92, 2004.

LO, Y. L. S.; BECKHOUSE, A. G.; BOULUS, S. L.; WELLS, C. A. Diversification of TOLLIP isoforms in mouse and man. **Mammalian Genome**, v. 20, n. 5, p. 305-314, 2009.

LORD, K. A.; HOFFMAN-LIEBERMANN, B.; LIEBERMANN, D. A. Nucleotide sequence and expression of a cDNA encoding MyD88, a novel myeloid differentiation primary response gene induced by IL6. **Oncogene**, v. 5, n. 7, p. 1095-1097, 1990.

MARGOLIASH, E. Primary structure and evolution of cytochrome c. **Proceedings of the National Academy of Sciences of the United States of America**, v. 50, n. 4, 672 p, 1963.

MARKS, J.; SCHMID, C. W.; SARICH, V. M. DNA hybridization as a guide to phylogeny: relations of the Hominoidea. **Journal of Human Evolution**, v. 17, n. 8, p. 769-786, 1988.

MARTIN, T. F. J. Phosphoinositide lipids as signaling molecules: common themes for signal transduction, cytoskeletal regulation, and membrane trafficking. **Annual review of cell and developmental biology**, v. 14, n. 1, p. 231-264, 1998.

MCLACHLAN, A. D. Repeating sequences and gene duplication in proteins. **Journal of molecular biology**, v. 64, n. 2, p. 417-437, 1972.

MEDZHITOV, R.; PRESTON-HURLBURT, P.; KOPP, E.; STADLEN, A.; CHEN, C.; GHOSH, S.; JANEWAY, C. A. MyD88 is an adaptor protein in the hToll/IL-1 receptor family signaling pathways. **Molecular cell**, v. 2, n. 2, p. 253-258, 1998.

MEDZHITOV, R. Toll-like receptors and innate immunity. **Nature Reviews Immunology**, v. 1, n. 2, p. 135-145, 2001.

MITCHAM, J. L.; PARNET, P.; BONNERT, T. P.; GARKA, K. E.; GERHART, M. J.; SLACK, J. L.; GAYLE, M. A.; DOWER, S. K.; SIMS, J. E. T1/ST2 signaling establishes it as a member of an expanding interleukin-1 receptor family. **Journal of Biological Chemistry**, v. 271, n. 10, p. 5777-5783, 1996.

MUZIO, M.; NI, J.; FENG, P.; DIXIT, V. M. IRAK (Pelle) family member IRAK-2 and MyD88 as proximal mediators of IL-1 signaling. **Science**, v. 278, n. 5343, p. 1612-1615, 1997.

NORRIS, J. L.; MANLEY, J. L. Functional interactions between the pelle kinase, Toll receptor, and tube suggest a mechanism for activation of dorsal. **Genes & development**, v. 10, n. 7, p. 862-872, 1996.

OHNISHI, H.; TOCHIO, H.; KATO, Z.; ORII, K. E.; LI, A.; KIMURA, T.; HIROAKI H, KONDO N, SHIRAKAWA M. Structural basis for the multiple interactions of the MyD88 TIR domain in TLR4 signaling. **Proceedings of the National Academy of Sciences**, v. 106, n. 25, p. 10260-10265, 2009.

O'NEILL, L. A. The interleukin-1 receptor/Toll-like receptor superfamily: 10 years of progress. **Immunological Reviews**. v. 226, p. 10–18, 2008.

SCANGA, C. A.; ALIBERTI, J.; JANKOVIC, D.; TILLOY, F.; BENNOUNA, S.; DENKERS, E. Y.; MEDZHITOV, R.; SHER, A. Cutting edge: MyD88 is required for resistance to *Toxoplasma gondii* infection and regulates parasite-induced IL-12 production by dendritic cells. **The Journal of Immunology**, v. 168, n. 12, p. 5997-6001, 2002.

SIBLEY, C. G.; COMSTOCK, J. A.; AHLQUIST, J. E. DNA hybridization evidence of hominoid phylogeny: a reanalysis of the data. **Journal of Molecular Evolution**, v. 30, n. 3, p. 202-236, 1990.

TAKEUCHI, O.; HOSHINO, K.; AKIRA, S. Cutting edge: TLR2-deficient and MyD88-deficient mice are highly susceptible to *Staphylococcus aureus* infection. **The Journal of Immunology**, v. 165, n. 10, p. 5392-5396, 2000.

TEMPLETON, A. R. The phylogeny of the hominoid primates: a statistical analysis of the DNA-DNA hybridization data. **Molecular biology and evolution**, v. 2, n. 5, p. 420-433, 1985.

TEMPLETON, A. R. Further comments on the statistical analysis of DNA-DNA hybridization data. **Molecular biology and evolution**, v. 3, n. 3, p. 290-298, 1986.

THORNE, J. L.; KISHINO, H.; FELSENSTEIN, J. An evolutionary model for maximum likelihood alignment of DNA sequences. **Journal of Molecular Evolution**, v. 33, n. 2, p. 114-124, 1991.

WESCHE, H.; HENZEL, W. J.; SHILLINGLAW, W.; LI, S.; CAO, Z. MyD88: an adapter that recruits IRAK to the IL-1 receptor complex. **Immunity**, v. 7, n. 6, p. 837-847, 1997.

YAMAKAMI, M.; YOSHIMORI, T.; YOKOSAWA, H. Tom1, a VHS domain-containing protein, interacts with tollip, ubiquitin, and clathrin. **Journal of Biological Chemistry**, n. 278, v. 52, p. 52865-52872, 2003.

ZHANG, G.; GHOSH, S. Negative regulation of toll-like receptor-mediated signaling by Tollip. **Journal of Biological Chemistry**, v. 277, n. 9, p. 7059-7065, 2002.

### **Tollip or Not Tollip: What Are the Evolving Questions behind It?**

#### **ABSTRACT**

Tollip plays an important role in the interleukin-1 receptor IL-1R and Toll pathways. As a modulator of the immune pathway, it indirectly controls the amount of antimicrobial peptides. This could indicate a vital step in maintaining animal immune systems and preventing infection. Evolutionary questions are crucial to understanding the conservation and functioning of the biochemical pathways like the Tollip-mediated one. Through an analysis of 36 sequences of the Tollip protein from different animal taxa, downloaded from Kyoto Encyclopedia of Genes and Genomes (KEGG) databank, we inferred diverse evolutionary parameters, such as molecular selection and structure conservation, by analyzing residue by residue, beyond the canonical parameters to this type of study, as maximum likelihood trees. We found that Tollip presented different trends in its evolving history. In primates, the protein is becoming more unstable, just the opposite is observed in the arthropod group. The most interesting finding was the concentration of positively selected residues at amino terminal ends. Some observed topological incongruences in maximum likelihood trees of complete and curated Tollip data sets could be explained through horizontal transfers, evidenced by recombination detection. These results suggest that there is more to be researched and understood about this protein.

Key words: Tollip, Bioinformatic, Evolution, Immune, TLR

## INTRODUCTION

There is a lot of biological information deposited in online databases, but little of the data is analyzed properly [1], [2]. These data are largely used in bioinformatics, covering various areas such as computer science, mathematics and biological engineering several. Thus it is possible to optimize these studies, in a simple way [3]. The bioinformatic data can be used in phylogenetic analysis, as it is used in most branches of biology, such as phylogenetic trees for paralog genes [4], population analysis [5], evolution, epidemiology [6], [7], and genomic and metagenomic sequence comparison [8]. Protein phylogeny is used to indicate synonymous and non-synonymous substitutions along with the branches in order to identify cases of rapid changes of amino acids [9]. The analysis of different trees allows for the observation of topological incongruences, differences in the formation of taxa, and the relationship between nodes and trees [10], [11]. The complete phylogenetic inference at species level is presented in the Tree of Life (ToL) Web Project. ToL is a collaborative project of hundreds of phylogenetic researchers correlating diverse sources of information, including morphological, physiological, and molecular information. (This project is a work in progress [12]).

The presence of pathogens in the environment can interfere in the survival and reproduction of individuals in a population, leading to new evolutionary trends [13], [14]. Multicellular organisms have a rapid immune response to pathogens entering, named innate immunity. This response is performed by specialized cells, which have specific receptors for pathogen-associated molecular patterns (PAMPs) [15], [16], the most noticeable are Toll-Like Receptors (TLRs) [17]. Tollip (Toll-interacting protein) participates in the signaling pathway of the TLR with an endogenous modulatory role. Tollip has a target N-terminal Myb1 (Tom1) binding domain (TBD), a conserved core domain 2 (C2) and a C-terminal portion of coupling ubiquitin to endoplasmic reticulum degradation (CUE). In resting cells, Tollip controls the activation pathway of Myeloid differentiation primary response gene (88) (MyD88)-dependent NF- $\kappa$ B in two different ways. First, Tollip associates with IL-1R, TLR4 after LPS activation, inhibiting the immune response mediated by TLR [18], [19]. This association requires TLR-TIR domain and intact C-terminal region of Tollip, CUE domain. Second, Tollip binds directly to interleukin-1

receptor-associated kinase-1 (IRAK-1) by inhibiting an autophosphorylation but without promoting its degradation. Overexpression of Tollip leads to inhibition of TLR2, TLR4, and IL-1R signaling, confirming a modulatory role of Tollip in immune responses [20]–[23].

The main goal of this paper is to show the topological incongruences between Tollip protein sequence phylogenetic trees using ToL data as reference. Other goals are to determine the diversity in the evolution of this protein in different taxa, the possible horizontal gene transfers, and the correlation of molecular features in the sequences within primates and arthropod groups.

## MATERIAL AND METHODS

Thirty-six sequences of Tollip protein were downloaded from KEGG (Table 1). The phylogenetic reference used was the Tree of Life Web Project, ToL (<http://tolweb.org/tree/>), which were used in topological comparisons with Tollip generated data.

All evolutionary analyses were carried out on the MEGA 5 software [24]. Maximum likelihood phylogenetic trees were obtained for Tollip using a Muscle alignment and G Blocks curation with default sets at PhyML 3.0 [25], using the most appropriate model of amino acid substitution and likelihood scores assessed by TOPALi V2.5 [26], [27]. The best model was determined by using the Akaike Information Criterion (AIC) [28], [29]. Supports for the nodes were evaluated by bootstrapping with 1000 pseudoreplicates.

The effect of reticulate evolutionary events was analyzed through a neighbor-net analysis [30] and converted into a splits graph using the drawing algorithms implemented in SplitsTree4 software – version 4.10 [31]. The neighbor-net method was based on the pairwise distance matrices of Tollip complete sequences alignment with deletion of gaps and non-informative parsimony sites; the matrices were calculated and corrected with the Poisson distribution model [32].

The isoelectric point (pI) and molecular weight (MW) of each protein sequence was inferred with the Compute pI/Mw tool [33], the variability present in sequences was accessed through the Protein Variability Server [34], and the main protein characteristics as instability index (ININ), aliphatic index (AI) and GRAVY (grand average of hydropathicity) were evaluated with the ProtParam tool [35]. Tests of correlation between collected data and statistical treatments were made with GraphPad Prism version 5.0 software [36].

To check if selection affected the patterns of genetic diversity, we tested if the protein was under positive selection. Tajima's D statistic [37] was calculated by testing the mutation neutrality hypothesis [38], as previously described by Coscollá and colleagues [39]. In order to investigate the presence of positively selected codons, the estimation of both positive and purifying selection at each amino acid sites was calculated from the ratio of non-synonymous to synonymous



substitutions,  $\omega$ , as previously described [40]. Analyses were conducted using the Selecton version 2.1 software [41], [42]. The significance of scores was obtained by using a Likelihood Ratio Test that compares two nested models: a null model that assumes no selection (M8a) [43] and an alternative model that does (M8) [44].

Several approaches were used to determine the extent of recombination in the Tollip data set. First, Tollip protein sequences previously aligned at Clustal W2 [45] were back-translated at BioEdit [46] package using standard genetic code, to normalize the codon frequencies and bring/make the comparisons more accurate. Once recombination eventually creates mosaic sequences in which evolutionary history at each site may be different. Then, GARD method [47] available in Datamonkey server [48] was also used to search for evidence of phylogenetic incongruences, and to identify the number and location of breakpoints corresponding to recombination events. In order to confirm GARD results, the recombination was assessed using a recombination cost “delta dirac” and mutation cost “Hamming”, implemented in the Recco program [49]. The gap extension cost was fixed to 0.2 and the statistical significance of the analysis was obtained after 1000 permutations. Validation of the previously obtained results was performed with the six methods implemented in the RDP3 program [50]: RDP [51], GENECONV [52], BootScan [53], Maximum Chisquared Test [54], CHIMAERA [55] and Sister Scan [56]. The analysis was performed with default settings for the detection methods, a Bonferroni corrected P-value cut-off of 0.05, and a requirement that each potential event had to be detected simultaneously by four or more methods.

A tridimensional model was generated starting from hsa protein sequence, evaluating I-Tasser server [57], using default sets. This approach was used in order to assess the potential implications of our findings in the tertiary protein structure. Tests to search for ligands and hot spots in the protein were ran using the Profunc application [58] at PDBSum [59].

## RESULTS AND DISCUSSION

The maximum likelihood composite trees generated are shown at Figure 1. The incongruent topology is evidenced by different branch sorting between phylogenetic trees of Tollip complete and G blocks trees, and between the current phylogenetic organization available at ToL. The most appropriate model for explaining the evolution of Tollip was found to be mtREV24 [60], [61] with following the parameters: BIC of 5424.48, AICc of 5006.19, lnL of -2432.61, Invariant sites n/a, and Gamma parameters of n/a.

The groups were split based on the Tollip protein sequence, confirming the evolutionary relatedness constructed in the ToL project. Despite this, a new phylogenetic array suggests other relationships between protein sequences of these animals. In Figure 1B, notice a node formed by subgroups 13 to 17, which include porifera (subgroup 17) and cnidaria (subgroup 16), together with bilateria, subgroups 13 to 15. The other branch is composed just by vertebrata, group I, which remains like a monophyletic group in all the trees, confirmed by bootstrap values in Figures 1B and 1C, respectively 81% and 89%. The separated groups, based in the Tollip sequence reinforces the evolutionary relatedness constructed in the ToL project. Although, other relationships between the vertebrates are suggested. Indeed, the group I suggests a characteristic function of the protein in order of higher levels of organisms complexity, requiring a less stable molecule for the modulation of information which will be discussed later. In the group I, the primates (subgroup 1) are in two arranges, in Figure 1B the ptr and ggo (bootstrap value of 100%) are far the other primates (bootstrap value of 94%) suggesting one difference in complete sequence, the non conserved sites difference ptr and ggo from the others. The conserved sites separated the two primates too, Figure 1C, the two sequences hold similarities between them in this molecular level.

In Figure 1C, there are two branches, one of them is formed by groups I to V and the other just by group III, except for subgroup 14. This configuration is due to the alignment of conserved sites in the sequences, showing the variability of Tollip sequences between different organisms. When the Tollip complete sequence was analyzed (Figure 1B), this configuration changed. The group III remains like a monophyletic group, the alignment of non conserved parts does not

change significantly the branch, subgroup 14 returns to the group III and the subgroup 15 becomes paraphyletic relatively to group III. This allows us to consider group III as close enough to be relatively consistent throughout the entire analysis.

The average length of Tollip was 262 amino acids with a standard deviation of 64 amino acids; and the molecular weight average was 31.33 kDa with a standard deviation of 6.78 kDa. Through splits decomposition and analysis of alignment, after block curing, we could estimate the proportion of invariant sites as 68.54% and the segregant sites counted was 77 in amino acid base. These findings suggest a tendency of recurrent duplications and/or insertions, as well as deletions evidenced by variations in the length and mass of protein ranging between approximately 25% and 4.62 fold, respectively. But it is important to stress that the protein activity seems be intact or just slightly reduced, once its activity is essential to keep the health in animals. Tollip participates in several immune pathways, mediated or not by Myd88 [18], [19], modulating the responses and the loss or reduction of its affinity to molecular complexes made between it and several other compounds, like IRAK-1 [20]–[23], could trigger an exaggerated response of the immune system, leading to death in some cases. Though, there are studies, like Didierlaurent's[62], which affirm that mice lacking Tollip become healthy and fertile.

Despite all identified polymorphisms and mutations of Tollip, we could not make any inferences about its role in the TLR-triggering activation of dendritic cells, without more in vitro and in vivo tests. Although, some studies [22], [62] have revealed that Tollip does not have a fundamental function in the TLR-triggering activation pathway of dendritic cells. Mutant mice lacking the tollip gene, when compared with wild-type mice, have been shown to not have significant differences. Therefore, mutations in key-residues for Tollip activity does not imply differences at the activation level of dendritic cells.

The characteristics of proteins were evaluated (Table 1) and the distribution of instability index (ININ), aliphatic index (AI) and GRAVY followed the normal distribution (P Kolmogorov Smirnov test >0.05). The molecular weight showed a statically significant correlation coefficient with aliphatic index ( $p = 0.014$ ;  $r = -0.4065$ ); another characters showed correlations between the aliphatic index and

the instability index ( $p < 0.001$ ;  $r = -0.5829$ ), and between GRAVY and the aliphatic index ( $p < 0.001$ ;  $r = 0.6714$ ). These data are shown in Table 2. The protein variability (Figure 2), was measured by the Shannon coefficient. We observed the Tollip G-block cured proteins, and the regions that have continuous conserved residues are probably responsible for the catalytic reactions or binding interactions.

The variability of sequence lengths implies a complex organization of Tollip function or adjustment in diverse immune pathways. The standard deviation of the number of residues in a protein probably reflects a process of tertiary structure "stabilization", evidenced by increasing AI values, which were positively correlated ( $r = 0.366$ ;  $p = 0.14$ ) with arthropod group. Higher molecular weights showed higher hydrophobicity by AI ( $p = -0.407$ ;  $r = 0.007$ ) and GRAVY results showed similar findings, being correlated with AI too ( $r = 0.671$ ;  $p = 3.7 \times 10^{-6}$ ). These aliphatic residues seem to be essential to the evolving process of this protein. The ININ revealed by itself a tendency of accumulation of instabilities in all groups except the arthropods, once a positive correlation of these values was shown between primates group ( $r = 0.515$ ,  $p = 0.001$ ) and another negative was shown between arthropods and ININ ( $r = -0.413$ ;  $p = 0.006$ ). These tendencies are related with a hydrophobization of the entire molecule, which increase the molecule lifetime, being advantageous for their group due to rapid molecular ratio and molecular turnover [63]–[65].

Primates revealed just a tendency to increase instability of protein (ININ;  $r = 0.515$ ,  $p = 0.001$ ); this is related to lower half-life in this protein. It is in agreement with the higher available energy in these animals, in opposite that observed in arthropods or small animals. The cell environment of superior chordata can be very unstable which enables physiological reactions, with rapid and efficient beginnings and ends. The Tollip has sites for ubiquitination [22], which considerably reduces its life-time. In these groups of animals, it seems that Tollip has more sites available or sites with more affinity to ubiquitin.

Starting from a virtual model of this protein, made from hsa sequence, at I-Tasser server using default variables had an estimated accuracy measured through TM-score of  $0.34 \pm 0.12$  and a RMSD of  $14.1 \pm 3.8$  Å. We tried to identify the pattern of hydrophobic pockets, but it was seen that aliphatic residual apolarity is

homogenously distributed along the entire sequence (Figure 3.A and B). The main residues, evidenced by conservation (Figure 2.A), likely construct the reaction pockets and in the order of the modular characteristics of this protein [66], the alpha-helices and beta-sheets are domains for binding to specific ligands. Some ions showed to be important to conservation of tertiary structure as calcium II (BS Score 1.34–1.39, RMSD 3.00, TM-Score 0.349). It contacts with G69-D74-D121-E122-R123 residues, as can be seen at Figure 3.C, that are relatively conserved. An organic compound, ligand 768 (1-(2,4-dichlorophenoxy)-3-{2-imino-3-[2-(1-piperidiny)ethyl]-2,3-dihydro-1H-benzimidazol-1-yl}-2-propanol), was found to be a ligand which contacts with E118-I131-A132-W133-L154 residues, as can be seen at Figure 3.D, with a RMSD of 5.51, TM-Score of 0.25 and BS Score of 0.86. This ligand 768 is related with inhibition of calcium-dependent membrane binding activity of prothrombin and of factors Va, VIII and Xa of human coagulation pathway [67] by interaction with C2 domain. This interaction is consistent with Tollip modular criteria and its functions, revealing a potential need of Ca<sup>2+</sup> to maintain the C2 domain structure and could be potentially inhibited by 768 ligand.

Splits graph (Figure 4) using a neighbor-net analysis, excluding parsimony uninformative sites, gaps and constant sites, showed a concentrated reticulation in the evolutionary history of Tollip, mainly disposed in superior animals. These groups present a complex diversification history. Indeed, these incongruences evidenced by trees (Figure 1) reveal an interesting clustering pattern in this protein, stressing the diversification of arthropod in detriment of others. This division is due to the diverse pathogenic elements that eventually could enter in contact with the arthropod and the ubiquity of their presence in almost all possible environments (earth, air and water) could make this process more efficient and fast. The high bootstrap values evidence the strong support of presented nodes and clusters.

GARD found at least 3 breaking-points with statistical significance ( $p < 0.001$ , KH test) and these findings were supported by Recco analysis with 1000 bootstraps and by at least four different algorithms at RDP software (Table 3). RDP showed the same breaking-points (Figure 5) which comprised the hypothesis that recombinational events generated or could isolate some groups bringing new specific pathways. Some incongruences, as discussed before, could be explained

by these events. Owing to a strange pattern of recombination found, the most probable hypothesis to support our findings is the horizontal transfer mediated by viruses or bacteria [68], that lived in the same environment of the two species, as some donors could not be identified with a high-threshold confidence level, these events could take part of very long and intrinsic evolutionary histories.

The molecular clock evaluated with the sequences (Table 4) showed that the sequences really presented different evolving patterns, where the sequences have increasing patterns of substitution when the complexity of the organisms become higher. These findings suggested that the Tollip evolutionary pattern is related with successive insertions and deletions that change the protein primary structure in order to bring less stable products; this is explained by the protein turnover that turns higher when the available energy and size of the animal increases [63]–[65].

Tajima's D statistic was 1.8226, meaning a tendency to positive selection. To assess the positive selection, we normalized the sequences using BioEdit through the back translation device, once the problem of codon preferences for each species could interfere in posterior analyses. To evaluate the results showed at Tajima's D statistic, the Selecton server was used and the results (Figure 6) showed a positive selection operating in almost all residues (49.33%) with statistical significance, M8 versus M8a as null model, evidenced by  $\Delta\text{LnL}$  value of 179.6 ( $p < 0.001$ ).

These findings are consistent with the data presented by analysis of the segregant and conserved sites, where it was determined that the protein is variable and presents a very active process of restructuring. The protein domains from amino-terminal ends are under a high positive selection, indicating that these parts of protein are variable and become higher adaptative values with more variability. Several consecutive amino acids presented in the second domain in the sequence a relative conservation, including a tendency to negative selection. These residues are related with the activity of the protein. Indeed, they could participate in the Tollip protein-protein and protein-lipid interactions, crucial to the right performance on the pathway.

The modulatory activity of Tollip is directly related to their association with different intracellular factors, such as other proteins and calcium ions. We have

noted that these residues responsible for such interactions suffer broadly neutral to negative selection, which in fact was obviously expected in order to keep their functionality.

Tollip polymorphisms were correlated with several human diseases like atopic dermatitis [69], inflammatory bowel disease [70], tuberculosis [71] and other. In atopic dermatitis (AD), Single-strand conformation polymorphism (SSCP) of the tollip sequence is correlated to AD. We could infer that amino acid exchanges of A (Ala) to S (Ser) occur at residue 222. Ser222 has a higher correlation to healthier controls (5.4%) when compared with Ala222 (2.7%) [69]. Residue 222 is occupied by A in 52.78% of sequences and is largely distributed in vertebrate sequences (subgroups 1–10), excluding oaa sequences. We found that it suffers a strong positive selection, through Selecton analysis (Fig. 6), shown as residue 245. In this case, seems that residue is conserved through vertebrata, subgroups 1 to 10, and we are inclined to believe that this is a trait which has become from an ancestor at amphibian level, and this mutation could be benefic to them too.

Tollip, in inflammatory bowel disease, suffers an inactivation that makes the intestinal epithelium unable to inhibit LPS-induced NF-kB activation [70], through a mutational amino acid exchange of lys150glu. In this sense, all primates present the residue K (Lys) at this position (except the ggo and ptr, which present R (Arg) and G (Gly), respectively), despite the common trend to present residue E (Glu) in other animals (55.56%). Residue D (Asp) in this position is important for insects (except for tsp and phu sequences, presenting E and Q (Gln) residues, respectively). This residue is under positive selection (position 166 at Selecton file, Figure 6). This mutation is apparently negative to health in primates and is a trait which was largely incorporated by other animals, reflecting a directional selection in this group, as occurs among insects.

In addition, a study involving tuberculosis (TB) and tollip [71] reveals that some synonymous polymorphisms or some that occur at noncoding regions (intronic regions or 3'UTR) could trigger different levels of risk of TB, not identifiable with protein analysis. This study also shows an association between minor homozygote of single-nucleotide polymorphism (SNP), named rs5743899, and a trend of reduced levels of Tollip mRNA in comparison with heterozygotes

and major homozygotes, driving down the Tollip expression levels regulating cytokine response. Still, another SNP (rs3750920) was associated with increased levels of Tollip mRNA, providing protection to the organism against TB. The hypofunctional Tollip genotype has an association with increased levels of proinflammatory cytokines and increased risk of TB, as well as production of augmented proinflammatory cytokines. However, the assessed SNPs were related to synonymous variations or mutations in non-coding regions, and therefore, our data could not reveal any kind of correlation with that. Shah et al. [71] finds that tollip's effect in murine models are not applicable because tollip behaves differently in humans. There is a need for more research in this area.



## CONCLUSIONS

Tollip presents diverse evolutionary tendencies and several of them are indicating successive modifications in the protein structure, in order to stabilize the tertiary structure accumulating aliphatic residues. Primates generally have more unstable proteins, while arthropods have more stability at ININ, AI and GRAVY level. Size was not correlated with any groups and seems to be highly variable in all groups. In/del trends were seen as very frequent. The three dimensional structure analysis revealed the modular characteristic of this protein and the necessity of  $\text{Ca}^{2+}$  to keep the correct pocket of C2 domain. Ligand association studies revealed that 768 ligand probably could inhibit the Tollip activity. Positively selected residues were found in almost all domains, but a considerable part of them are relatively conserved, indicating a conservation of active pockets, which is consistent with maintaining protein right activity. The tested animal groups were differentially grouped, when studied with parsimonious and non-parsimonious residues, and revealed through molecular clock analysis that they present different selection and evolving speeds. The recombination supports diverse incongruences observed in the phylogenetic trees obtained with complete and cured Tollip data sets. There are no evidences that support a homogeneity in this immunologic pathway, once Tollip presented evolving trends that are not constant for all groups. Summarizing, some groups are highly evolutionary closed, as arthropods and primates, but when compared between them are totally non consistent.

In conclusion, differences in Tollip structure among vertebrates could be detected as well as changes occurring in the primary structure through evolutionary processes. Once these changes occur in Tollip's structure, the same must occur with other structures in the IL-1R and TLR pathway. Adaptive immunity is commonly seen as the most evolved aspect of the immune systems of these organisms, but our data suggest that innate immunity in vertebrates could also be evolving differently among the species in order to promote a better adaptation to their reality.

## **ACKNOWLEDGMENTS**

We acknowledge to Joana Cardoso Costa of University of Coimbra, Portugal, for all patience and attention dispensed in the teaching of these valuable techniques. We thanks for LABGEN team.

Specie	Code	Tollip protein N. of access	pl	MW (Da)	Instability Index	Aliphatic Index	GRAVY
<i>Anolis carolinensis</i>	acs	100562725	4.99	31238.86	54.21	79.6	−0.295
<i>Anopheles gambiae</i>	aga	AGAP003615	5.37	30424.6	42.29	85.57	−0.319
<i>Apis mellifera</i>	ame	552034	5.63	31855.55	55.02	79.36	−0.334
<i>Ailuropoda melanoleuca</i>	aml	100463737	4.75	35361.18	47.73	83.28	−0.212
<i>Acyrtosiphon pisum</i>	api	100167693	5.31	30628.78	37.88	83.79	−0.453
<i>Amphimedon queenslandica</i>	aqu	100616078	5.55	33666.77	47.75	66.31	−0.706
<i>Bos taurus</i>	bta	539480	5.3	30107.58	43.7	84.25	−0.191
<i>Canis familiaris</i>	cfa	483657	5.05	30068.36	44.38	81.79	−0.211
<i>Ciona intestinalis</i>	cin	100178902	4.96	33008.67	41.76	77.63	−0.438
<i>Danio rerio</i>	dre	336876	4.93	30435.85	45.65	80.54	−0.232
<i>Equus caballus</i>	ecb	100051383	6.02	45675.89	51.53	77.7	−0.408
<i>Felis catus</i>	fca	101095614	5.18	29976.31	42.47	82.52	−0.186
<i>Gallus gallus</i>	gga	423099	5.03	30539.94	48.15	78.61	−0.317
<i>Gorilla gorilla gorilla</i>	ggo	101153660	6.51	32640.55	51.84	82.73	−0.134
<i>Hydra magnipapillata</i>	hmg	100204873	4.91	24690.68	42.49	85.37	−0.226
<i>Homo sapiens</i>	hsa	54472	5.68	30281.81	54.72	82.19	−0.243
<i>Ixodes scapularis</i>	isc	ISCW002962	6.79	21534.59	24.65	92.49	−0.073
<i>Macaca mulatta</i>	mcc	702587	5.68	30311.83	54.72	81.82	−0.252
<i>Monodelphis domestica</i>	mdo	100011810	5.04	30174.46	48	84.31	−0.232
<i>Meleagris gallopavo</i>	mga	100543926	4.89	31330.93	46.17	82.14	−0.279
<i>Mus musculus</i>	mmu	54473	5.04	30344.82	52.35	76.86	−0.29
<i>Nasonia vitripennis</i>	nvi	100115445	5.62	31502.04	45.79	81.29	−0.4
<i>Ornithorynchus anatinus</i>	oaa	100078225	8.9	23160.41	47.24	72.92	−0.498
<i>Oryzias latipes</i>	ola	101163379	5.05	30281.61	46.64	81.92	−0.192
<i>Pediculus humanus corporis</i>	phu	PHUM008100	5.8	29271.21	42.43	82.81	−0.362
<i>Pongo abelii</i>	pon	100442943	5.68	30281.81	54.72	82.19	−0.243
<i>Pan paniscus</i>	pps	100994067	5.08	22812.3	62.26	79.41	−0.229
<i>Pan troglodytes</i>	ptr	450943	5.79	63790.54	53.35	71.86	−0.263
<i>Rattus norvegicus</i>	rno	361677	5.04	30314.79	52.35	77.23	−0.281
<i>Strongylocentrotus purpuratus</i>	spu	756664	5.61	33741.14	60.18	68.27	−0.494
<i>Tribolium castaneum</i>	tca	664056	5.9	29687.86	42.98	85.75	−0.339
<i>Taeniopygia guttata</i>	tgu	100220909	5.03	30394.82	46.29	79.34	−0.268
<i>Takifugu rubripes</i>	tru	101078355	4.93	30363.69	45.7	80.8	−0.218
<i>Trichinella spiralis</i>	tsp	3457	4.78	27893.53	36.6	92.42	−0.247
<i>Xenopus laevis</i>	xla	443846	4.93	29983.24	47.91	77.21	−0.313
<i>Xenopus tropicalis</i>	xtr	448770	5.26	30069.42	43.11	76.47	−0.322

These data are representative from several taxa and would permit to infer about the evolutionary history of Tollip in diverse animal groups. Main protein features were inferred through diverse web servers from amino acid sequences.  
doi:10.1371/journal.pone.0097219.t001

**Table 1.** Tollip downloaded reference data from KEGG and principal protein features.

doi:10.1371/journal.pone.0097219.t001

	MW	Instability Index	Aliphatic Index	GRAVY	pI	Arthropod	Primates
MW		<b>0.04</b>	<b>0.01</b>	0.21	0.36	0.19	0.07
Instability Index	<b>0.28</b>		<b>9.55E-05</b>	0.18	0.42	<b>0.01</b>	<b>0.001</b>
Aliphatic Index	<b>-0.41</b>	<b>-0.58</b>		<b>3.68E-06</b>	0.19	<b>0.01</b>	0.40
GRAVY	-0.14	-0.16	<b>0.67</b>		0.10	0.24	0.06
pI	-0.06	-0.04	-0.15	-0.22		0.08	0.16
Arthropod	-0.15	<b>-0.41</b>	<b>0.36</b>	-0.12	0.24		0.10
Primates	0.25	<b>0.51</b>	-0.04	0.27	0.17	-0.22	

These characters were evaluated under ANOVA 1 Factor test, followed by Bartlett's test for equal variances at 5% of significance. To evaluate the correlation coefficients and p-value was made using the Bonferroni's Multiple Comparison test at 5% of significance. R scores of correlation are shown under table's diagonal and p-value of each one are shown above table's diagonal. Significant values are in bold.  
doi:10.1371/journal.pone.0097219.t002

**Table 2.** Correlation between main assessed protein characteristics of Tollip.

doi:10.1371/journal.pone.0097219.t002

Recombinant Event (RE)	Recombinant	Major Parent	Minor Parent	Breaking Point	P - value	Recombination Detection Tests								
						RPD	GENCONV	BootScan	MaxChi	Chimaera	SiScan	PhylPro	LARD	3Seq
1	mgp	gga_423099 (100% similarity)	Unknown (cin_100178902)	839–31	$4.544 \times 10^{-9}$ (GENECONV)	1	2	-	2	1	-	-	-	-
2	acs	gga_423099 (92.9% similarity)	Unknown (cin_100178902)	5–30	$2.990 \times 10^{-6}$ (RDP)	1	2	-	2	1	-	-	-	-
3	ptr	mcc_702587 (99.6% similarity)	Unknown (bta_539480)	1–78	$1.544 \times 10^{-7}$ (SiScan)	-	-	1	1	-	1	-	-	1
4	ptr	aml_100463737 (90.2% similarity)	Unknown (oaa_100078225)	276–496	$6.285 \times 10^{-28}$ (RDP)	1	1	1	1	1	1	-	-	1
5	ecb	fca_101095614 (95.2% similarity)	Unknown (spu_756664)	817–45	$1.627 \times 10^{-11}$ (RDP)	1	1	-	1	1	-	-	-	1
6	oaa	fca_101095614 (97% similarity)	Unknown (api_100167693)	194–528	$3.241 \times 10^{-21}$ (GENECONV)	1	1	1	1	1	1	-	-	1

doi:10.1371/journal.pone.0097219.t003

doi:10.1371/journal.pone.0097219.t003

**Table 3.** Potential recombinant events identified in Tollip with RDP.

doi:10.1371/journal.pone.0097219.t003

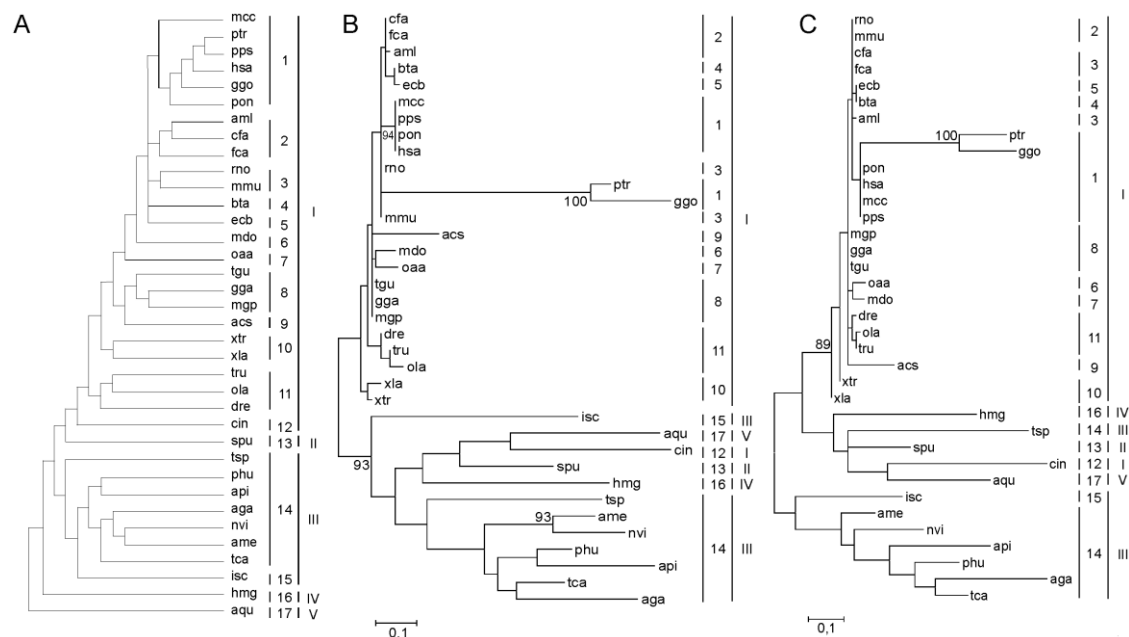
Complete Tollip Sequences				
	<i>lnL</i>	Parameters	(+G)	(+I)
With Clock	2648.884	35	n/a	n/a
Without Clock	2504.758	69	n/a	n/a
Curated G-Block Tollip Sequences				
	<i>lnL</i>	Parameters	(+G)	(+I)
With Clock	2375.702	35	n/a	n/a
Without Clock	2204.713	69	n/a	n/a

The molecular clock test was performed by comparing the ML value for the given topology with and without the molecular clock constraints under Poisson model [32]. The null hypothesis of equal evolutionary rate throughout the trees were rejected at a 5% significance level ( $P < 0.05$ ). The analysis involved 36 amino acid sequences. All positions containing gaps and missing data were eliminated. There were a total of 88 positions in the final dataset. Evolutionary analyses were conducted in MEGA5 [24].

doi:10.1371/journal.pone.0097219.t004

**Table 4.** Results from a test of molecular clocks using the Maximum Likelihood method.

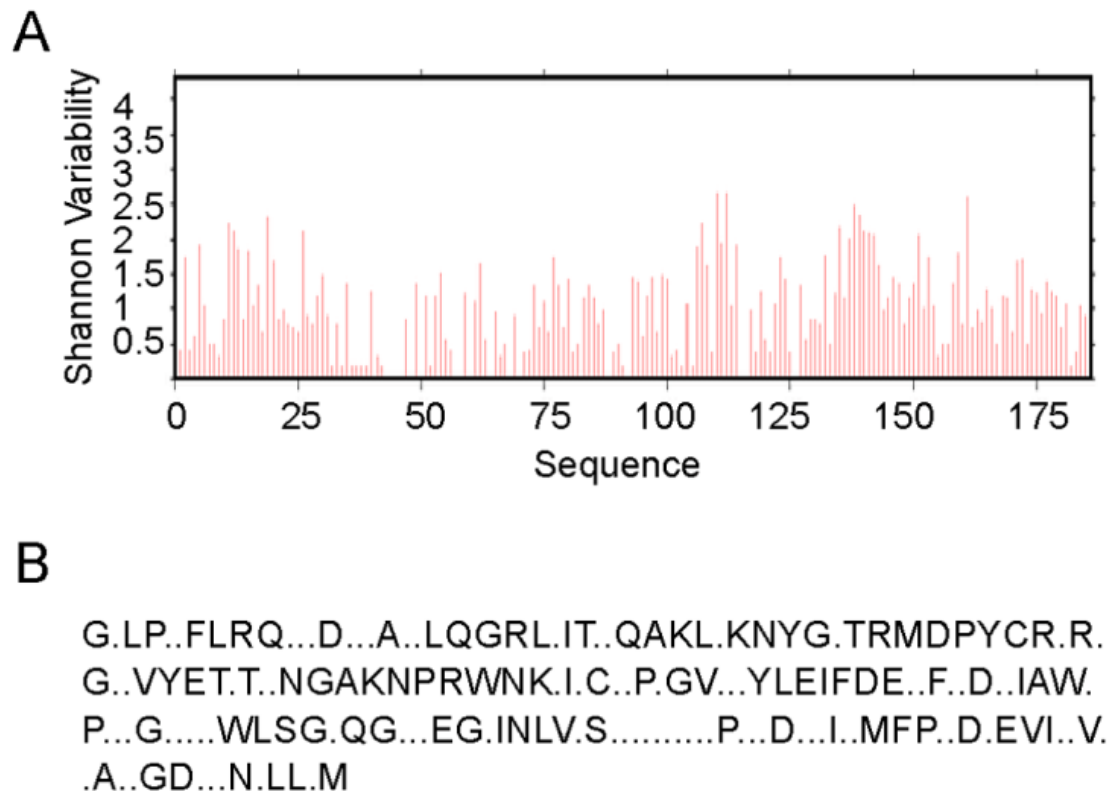
doi:10.1371/journal.pone.0097219.t004



**Figure 1.** Molecular Phylogenetic analysis by Maximum Likelihood method.

The evolutionary history of ToL (A), complete Tollip protein (B) and G-block cured Tollip protein (C) are shown. Trees B and C are drawn to scale, with branch lengths measured in the number of substitutions per site. The analysis involved 36 amino acid sequences. All positions containing gaps and missing data were eliminated. Group sorting was made in roman numerals (I - vertebrata, II - echinodermata, III - arthropod, IV - cnidaria, V - porifera) and the subgroups were coded in arabic numerals (1 - primates, 2 - carnivora, 3 - rodentia, 4 - bovidae, 5 - equidae 6 - marsupialia 7 - monotremata, 8 - birds, 9 - reptilia, 10 - amphibia, 11 - actinopterygii, 12 - ascidiacea, 13 - echinoidea, 14 - hexapoda, 15 - arachnida, 16 - hydrozoa, 17 - demospongiae).

doi:10.1371/journal.pone.0097219.g001

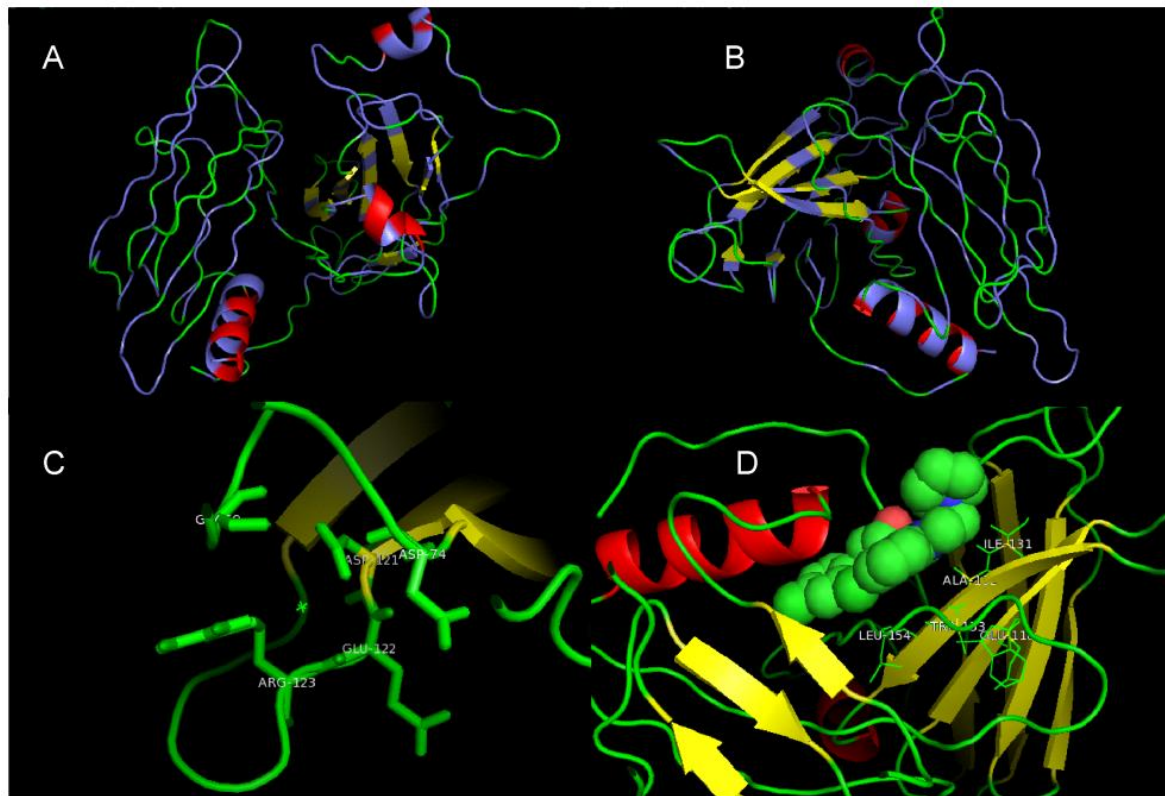


**Figure 2.** Variability of Tollip protein.

The variability residue per residue measured with Shannon Index (A). Protein conserved residues are disposed at (B), with variable positions as ".", these positions reflect some essential molecular arrangement to Tollip function. All analyses were made with Tollip G-block cured, to avoid the gaps and non-informative parsimony site.

doi:10.1371/journal.pone.0097219.g002

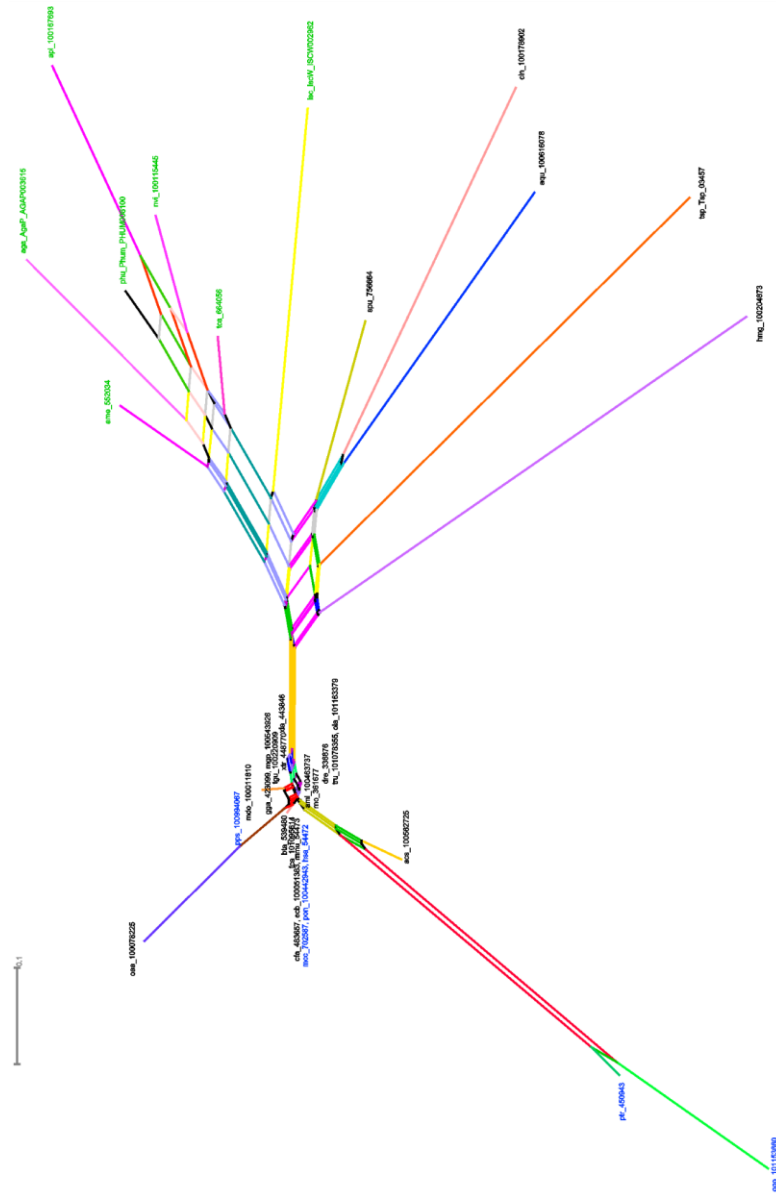




**Figure 3.** Human (hsa) Tollip tertiary structure.

This structure was modeled at I-Tasser server, using default sets. It could be seen the aliphatic residues distribution along all sequence (A and B). The ligands are shown arranged at lateral chain of the right residues, Ca<sup>2+</sup> (C) and 768 (D).

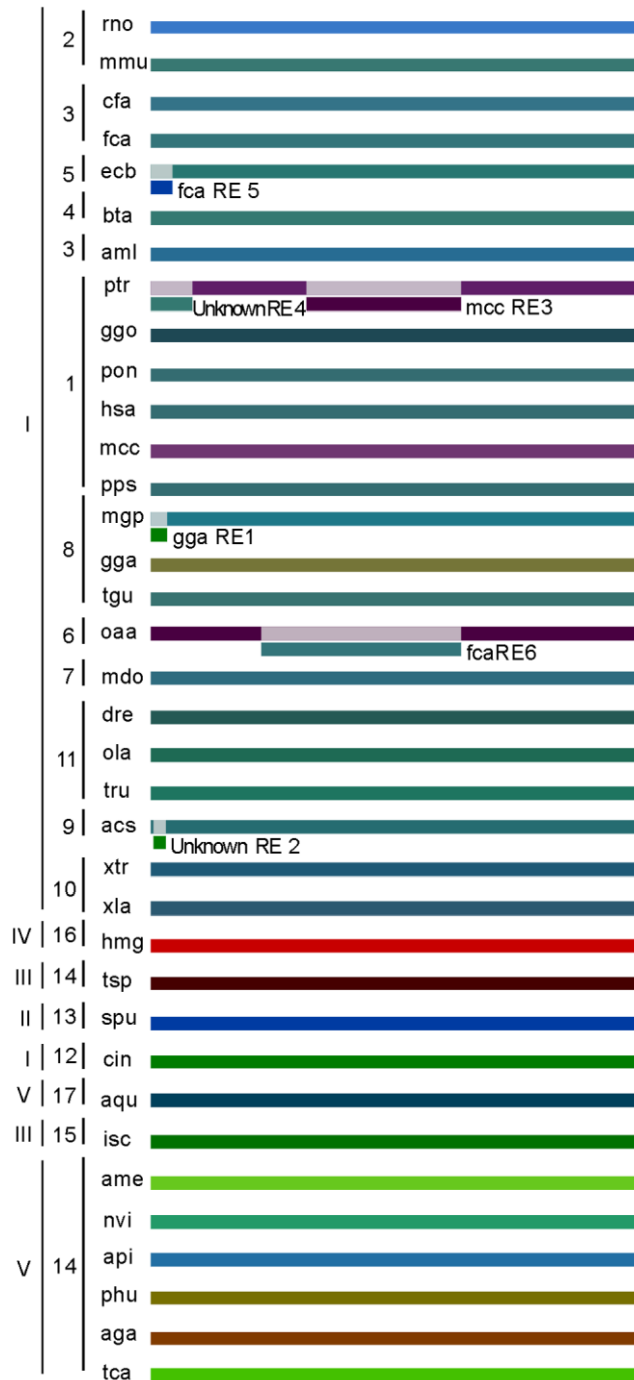
doi:10.1371/journal.pone.0097219.g003



**Figure 4.** Splits graph of complete Tolip protein alignment.

The parsimony uninformative sites, gaps and constant sites were excluded. There were 1000 pseudoreplicates performed as bootstraps to support the derivations, it was used as ProteinML distance correction the model mtREV24. Green operational taxonomic units represents arthropods groups, blue taxa represent primates and black represent the other groups.

doi:10.1371/journal.pone.0097219.g004



**Figure 5.** Recombinational events involved with Tollip evolution.

Each sequence are represented by a color and the recombination is evidenced by donor. All analyses were evaluated with RDP and the most significant P value to support the findings are shown at Table 3.

doi:10.1371/journal.pone.0097219.g005



**Figure 6.** Positive selections operating in each codon of Tollip, evidenced by Selecton algorithm. There were used two models which were evaluated separately, M8 and M8a, where the last is referenced as null model.  
 doi:10.1371/journal.pone.0097219.g006

## REFERENCES

1. Galperin MY, Fernández-Suárez XM (2012) The 2012 Nucleic Acids Research Database Issue and the online Molecular Biology Database Collection. *Nucleic Acids Res* 40: D1–D8.
2. Wren JD, Bateman A (2008) Databases, data tombs and dust in the wind. *Bioinformatics* 24: 2127–2128.
3. Mount D (2002) *Bioinformatics: Sequence and Genome Analysis*. Cold Spring Harbor Laboratory Press. pp. 480–526.
4. Mäser P, Thomine S, Schroeder JI, Ward JM, Hirschi K, et al. (2001) Phylogenetic relationships within cation transporter families of Arabidopsis. *Plant Physiol* 126: 1646–1667.
5. Edwards SV (2009) Is a new and general theory of molecular systematics emerging? *Evolution* 63: 1–19.
6. Marra MA, Jones SJM, Astell CR, Holt RA, Brooks-Wilson A, et al. (2003) The Genome sequence of the SARS-associated coronavirus. *Science* 300: 1399–1404.
7. Grenfell BT, Pybus OG, Gog JR, Wood JLN, Daly JM, et al. (2004) Unifying the epidemiological and evolutionary dynamics of pathogens. *Science* 303: 327–332.
8. Brady A, Salzberg S (2011) PhymmBL expanded: confidence scores, custom databases, parallelization and more. *Nat Methods* 8: 367.
9. Yang Z (2007) Adaptive molecular evolution. In: Balding D, Bishop M, Cannings C, editors. *Handbook of statistical genetics*. Wiley, New York. pp. 381–386.

10. Mason-Gamer RJ, Kellogg EA (1996) Testing for Phylogenetic Conflict Among Molecular Data Sets in the Tribe Triticeae (Gramineae). *Syst Biol* 45: 524–545.
11. Rokas A, Williams BL, King N, Carroll SB (2003) Genome-scale approaches to resolving incongruence in molecular phylogenies. *Nature* 425: 798–804.
12. Maddison D, Schulz K, Wayne P, Maddison W (2007) The Tree of Life Web Project. *Zootaxa* 1668: 19–40.
13. Price P (1980) *Evolutionary biology of parasites*. Princeton, NJ: Princeton University Press. 237 p.
14. Grenfell B, Dobson A (1995) *Ecology of infectious diseases in natural populations*. Cambridge, UK: Cambridge University Press. 536 p.
15. Franceschi C, Bonafè M, Valensin S (2000) Human immunosenescence: the prevailing of innate immunity, the failing of clonotypic immunity, and the filling of immunological space. *Vaccine* 18: : 1717–1720. Available: <http://linkinghub.elsevier.com/retrieve/pii/S0264410X99005137>. Accessed 2013 Dec 13.
16. Müller L, Fülöp T, Pawelec G (2013) Immunosenescence in vertebrates and invertebrates. *Immun Ageing* 10: : 12. Available: <http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3637519&tool=pmcentrez&rendertype=abstract>. Accessed 2013 Dec 13.
17. O'Neill LAJ, Golenbock D, Bowie AG (2013) The history of Toll-like receptors - redefining innate immunity. *Nat Rev Immunol* 13: : 453–460. Available: <http://www.ncbi.nlm.nih.gov/pubmed/23681101>. Accessed 2013 Dec 13.
18. Bulut Y, Faure E, Thomas L, Equils O, Arditi M (2001) Cooperation of Toll-like receptor 2 and 6 for cellular activation by soluble tuberculosis factor and *Borrelia*

burgdorferi outer surface protein A lipoprotein: role of Toll-interacting protein and IL-1 receptor signaling molecules in Toll-like receptor 2 signaling. *J Immunol* 167: 987–994.

19. Zhang G, Ghosh S (2002) Negative regulation of toll-like receptor-mediated signaling by Tollip. *J Biol Chem* 277: : 7059–7065. Available: <http://www.ncbi.nlm.nih.gov/pubmed/11751856>. Accessed 2013 Dec 13.

20. Burns K, Clatworthy J, Martin L, Martinon F, Plumpton C, et al. (2000) Tollip, a new component of the IL-1RI pathway, links IRAK to the IL-1 receptor. *Nat Cell Biol* 2: 346–351.

21. Li T, Hu J, Li L (2004) Characterization of Tollip protein upon Lipopolysaccharide challenge. *Mol Immunol* 41: 85–92. Available: <http://www.ncbi.nlm.nih.gov/pubmed/15140579>. Accessed 2013 Dec 13.

22. Brissoni B, Agostini L, Kropf M, Martinon F, Swoboda V, et al. (2006) Intracellular trafficking of interleukin-1 receptor I requires Tollip. *Curr Biol* 16: 2265–2270.

23. Piao W, Song C, Chen H, Diaz MAQ, Wahl LM, et al. (2009) Endotoxin tolerance dysregulates MyD88- and Toll/IL-1R domain-containing adapter inducing IFN-beta-dependent pathways and increases expression of negative regulators of TLR signaling. *J Leukoc Biol* 86: 863–875.

24. Tamura K, Peterson D, Peterson N, Stecher G, Nei M, et al. (2011) MEGA5: molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. *Mol Biol Evol* 28: 2731–2739.

25. Guindon S, Gascuel O (2003) A simple, fast, and accurate algorithm to estimate large phylogenies by maximum likelihood. *Syst Biol* 52: 696–704.

26. Milne I, Wright F, Rowe G, Marshall DF, Husmeier D, et al. (2004) TOPALi: software for automatic identification of recombinant sequences within DNA multiple alignments. *Bioinformatics* 20: 1806–1807.
27. Milne I, Lindner D, Bayer M, Husmeier D, McGuire G, et al. (2009) TOPALi v2: a rich graphical interface for evolutionary analyses of multiple alignments on HPC clusters and multi-core desktops. *Bioinformatics* 25: 126–127.
28. Akaike H (1974) A new look at the statistical model identification. *Autom Control IEEE Trans* 19: 716–723.
29. Posada D, Buckley TR (2004) Model selection and model averaging in phylogenetics: advantages of akaike information criterion and bayesian approaches over likelihood ratio tests. *Syst Biol* 53: 793–808.
30. Bryant D, Moulton V (2004) Neighbor-net: an agglomerative method for the construction of phylogenetic networks. *Mol Biol Evol* 21: 255–265.
31. Huson DH, Bryant D (2006) Application of phylogenetic networks in evolutionary studies. *Mol Biol Evol* 23: 254–267.
32. Zuckerkandl E, Pauling L (1965) Evolutionary divergence and convergence in proteins. In: Bryson V, Vogel H, editors. *Evolving genes and proteins*. New York, USA: Academic Press. pp. 97–166.
33. Bjellqvist B, Basse B, Olsen E, Celis JE (1994) Reference points for comparisons of two-dimensional maps of proteins from different human cell types defined in a pH scale where isoelectric points correlate with polypeptide compositions. *Electrophoresis* 15: 529–539.
34. Garcia-Boronat M, Diez-Rivero CM, Reinherz EL, Reche PA (2008) PVS: a web server for protein sequence variability analysis tuned to facilitate conserved epitope discovery. *Nucleic Acids Res* 36: W35–41.



35. Gasteiger E, Hoogland C, Gattiker A, Duvaud S, Wilkins MR, et al.. (2005) Protein Identification and Analysis Tools on the ExPASy Server. In: Walker JM, editor. In: The Proteomics Protocols Handbook, Humana Press.
36. GraphPad Prism version 5.04 for Windows, GraphPad Software, La Jolla California USA. Available: [www.graphpad.com](http://www.graphpad.com) Accessed 2013 Nov 7.
37. Tajima F (1989) Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. *Genetics* 123: 585–595.
38. Kimura M (1983) *The Neutral Theory of Molecular Evolution*. Cambridge, UK: Cambridge University Press. 367 p.
39. Coscollá M, Gosalbes MJ, Catalán V, González-Candelas F (2006) Genetic variability in environmental isolates of *Legionella pneumophila* from Comunidad Valenciana (Spain). *Environ Microbiol* 8: 1056–1063.
40. Costa J, Tiago I, Da Costa MS, Veríssimo A (2010) Molecular evolution of *Legionella pneumophila* dotA gene, the contribution of natural environmental strains. *Environ Microbiol* 12: 2711–2729.
41. Stern A, Doron-Faigenboim A, Erez E, Martz E, Bacharach E, et al. (2007) Selecton 2007: advanced models for detecting positive and purifying selection using a Bayesian inference approach. *Nucleic Acids Res* 35: W506–11.
42. Doron-Faigenboim A, Stern A, Mayrose I, Bacharach E, Pupko T (2005) Selecton: a server for detecting evolutionary forces at a single amino-acid site. *Bioinformatics* 21: 2101–2103.
43. Swanson WJ, Nielsen R, Yang Q (2003) Pervasive adaptive evolution in mammalian fertilization proteins. *Mol Biol Evol* 20: 18–20.

44. Yang Z, Nielsen R, Goldman N, Pedersen AM (2000) Codon-substitution models for heterogeneous selection pressure at amino acid sites. *Genetics* 155: 431–449.
45. Larkin MA, Blackshields G, Brown NP, Chenna R, McGettigan PA, et al. (2007) Clustal W and Clustal X version 2.0. *Bioinformatics* 23: 2947–2948.
46. Hall TA (1999) BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. *Nucleic Acids Symp Ser* 41: 95–98.
47. Kosakovsky Pond SL, Posada D, Gravenor MB, Woelk CH, Frost SDW (2006) GARD: a genetic algorithm for recombination detection. *Bioinformatics* 22: 3096–3098.
48. Delport W, Poon AFY, Frost SDW, Kosakovsky Pond SL (2010) Datamonkey 2010: a suite of phylogenetic analysis tools for evolutionary biology. *Bioinformatics* 26: 2455–2457.
49. Maydt J, Lengauer T (2006) Recco: recombination analysis using cost optimization. *Bioinformatics* 22: 1064–1071.
50. Martin DP, Lemey P, Lott M, Moulton V, Posada D, et al. (2010) RDP3: a flexible and fast computer program for analyzing recombination. *Bioinformatics* 26: 2462–2463.
51. Martin D, Rybicki E (2000) RDP: detection of recombination amongst aligned sequences. *Bioinformatics* 16: 562–563.
52. Padidam M, Sawyer S, Fauquet CM (1999) Possible emergence of new geminiviruses by frequent recombination. *Virology* 265: 218–225.

53. Martin DP, Posada D, Crandall KA, Williamson C (2005) A modified bootscan algorithm for automated identification of recombinant sequences and recombination breakpoints. *AIDS Res Hum Retroviruses* 21: 98–102.
54. Smith JM (1992) Analyzing the mosaic structure of genes. *J Mol Evol* 34: 126–129.
55. Posada D, Crandall KA (2001) Evaluation of methods for detecting recombination from DNA sequences: computer simulations. *Proc Natl Acad Sci U S A* 98: 13757–13762.
56. Gibbs MJ, Armstrong JS, Gibbs AJ (2000) Sister-scanning: a Monte Carlo procedure for assessing signals in recombinant sequences. *Bioinformatics* 16: 573–582.
57. Zhang Y (2008) I-TASSER server for protein 3D structure prediction. *BMC Bioinformatics* 9: 40.
58. Laskowski RA, Watson JD, Thornton JM (2005) ProFunc: a server for predicting protein function from 3D structure. *Nucleic Acids Res* 33: W89–W93.
59. Laskowski RA, Chistyakov VV, Thornton JM (2005) PDBsum more: new summaries and analyses of the known 3D structures of proteins and nucleic acids. *Nucleic Acids Res* 33: D266–D268.
60. Adachi J, Hasegawa M (1996) Model of amino acid substitution in proteins encoded by mitochondrial DNA. *J Mol Evol* 42: 459–468.
61. Yang Z, Nielsen R, Hasegawa M (1998) Models of amino acid substitution and applications to mitochondrial protein evolution. *Mol Biol Evol* 15: 1600–1611.

62. Didierlaurent A, Brissoni B, Velin D, Aebi N, Tardivel A, et al. (2006) Tollip regulates proinflammatory responses to interleukin-1 and lipopolysaccharide. *Mol Cell Biol* 26: 735–742.
63. Yeh GY, Eisenberg DM, Kaptchuk TJ, Phillips RS (2003) Systematic review of herbs and dietary supplements for glycemic control in diabetes. *Diabetes Care* 26: 1277–1294.
64. Speakman JR (2005) Body size, energy metabolism and lifespan. *J Exp Biol* 208: 1717–1730.
65. White CR, Seymour RS (2005) Allometric scaling of mammalian metabolism. *J Exp Biol* 208: 1611–1619.
66. Azurmendi HF, Mitra S, Ayala I, Li L, Finkielstein CV, et al. (2010) Backbone (1)H, (15)N, and (13)C resonance assignments and secondary structure of the Tollip CUE domain. *Mol Cells* 30: 581–585.
67. Segers K, Sperandio O, Sack M, Fischer R, Miteva MA, et al. (2007) Design of protein membrane interaction inhibitors by virtual ligand screening, proof of concept with the C2 domain of factor V. *Proc Natl Acad Sci U S A*. 104: 12697–12702.
68. Keeling PJ, Palmer JD (2008) Horizontal gene transfer in eukaryotic evolution. *Nat Rev Genet* 9: 605–618.
69. Schimming TT, Parwez Q, Petrasch-Parwez E, Nothnagel M, Epplen JT, et al. (2007) Association of toll-interacting protein gene polymorphisms with atopic dermatitis. *BMC Dermatol* 7: 1–8.
70. Ishihara S, Aziz MM, Yuki T, Kazumori H, Kinoshita Y (2009) Inflammatory bowel disease: review from the aspect of genetics. *J Gastroenterol* 44: 1097–1108.

71. Shah JA, Vary JC, Chau TTH, Bang ND, Yen NTB, et al. (2012) Human TOLLIP regulates TLR2 and TLR4 signaling and its polymorphisms are associated with susceptibility to tuberculosis. *J Immunol* 189: 1737–1746.

## **CAPÍTULO 3**

---

### **Aspectos evolutivos da proteína MyD88 e seus domínios**

#### **RESUMO**

A proteína MyD88 de vertebrados participa da via do NF- $\kappa$ B sinalizada pelos receptores da família TLR e IL-R, atuando na resposta imunológica inata. Essas proteínas possuem dois domínios, DM que interage com proteínas citoplasmáticas da família IRAK, TIR que interage com os receptores de membrana e uma região INT que se liga entre os dois domínios. Este trabalho teve como objetivo identificar aspectos evolutivos das proteínas MyD88 presentes em 46 espécies animais. Os resultados indicam que ambos os domínios e sobretudo, o domínio TIR, apresentam um alto grau de conservação, incluindo os resíduos descritos como essenciais na interface de interação da proteína. Entre as espécies, as proteínas MyD88 de primatas demonstram ter maior identidade entre si.

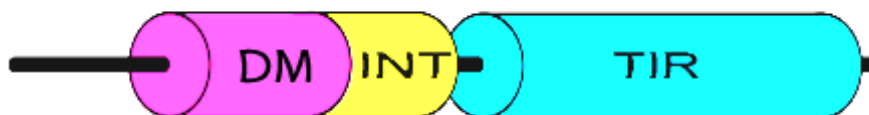
Palavras-chave: MyD88, domínio, DM, TIR, Identidade

## INTRODUÇÃO

Inicialmente o gene MyD88 (gene 88 da resposta primária de Diferenciação Mieloide) foi identificado em células mieloides murinas como importante na via de maturação celular (Lord et al., 1990a; 1990b).

Através de análises bioquímicas da MyD88 foi possível identificar relações dos domínios TIR (Receptor Toll/IL-1) que está presente tanto nela quanto nos receptores IL-R (Receptor de Interleucina) e Toll. Posteriormente foi apontado uma relação dos domínios DM (Domínio de Morte) entre a MyD88 e o mesmo domínio em IRAK (IL-1 Receptor Associado a Quinase) de mamíferos e Pelle de *Drosophila*, o que sugere que a proteína recruta essas moléculas em suas vias de sinalização por meio dessas interações, TIR/TIR e DM/DM (Hultmark, 1994; Muzio et al., Wesche et al., 1997).

Entre esses domínios há uma região intermediária (INT) que liga esses dois domínios como mostrado na **figura 1**. (Medzhitov et al., 1997; Muzio et al., 1997; Wesche et al., 1997).



**Figura 1.** Disposição das porções DM, INT e TIR na molécula MyD88.

Realizando estudos de deleção do gene MyD88 em ratos, confirmou-se a sua importância na ativação da via NF- $\kappa$ B (Fator de Transcrição Nuclear kappa B) e na produção de citocinas (Adachi et al., 1998; Bazan et al., 1996; Robinson et al., 1997). Estudos de super expressão mostraram ainda que MyD88 é importante para a sinalização dos receptores TLR (Receptor tipo Toll) (Medzhitov et al., 1997), sendo confirmado *in vivo* utilizando ratos deficientes para esse gene (Kawai et al., 1999).

Na via NF- $\kappa$ B ativada pelos TLRs o MyD88 tem o papel de acoplamento, formando assim um complexo protéico. A proteína Myd88 possui um domínio TIR C-terminal e um DM N-terminal. Em mamíferos a porção TIR interage com a porção de mesmo nome no receptor celular, tanto da família TLR quanto IL-R,

enquanto que a porção DM interage com a região de mesmo nome encontrada também na proteína citoplasmática IRAK. Deste modo, o estímulo que se origina fora da célula, em seus receptores, é passado para dentro da célula, seguindo a via (Warner e Núñez, 2013).

Com o avanço da bioinformática, temos muitas informações em banco de dados e muitas ferramentas computacionais que nos possibilitam analisar vários aspectos relacionados às sequências já depositadas (Galperin e Fernández-Suárez, 2012; Wren e Bateman, 2008). Através de análises das sequências de proteínas e ácidos nucleicos podemos estabelecer relações evolutivas entre diversas espécies, analisar aspectos de regiões específicas a nível de domínios e resíduos dessas sequências. Esse estudo tem como objetivo analisar dados referentes à proteína MyD88 e seus domínios e região, TIR, INT e DM em relação a sua funcionalidade na proteína entre as espécies.



## MATERIAIS E MÉTODOS

As sequências de MyD88 foram retiradas do banco de dados KEGG (<http://www.genome.jp/kegg/>), um total de 46 ortólogos, as siglas, nome das espécies e números dos bancos de dados das sequências estão listados na **tabela 1**. Foram selecionados as sequências completas e três regiões dentro da sequência, DM, INT e TIR, devidamente delimitadas no banco de dados do NCBI (<http://www.ncbi.nlm.nih.gov/>), e confirmadas nos bancos Uniprot (<http://www.uniprot.org/>) e Ensembl (<http://www.ensembl.org/>).

Para calcular a identidade entre cada uma das regiões da MyD88, nós realizamos um alinhamento par-a-par utilizando o algoritmo de Needleman-Wunsch (Needleman e Wunsch, 1970) com uma matriz de substituição de identidade (aminoácidos idênticos  $\rightarrow$  1, aminoácidos diferentes ou gaps  $\rightarrow$  0). A pontuação total do alinhamento foi normalizada pelo comprimento da maior sequência a ser alinhada. A representação gráfica do resultados foram feitas utilizando o software Mathematica, Versão 10.3 (Wolfram Research, 2015) (<http://www.wolfram.com/>).

Posteriormente as sequências das regiões DM e TIR foram alinhadas utilizando o programa MUSCLE (Edgar, 2004a, 2004b) disponível dentro do programa MEGA6 (Tamura et al., 2013). O padrão de substituição de aminoácidos foi selecionado usando o mesmo programa, a topologia da árvore construída por máxima verossimilhança com um bootstrap de 1000 replicatas. Para a comparação de árvores foi criado uma topologia de árvores baseado nos dados do projeto Tree of Life, ToL (<http://tolweb.org/tree/>) (Maddison et al., 2007).

Um alinhamento de todas as sequências foi realizado utilizando o programa T-coffee Expresso (disponível em: <http://tcoffee.crg.cat/apps/tcoffee/do:expresso>) (Armougom et al., 2006) e os resíduos descritos como funcionalmente importantes foram comparados entre as espécies.

Sigla	Espécie	Nº BD	Sigla	Espécie	Nº BD
<b>amj</b>	<i>Alligator mississippiensis</i>	XP_006264638	<b>ngi</b>	<i>Nannospalax galili</i>	XP_008844716
<b>aml</b>	<i>Ailuropoda melanoleuca</i>	D2GZQ7	<b>nle</b>	<i>Nomascus leucogenys</i>	XP_003256906
<b>apla</b>	<i>Anas platyrhynchos</i>	XP_005031349.1	<b>oas</b>	<i>Ovis aries</i>	C8BKC7
<b>bacu</b>	<i>Balaenoptera acutorostrata scammoni</i>	XP_007191531	<b>ocu</b>	<i>Oryctolagus cuniculus</i>	XP_002723915
<b>bom</b>	<i>Bos mutus</i>	XP_005901199	<b>ola</b>	<i>Oryzias latipes</i>	H2M1J7
<b>bta</b>	<i>Bos taurus</i>	Q599T9	<b>pale</b>	<i>Pteropus alecto</i>	L5KRQ9
<b>cfa</b>	<i>Canis familiaris</i>	XP_534223	<b>phd</b>	<i>Pantholops hodgsonii</i>	XP_005975937
<b>cge</b>	<i>Cricetulus griseus</i>	G3HQV3	<b>phi</b>	<i>Pseudopodoces humilis</i>	XP_005527867
<b>cjc</b>	<i>Callithrix jacchus</i>	XP_002759780	<b>pon</b>	<i>Pongo abelii</i>	H2PB58
<b>cmk</b>	<i>Callorhinchus milii</i>	XP_007908574	<b>pps</b>	<i>Pan paniscus</i>	B3Y679
<b>dre</b>	<i>Danio rerio</i>	Q5XJ85	<b>pss</b>	<i>Pelodiscus sinensis</i>	E5KBR3
<b>ecb</b>	<i>Equus caballus</i>	F6TAV5	<b>ptr</b>	<i>Pan troglodytes</i>	B3Y678
<b>fca</b>	<i>Felis catus</i>	XP_003992302	<b>ptg</b>	<i>Panthera tigris altaica</i>	XP_007081109
<b>gga</b>	<i>Gallus gallus</i>	A5HNF6	<b>rno</b>	<i>Rattus norvegicus</i>	Q6Y1S1
<b>hgl</b>	<i>Heterocephalus glaber</i>	G5C5P6	<b>shr</b>	<i>Sarcophilus harrisii</i>	XP_003762699
<b>hsa</b>	<i>Homo sapiens</i>	Q99836	<b>ssc</b>	<i>Sus scrofa</i>	A2TF48
<b>lcm</b>	<i>Latimeria chalumnae</i>	H3BB23	<b>tgu</b>	<i>Taeniopygia guttata</i>	H0YQ76
<b>lve</b>	<i>Lipotes vexillifer</i>	XP_007462751	<b>tru</b>	<i>Takifugu rubripes</i>	A8QMS7
<b>mcc</b>	<i>Macaca mulatta</i>	B3Y683	<b>tup</b>	<i>Tupaia chinensis</i>	L8Y2R9
<b>mcf</b>	<i>Macaca fascicularis</i>	I7GLV7	<b>umr</b>	<i>Ursus maritimus</i>	XP_008708432
<b>mgp</b>	<i>Meleagris gallopavo</i>	XP_010710631	<b>xla</b>	<i>Xenopus laevis</i>	Q9DF60
<b>mmu</b>	<i>Mus musculus</i>	Q3U7M4	<b>xma</b>	<i>Xiphophorus maculatus</i>	XP_005797881
<b>mze</b>	<i>Maylandia zebra</i>	XP_004546813	<b>xtr</b>	<i>Xenopus tropicalis</i>	Q28DJ2

**Tabela 1.** Sigla, nome das espécies e número no banco de dados das espécies analisadas.

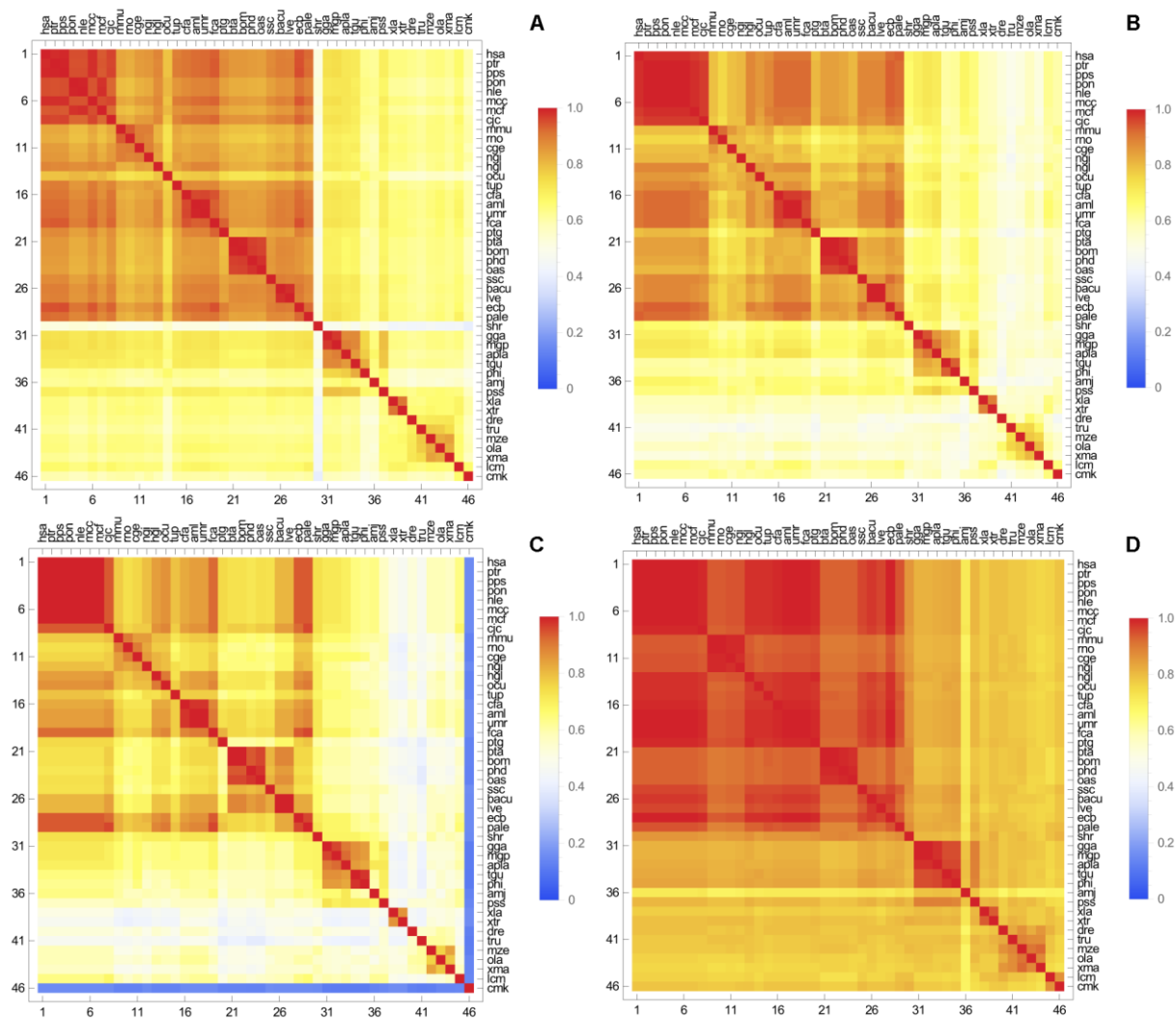
## RESULTADO E DISCUSSÃO

Como mencionado anteriormente, as sequências de MyD88 possuem dois domínios, o domínio DD e o domínio TIR, unidos por um segmento INT. Devido as características distintas dessas regiões, como o maior número de parceiros de interação do domínio TIR em relação às demais regiões, e a um possível maior grau de liberdade conformacional da região interdomínios, nós decidimos utilizar a identidade sequencial entre essas regiões da MyD88 para avaliar como essas características poderiam influenciar na conservação desta proteína multidomínio.

Conforme descrito nos métodos, a aplicação desse alinhamento par-a-par em todas as sequências de MyD88 estudadas nesse trabalho resultou em uma matriz, representada graficamente pela figura 2, para a sequência completa e cada uma das regiões.

A figura 2 possui uma escala de cores onde o vermelho apresenta 100% de identidade e o azul representa 0. Em todas as figuras o grupo que apresentou maior identidade foram os primatas, próximo a 100%. Em (D) vemos uma grande identidade entre todas as espécies, embora a espécie *Alligator mississippiensis* (amj) apresenta uma coloração mais próxima do amarelo, assim uma menor identidade que os demais. Ao passo que o domínio INT (C) apresenta menor conservação de modo geral. A espécie *Callorhinchus milii* (cmk) apresenta identidade menor (< 20%) devido a sua região INT ser menor (10aa) em relação às demais sequências INT das outras espécies ( $\bar{x}$  = 51,6aa) e também está mais distante evolutivamente das demais analisadas, formando um outgroup na árvore ToL (fig. 4.A) que discutiremos mais adiante. Enquanto em (B) todos os mamíferos apresentam identidade entre (80% a 90%). Em (A) a espécie *Sarcophilus harrisii* (shr) apresenta uma identidade menor (< 60%) aparentemente pelo tamanho de sua sequência MyD88 completa (350aa), que está acima da média referente às outras espécies ( $\bar{x}$ =297,7aa), essa diferença é devido às mudanças relacionadas com duplicações e deleções no processo evolutivo do gene das espécies analisadas. Este resultado nos mostra que o MyD88 de mamíferos, tanto da sequência completa, quanto dos seus domínios separados, principalmente primatas, possuem mais identidade entre eles do que as demais espécies analisadas.

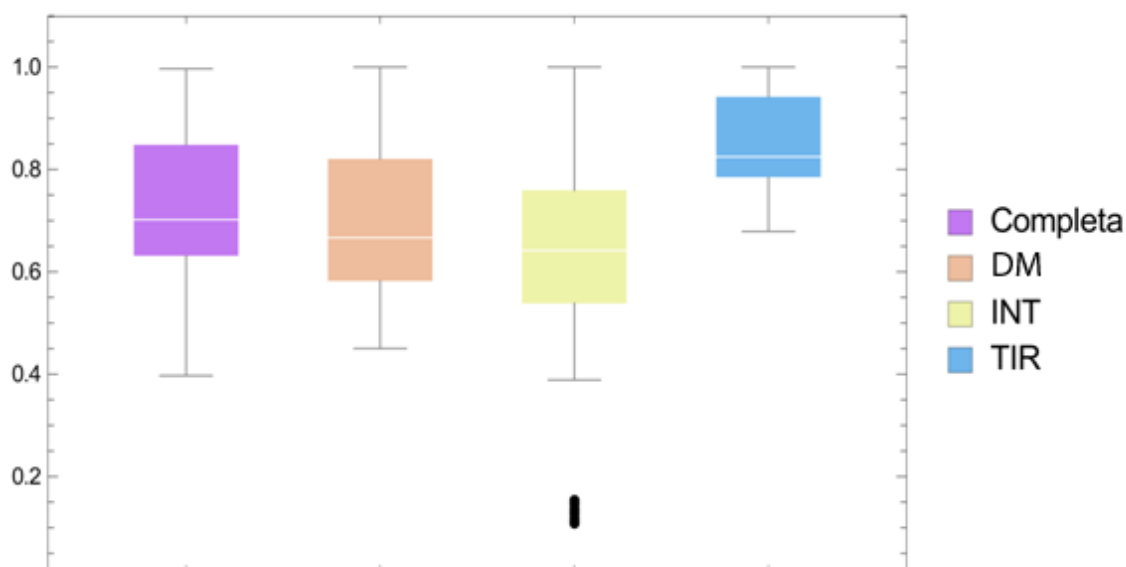
O alinhamento múltiplo das proteínas MyD88 mostrou que há uma alta similaridade entre as sequências de aminoácidos. Uma parcela desses aminoácidos são idênticas entre as sequências indicando uma possível importância funcional, por exemplo, como componentes da interface de interação com outras proteínas, ou estrutural, sendo essenciais para o enovelamento correto das MyD88.



**Figura 2.** Heat Map da sequência completa (A), domínio DM (B), região INT (C) e domínio TIR (D).

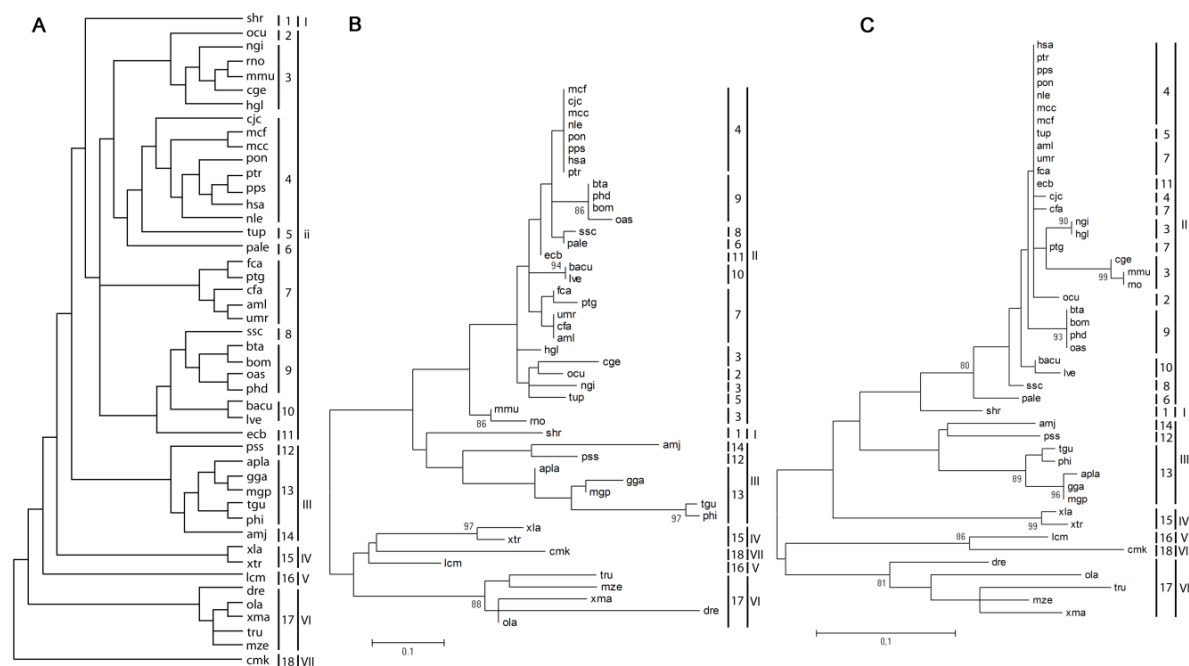
Confirmamos em gráfico BoxPlot a distribuição da identidade sequencial entre as proteínas completas e suas regiões nas 46 espécies comparadas (figura

3.)



**Figura 3.** BoxPlot da distribuição de identidade entre a sequência e regiões. O gráfico mostra a maior identidade entre os domínios TIR, entre (60 e 100%), seguido da sequência completa (40 a 100%), domínio DM (45% a 100%) e a região INT 40% e 100% mais alguns outliers em preto.

Os gráficos das porcentagens de identidade entre as regiões demonstram uma maior conservação na região do domínio TIR em relação aos demais, o que é ainda mais evidente analisando as distribuições de identidade entre os domínios (fig. 3). Não temos conhecimento de trabalhos na literatura que comparam a diferença de conservação entre os domínios de proteínas multidomínio, mas acreditamos que, pelo menos nesse caso específico, a conservação possa ser uma consequência do maior número de parceiros de interação desse domínio.



**Figure 4.** Análise Filogenética Molecular pelo método de Máxima Verossimilhança. Histórico evolutivo de ToL (A), Região DM da proteína MyD88 (B) e Região TIR da proteína TIR (C) são mostrados. Árvores B e C são construídas em escala, com medidas de comprimento de ramos em número de substituição por sítios. As análises envolveram 46 seqüências de aminoácidos. Todas posições que continham intervalos e dados faltante foram eliminados. Os grupos foram classificados em números romanos (I - marsupialia, II - eutheria, III - reptilia, IV - amphibia, V - coelacanthimorpha, VI - neopterygii, VII - chondrichthyes) e os subgrupos foram classificados em números arábicos (1 - dasyuromorphia, 2 - logomorpha, 3 - roedentia, 4 - primatas, 5 - scandentia, 6 - chiroptera, 7 - carnivora, 8 - suidae, 9 - ruminantia, 10 - cetacea, 11 - perissodactyla, 12 - testudines, 13 - neornithes, 14 - crocodylomorpha, 15 - amphibia, 16 - coelacanthimorpha, 17 - neopterygii, 18 - chondrichthyes).

As árvores dos domínios de MyD88 (Fig. 4) tanto da região DM (B) quanto da região TIR (C) apresentaram similaridades de ramos em relação à árvore da vida (A). Os modelos foram selecionados pelo programa MEGA6. O modelo mais apropriado para explicação da evolução da árvore (B) foi o JTT. com os seguintes parâmetros: BIC de 3016,465; AICc de 2492,804; lnL de -1153,224. Sítios

invariantes n/a, e parâmetros Gamma de 1,08 (Jones et al., 1992). O modelo na árvore (C) foi o Dayhoff com os seguintes parâmetros: BIC de 3173,528; AIC de 2585, 871; lnL de -1201,346. Sítios invariantes n/a, e parâmetros Gamma de 0,29 (Schwarz e Dayhoff, 1979). Todas as espécies que tiveram sua sequência analisada são pertencentes ao grupo dos Teleostomi.

GRUPOS											
Árvores	18, 16	3 <sup>a</sup>	3 <sup>b</sup>	9	10	13 <sup>c</sup>	13 <sup>d</sup>	13 <sup>e</sup>	15	17	II
DM	-	-	86	86	94	-	< 80	97	97	88	< 80
TIR	86	99	< 80	93	< 80	96	89	< 80	99	81	80

**Tabela 2.** Valores de bootstraps dos grupos formados na árvore B (DM) e C (TIR).

<sup>a</sup> mmu, rno, cge. <sup>b</sup> mmu, rno.

<sup>c</sup> apla, gga, mgp, tgu, phi. <sup>d</sup> apla, gga, mgp. <sup>e</sup> tgu, phi.

As sequências dos domínios DM e TIR de ruminates (9), anfíbios (15) e neopterygii (17) analisados demonstraram uma forte relação (bootstrap > 80) com uma sequência ancestral comum. O domínio TIR mostrou formar mais grupos monofiléticos com força de nós acima de 80 do que o domínio DM.

A região TIR se acopla aos receptores TLRs na via NF-κB, exceto o TLR3, e também IL-1, IL-18, e IL-33 (Dunne e O'Neill, 2003), essa variação de TIR e TLRs se torna mais evidente ao sabermos que Deuterostômios possuem 222 TLRs e até 26 adaptadores TIR potenciais, e o anfioxo que é cephalochordata possui pelo menos 39 TLRs e 40 adaptadores TIR (Huang et al., 2008). Com essa alta variabilidade entre o próprio TIR e seus ligantes nos receptores dentre as espécies animais podemos prever, juntamente com nossa análise, que esses domínios mantiveram sua conservação ao longo da evolução, e foram selecionados ao passo que mantinham sua funcionalidade dentro da via. A porção DD no Myd88 é responsável por sua ligação à proteína IRAK4 juntamente com IRAK2 ou IRAK1 (Deguine e Barton, 2014, Loiarro et al., 2007), essa porção mostrou nas nossas análises uma relação monofilética nos grupos 9, 10, 15, 17, 3<sup>b</sup> e 13<sup>e</sup>. Isso nos mostra que os dois domínios importantes na sequência MyD88 dos muridae, galloanserae, ruminantia, amphibia e neopterygii estudados

mantiveram uma forte relação de parentesco ao longo da evolução. Embora os domínios DM e TIR com grande identidade e conservação em primatas (fig. 2) não mostraram uma relação monofilética nas árvores referente aos domínios (fig. 4B e C), mostrando a importância de diferentes tipos de análise para estudar a relação entre ortólogos nas diferentes espécies. A região INT não foi analisado filogeneticamente devido ao tamanho curto da sequência não apresentando uma árvore concisa.

Em estudos com sequências mutantes, foram descritos 6 resíduos presentes na superfície do domínio DM (V43, A44, E52, Y58, I61 e R62) como essenciais ao recrutamento da molécula IRAK4 e formação do Myddosome (Lin et al., 2010). Outro trabalho mostrou no domínio TIR resíduos importante na formação de homodímero de MyD88 (TIR-TIR), Dentre os 11 mutantes analisados (D162A, E183A, N186A, R188A, D195A, S242A, S244A, H248A L252P, K281I, R288G e K291M), três deles, E183A, S244A e R288G impediram a formação do homodímero; a mutação do resíduo L252A aumentou a formação do dímero, todos presentes na interface da estrutura (Loiarro et al., 2013).

	DM						TIR			
SIGLA	43	44	52	58	61	62	188	244	252	288
hsa	V	A	E	Y	I	R	R	S	L	R
ptr	V	A	E	Y	I	R	R	S	L	R
pps	V	A	E	Y	I	R	R	S	L	R
pon	V	A	E	Y	I	R	R	S	L	R
nle	V	A	E	Y	I	R	R	S	L	R
mcc	V	A	E	Y	I	R	R	S	L	R
mcf	V	A	E	Y	I	R	R	S	L	R
cjc	V	A	E	Y	I	R	R	S	L	R
mmu	V	A	E	Y	I	R	R	S	L	R
rno	A	A	E	Y	I	R	R	S	L	R
cge	V	A	E	Y	I	R	R	S	L	R
ngl	M	A	E	Y	I	R	R	S	L	R
hgl	V	A	E	Y	I	R	R	S	L	R
ocu	V	A	E	Y	I	R	R	S	L	R
tup	V	A	E	Y	I	R	R	S	L	R
cfa	V	A	E	Y	I	R	R	S	L	R
aml	V	A	E	Y	I	R	R	S	L	R
umr	V	A	E	Y	I	R	R	S	L	R
fca	V	A	E	Y	I	R	R	S	L	R
ptg	V	A	E	Y	I	R	R	S	L	R



bta	V	A	E	Y	I	Q	R	S	L	R
bom	V	A	E	Y	I	Q	R	S	L	R
phd	V	A	E	Y	I	Q	R	S	L	R
oas	V	A	E	Y	I	Q	R	S	L	R
ssc	V	A	E	Y	I	R	R	S	L	R
bacu	V	A	E	Y	I	R	R	S	L	R
lve	V	A	E	Y	I	R	R	S	L	R
ecb	V	A	E	Y	I	R	R	S	L	R
pale	V	A	E	Y	I	R	R	S	L	R
shr	V	S	E	Y	I	L	R	C	L	R
gga	T	A	E	Y	I	R	K	S	L	R
mgp	T	A	E	Y	I	R	K	S	L	R
apla	A	A	E	Y	I	R	K	S	L	R
tgu	A	A	E	F	I	R	K	S	L	R
phi	A	A	E	F	I	R	K	S	L	R
amj	V	A	E	F	I	R	K	S	L	K
pps	V	A	E	Y	I	K	K	S	L	R
xla	V	A	E	Y	I	R	K	G	L	K
xtr	V	A	E	Y	I	K	K	G	L	R
dre	V	A	E	Y	I	K	N	C	L	R
tru	V	A	E	F	I	K	K	S	L	R
mze	V	A	E	Y	I	K	K	C	L	R
ola	V	A	E	Y	I	K	K	C	L	R
xma	V	A	E	Y	I	K	N	C	L	R
lcm	V	A	E	Y	I	K	K	C	L	R
cmk	V	A	E	Y	I	R	H	C	L	R

**Tabela 3.** Resíduos alinhados através da ferramenta T-coffee Expresso. Sítios com grande importância no funcionamento dos domínios DM e TIR de MyD88. Os resíduos em negro foram conservado em todas sequências ao longo da evolução, em vermelho estão os resíduos que sofreram substituições no processo evolutivo.

Os sítios 52 e 61 mostraram conservação nas 46 sequências analisadas, já os sítios 43,44,58 e 62 demonstraram poucas variações. Esses resíduos do domínio DD apresentaram grande conservação, mesmo naqueles que sofreram substituição ao longo da evolução, confirmando sua importância na interação MyD88-IRAK4. Na porção TIR, 3 resíduos (188, 244 e 288), já citados como importantes na homodimerização entre dois domínios TIR, quando estes sofrem mutações não sinônimas podem impedir essa dimerização, O resíduo L252 apresentou conservação em todos organismos. Como esperado, resíduos descritos como funcionalmente importantes assim como regiões de interface nas proteínas homólogas apresentaram um alto grau de conservação (Zhang et al.,

2010). Entretanto, observamos alguns padrões de substituições de aminoácidos entre as espécies, que podem ocorrer por diversos fatores, como por exemplo: presença/ausência de alguns parceiros de interação; variação de resíduos nos parceiros de interação (coevolução); ou diferenças de afinidade na interação que podem ser importantes na regulação da via nas diferentes espécies.

## CONCLUSÃO

A sequência completa e os domínios de MyD88 em primatas possuem maior identidade em relação às outras espécies analisadas. Foi reforçado o fato entre alta similaridade de alguns aminoácidos da interface com a funcionalidade da proteína. O domínio TIR apresentou maior identidade entre os domínios e uma maior ancestralidade da sua sequência em relação à região DM, podendo sugerir uma relação com o maior número de parceiros (receptores) de interação. Apesar de encontrar um padrão de substituições não sinônimas nos resíduos críticos das regiões do MyD88 nas 46 espécies analisadas, eles em sua maioria se mostram conservados ao longo da evolução. Há uma grande necessidade de mais estudos que visam analisar diferenças de conservação de domínios em proteínas multidomínio.

## REFERÊNCIAS

Adachi, O., Kawai, T., Takeda, K., Matsumoto, M., Tsutsui, H., Sakagami, M., Nakanishi, K., Akira, S., 1998. Targeted disruption of the MyD88 gene results in loss of IL-1- and IL-18-mediated function. *Immunity*. 9, 143–150.

Armougom, F., Moretti, S., Poirot, O., Audic, S., Dumas, P., Schaeli, B., Keduas, V., Notredame, C., 2006. Espresso: automatic incorporation of structural information in multiple sequence alignments using 3D-Coffee. *Nucleic Acids Res.* 1;34 (Web Server issue):W604-8. Acessado em 12 de novembro de 2015.

Bazan, J.F., Timans J.C., Kastelein, R.A., 1996. A newly defined interleukin-1? *Nature*. 379, 591.

Deguine, J., Barton, G.M., 2014. MyD88: a central player in innate immune signaling. *F1000Prime Rep.* 6, 97, 1-7.

Dunne, A., O'Neill L.A., 2003. The interleukin-1 receptor/Toll-like receptor superfamily: signal transduction during inflammation and host defense. *Sci STKE*. 177, re3.

Edgar, R.C., 2004a. MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res.* 32, 1792–1797.

Edgar, R.C., 2004b. MUSCLE: a multiple sequence alignment method with reduced time and space complexity. *BMC Bioinformatics*. 5, p. 113.

Galperin, M.Y., Fernández-Suárez, X.M., 2012. The 2012 Nucleic Acids Research Database Issue and the online Molecular Biology Database Collection. *Nucleic Acids Res.* 40, D1–D8.

Huang, S., Yuan, S., Guo, L., Yu, Y., Li, J., Wu, T., Liu, T., Yang, M., Wu, K., Liu H., Ge, J., Yu, Y., Huang, H., Dong, M., Yu, C., Chen, S., Xu, A., 2008. Genomic

analysis of the immune gene repertoire of amphioxus reveals extraordinary innate complexity and diversity. *Genome Res.* 18, 1112–1126.

Hultmark, D., 1994. Macrophage differentiation marker MyD88 is a member of the Toll/IL-1 receptor family. *Biochem. Biophys. Res. Commun.* 199, 144–146.

Jones, D.T., Taylor, W.R., Thornton, J.M., 1992. The rapid generation of mutation data matrices from protein sequences. *Comput. Appl. Biosci.* 8, 275-282.

Kawai, T., Adachi, O., Ogawa, T., Takeda, K., Akira, S., 1999. Unresponsiveness of MyD88-deficient mice to endotoxin. *Immunity.* 11, 115–122.

Lin, S.-C., Lo, Y.-C., Wu, H., 2010. Helical assembly in the MyD88:IRAK4:IRAK2 complex in TLR/IL-1R signaling. *Nature*, 465, 885–890.

Loiarro, M., Capolunghi, F., Fantò, N., Gallo, G., Campo, S., Arseni, B., Carsetti, R., Carminati, P., De Santis, R., Ruggiero, V., Sette, C., 2007. Pivotal Advance: Inhibition of MyD88 dimerization and recruitment of IRAK1 and IRAK4 by a novel peptidomimetic compound. *J. Leukoc. Biol.* 82, 801–810.

Loiarro, M., Volpe, E., Ruggiero, V., Gallo, G., Furlan, R., Maiorino, C., Battistini, L., Sette, C., 2013. Mutational analysis identifies residues crucial for homodimerization of Myeloid Differentiation Factor 88 (MyD88) and for its function in immune cells. *J. Biol. Chem.* 288, 30210-30222.

Lord, K.A., Abdollahi, A., Hoffman-Liebermann, B., Liebermann, D.A., 1990a. Dissection of the immediate early response of myeloid leukemia cells to terminal differentiation and growth inhibitory stimuli. *Cell Growth Differ.* 1, 637-645.

Lord, K.A., Hoffman-Liebermann, B., Liebermann, D.A., 1990b. Nucleotide sequence and expression of a cDNA encoding MyD88, a novel myeloid differentiation primary response gene induced by IL6. *Oncogene.* 5, 1095-1097.

Maddison, D. R., Schulz, K.-S., Maddison, W.P., 2007. The Tree of Life Web Project. *Zootaxa*.1668, 19-40.

Medzhitov, R., Preston-Hurlburt, P., Janeway Jr, C.A., 1997. A human homologue of the *Drosophila* Toll protein signals activation of adaptive immunity. *Nature*. 388, 394–397

Medzhitov, R., Preston-Hurlburt, P., Kopp, E., Stadlen, A., Chen, C., Ghosh, S., and C. A. Janeway Jr. 1998. MyD88 is an adaptor protein in the hToll/IL-1 receptor family signaling pathways. *Mol. Cell*. 2, 253–258.

Muzio, M., Ni, J., Feng, P., Dixit, V.M., 1997. IRAK (Pelle) family member IRAK-2 and MyD88 as proximal mediators of IL-1 signaling. *Science*. 278, 1612–1615.

Needleman, S.B., Wunsch, C.D., 1970. A general method applicable to the search for similarities in the amino acid sequence of two proteins. *JMol. Biol.* 48, p. 443.

Robinson, D., Shibuya, K., Mui, A., Zonin, F., Murphy, E., Sana, T., Hartley, S.B., Menon, S., Kastelein R., Bazan F., O'Garra, A., 1997. IGIF does not drive Th1 development but synergizes with IL-12 for interferon- $\gamma$  production and activates IRAK and NF $\kappa$ B. *Immunity*. 7, 571–581.

Schwarz, R., Dayhoff, M., 1979. Matrices for detecting distant relationships. In Dayhoff, M. (Eds.), *Atlas of protein sequences*. NABR, pp. 353 - 358.

Tamura, K., Stecher, G., Peterson, D., Filipski, A., Kumar, S., 2013. MEGA6: Molecular Evolutionary Genetics Analysis Version 6.0. *Mol. Biol. Evol.* 30, 2725–2729.

Warner, N., Núñez G., 2013. MyD88: A Critical Adaptor Protein in Innate Immunity Signal Transduction. *The JI*. 190, 3-4.

Wesche, H., Henzel, W.J., Shillinglaw, W., Li, S., Cao, Z., 1997. MyD88: an adapter that recruits IRAK to the IL-1 receptor complex. *Immunity*. 7, 837–847.

Wesche, H., W. J. Henzel, W. Shillinglaw, S. Li, Z. Cao., 1997. MyD88: an adapter that recruits IRAK to the IL-1 receptor complex. *Immunity*. 7, 837–847.

Wolfram Research, Inc., 2015. Mathematica, Version 10.3, Champaign, Illinois. <http://www.wolfram.com/> Acessado em 11 de novembro de 2015.

Wren, J.D., Bateman, A., 2008. Databases, data tombs and dust in the wind. *Bioinformatics*. 24, 2127–2128.

Zhang, Q.C., Petrey, D., Norel, R., Honig, B.H., 2010. Protein interface conservation across structure space. *PNAS*. 107, 10896–10901.