



**UNIVERSIDADE FEDERAL DE MINAS GERAIS
INSTITUTO DE CIÊNCIAS BIOLÓGICAS
DEPARTAMENTO DE BIOQUÍMICA E IMUNOLOGIA
PROGRAMA DE PÓS-GRADUAÇÃO EM BIOINFORMÁTICA
DISSERTAÇÃO DE MESTRADO**

**Estudo da ancestralidade de genes componentes do sistema
imune**

Verônica Rodrigues de Melo Costa

**Belo Horizonte
Janeiro
2015**

Dissertação de Mestrado
Estudo da ancestralidade de genes componentes do sistema imune

Dissertação de mestrado apresentada ao Programa de Pós-Graduação em Bioinformática do Instituto de Ciências Biológicas da Universidade Federal de Minas Gerais como requisito parcial para obtenção do título de Mestre em Bioinformática.

Orientador: Prof^o Dr. José Miguel Ortega

Belo Horizonte
Janeiro
2015

Dedico este trabalho à Mitsy, Harry e Cogo.

Agradecimentos

Começo agradecendo a todos os membros do Laboratório de Biodados pelos anos de convivência tão agradável. Agradeço especialmente ao Gabriel Fernandes, por me apresentar a bioinformática, ao Rafael Guedes, pela paciência, ao Henrique Velloso pelo carinho, amizade e apoio, e ao Diego Souza, Tetsu Sakamoto e Ricardo Vialle pelas incontáveis vezes em que estiveram a disposição para me ajudar.

Ao Super KRAV, pelo companheirismo, amizade e muitas risadas. Vocês são demais!

Ao apoio, ensinamentos, boas ideias e auto astral do meu orientador José Miguel Ortega.

Ao Programa de Pós-Graduação em Bioinformática da UFMG pela oportunidade e, especialmente, à Sheila e Ana Paula que sempre me receberam com carinho e disposição.

Ao grupo do Dr. Miguel Andrade do MDC, na Alemanha, que me recebeu muito bem e me ensinaram muito enquanto estive lá.

Agradeço aos amigos da biologia pela amizade e pelos vários momentos de descontração, em especial ao Mateus Scarpelli, Thomaz Luscher e Julia Thompson.

À Carolina Andrade, Karla Pollyanna e Tiago Bruno por estarem sempre presentes.

À Flávia Motta, Lucas Reuter, Pedro Reuter e Rafael Reuter pelo companheirismo e carinho, mesmo à distância.

À Luisa Padua, Luíza Bittencourt, Laura Velloso, Monick Ribeiro, Stéphanie Resende e Kleana de Paula pelos anos de amizade e por compreenderem a minha ausência nos momentos em que foram substituídas pelo trabalho.

Agradeço imensamente às minhas avós, tios, tias, primos, primas e Daniel por acreditarem em mim e sonharem junto comigo.

Por fim, termino agradecendo imensamente àqueles a quem dedico este trabalho: ao meu pai Márcio, à minha mãe Rose e ao meu irmão Victor. Muito obrigada por todo amor e confiança que tiveram em mim. Se hoje finalizo mais essa etapa, é porque vocês sempre estiveram ao meu lado. Amo vocês!

*"Much to learn you still have... my old padawan.
This is just the beginning!"*

- Mestre Yoda

Resumo

O sistema imune evoluiu de modo a formar uma rede de células, tecidos e órgãos que protegem o organismo contra microrganismos patogênicos, células tumorais e de outros antígenos. A defesa dos organismos contra microrganismos é basicamente mediada pelas reações da imunidade natural (nativa ou inata) e pelas reações da imunidade adquirida (específica ou adaptativa). O presente trabalho irá mostrar um estudo sobre a origem de genes componentes de importantes subsistemas do sistema imune, sendo eles as vias de sinalização de interferons, receptores de células T e B. O mesmo foi feito para marcadores de células *Natural Killer* e para os receptores semelhantes a *Toll*. Com o objetivo de delinear a construção deste sistema, buscaram-se, a partir de curadoria manual, as sequências de aminoácidos dos representantes humanos e estas foram usadas como *Seed* para procurar homólogos, uma vez que grupos de proteínas homólogas são importantes fontes para propagação da informação biológica, bem como para a inferência da origem dos genes que as codificam. Ao examinar a distribuição taxonômica desses homólogos, foi determinado o *Lowest Common Ancestor* (LCA) desses genes, ou seja, foi possível retratar os clados, na linhagem humana, onde cada proteína ocorreu pela primeira vez.

O complexo receptor de células T (TCR) aparece tardiamente, em Eutheria, e todas as cadeias de CD3 têm as suas subunidades que aparecem ainda mais recentemente, em Euteleostomi (cadeia zeta) e Amniota (cadeias delta e epsilon). Este complexo liga-se a antígenos virais e tumorais, que estão associados com MCH classe I, enquanto agentes patogênicos bacterianos e outros antígenos são encontrados associados a MHC de classe II; os componentes para ambos MHC surgiram em Euteleostomi. No receptor de células B (BCR), o heterodímero Ig-alfa/Ig-beta apareceu também em Euteleostomi. Muitos outros receptores transmembrana são conhecidos por modular elementos específicos da sinalização de BCR e eles surgiram em diferentes clados: CD45 em Bilateria e FcGR1IB em Eutheria; há também um complexo de correceptores constituído por CD21 e CD19, originados em Amniota e Eutheria, respectivamente. Em relação à resposta imune anti-viral, os interferons (IFNs) e os seus receptores cognatos são distribuídos entre os clados Euteleostomi (IFN α , IFN γ e IFNAR1), Tetrapoda (IFNAR2, IFNGR1, IFNGR2, IL10RB e

IL28RA), Amniota (IFN β , IFN κ , IFN ϵ e INFs tipo III) e Eutheria (IFN ω). Assim, aparentemente a origem de ligantes e receptores cognatos é separada no tempo, embora o intervalo seja pequeno. Para a imunidade natural, os principais componentes apareceram entre os clados Bilateria e Euarchontoglires. Em linhas gerais, o LCA de 385 genes foi inferido, sendo 24 deles componentes extracelulares, 84 membros da membrana plasmática, 133 do citosol, 21 do núcleo e 63 genes que são ativados com a ativação das vias de IFN. De um modo geral, observa-se que a sinalização na membrana aparece mais tardiamente, em momentos críticos do início do funcionamento do sistema. Estes genes foram escolhidos devido a sua importância central nas respostas imunológicas. Em conclusão, o agrupamento de homólogos seguido da inferência de ancestralidade retratou a evolução do sistema imunológico, demonstrando que o estabelecimento da resposta engloba, de um modo geral, o período que se inicia juntamente com a origem dos ossos, em Euteleostomi, até o advento dos Eutheria, sendo que a fixação do sistema sempre se conclui com o aparecimento dos elementos de membrana.

Abstract

The immune system evolved to form a network of cells, tissues and organs that protect the organism against pathogenic microorganisms, tumor cells and other antigens. The organisms defense against microorganisms is basically mediated by the reactions of natural immunity (native or innate) and by the reactions of acquired immunity (specific or adaptive). will display a study of the origin of genes of important components subsystems of the immune system, namely interferons, T and B cell receptors signaling pathways. The same was performed for Natural Killer cells markers and the Toll-like receptors. Aiming to delineate the raising of this network, the amino acid sequences of the human representatives were mined and used as Seed sequences to search for homologues, since groups of homologous are important sources for the propagation of biological information and for inferring the origin of the genes that encode them. By examining the taxonomic distribution of the homologues, the Lowest Common Ancestor (LCA) of these genes was determined and it was possible to depict the clades of the human lineage where each protein has first occurred. The T-cell receptor complex (TCR) appears very late, in Eutheria, and all the chains of CD3 have its subunits appearing more recently, in Euteleostomi (zeta chain) and Amniota (delta and epsilon chains). This complex binds to viral and tumor antigens, which are associated with class I MHC, while bacterial and other pathogens antigens, are associated to class II MHC. The components for both MHC emerged very early, in Euteleostomi. In B-cell receptor complex (BCR), the Ig-Alpha/Ig-Beta heterodimer also appeared in Euteleostomi. Many other transmembrane receptors are known for modulating specific elements of BCR signaling and they arose in different times: CD45 in Bilateria and FcGRIIB in Eutheria; there is also a co-receptor complex composed of CD21 and CD19, originated in Amniota and Eutheria, respectively. In respect to the immune response to virus, the interferons (IFNs) and its cognate receptors are distributed between the clades Euteleostomi (IFN α , IFN γ e IFNAR1), Tetrapoda (IFNAR2, IFNGR1, IFNGR2, IL10RB e IL28RA), Amniota (IFN β , IFN κ , IFN ϵ e type III INFs) and Eutheria (IFN ω). Thereby, the origin of ligands and its cognate receptors are separated in time, although it is a short interval. For natural immunity, the main components appeared among clades Bilateria and Euarchontoglires. In general, the LCA of

385 genes were inferred, wherein 24 of them extracellular components, 84 were members of the plasma membrane, 133 of cytosol, 21 of nucleus and 63 genes that are activated due to the activation of the IFN pathways. In general, it was observed that the signal on the membrane appears later, in critical moments of the beginning of the system operation. These genes had been chosen due to their central role in immune responses. In conclusion, bioinformatics clustering of homologues followed by ancestry inference depicted the evolution of immune response, demonstrating that the establishment of the system mostly encompassed the period initiated with the origin of bones, in Euteleostomi, up to the advent of Eutheria.

I. SUMÁRIO

I. SUMÁRIO.....	1
II. ÍNDICE DE FIGURAS.....	4
III. ÍNDICE DE TABELAS.....	6
IV. ABREVIACÕES E SIGLAS.....	7
V. INTRODUÇÃO.....	9
1. O sistema Imune	9
2. Tipos de resposta imunológica.....	9
2.1 Resposta imunológica natural.....	11
2.1.1 Componentes do sistema imunológico natural.....	12
2.2 Resposta imunológica adquirida.....	14
2.2.1 Células e tecidos do sistema imunológico adquirido.....	15
3. Mecanismos associados aos linfócitos T.....	17
3.1 Reconhecimento de antígenos - MHC.....	17
3.2 Processamento e apresentação de antígenos.....	19
3.3 Receptor TCR e co-receptores.....	20
4. Mecanismos efetores da resposta imune.....	23
4.1 Imunidade mediada por células.....	23
4.1.1 Aspectos da imunidade anti viral.....	25
4.2 Imunidade humoral: ativação das células B e produção de anticorpos.....	26
5. Aspectos evolutivos do sistema imune.....	28
6. Homologias e conceitos.....	30
6.1 Métodos de agrupamento de sequências homólogas.....	33
7. Bancos de dados biológicos.....	35
7.1 Uniprot.....	35
7.2 KEGG.....	36
7.3 NCBI Taxonomy	38
7.4 Wikipathways.....	39
8. Ferramentas e serviços.....	40
8.1 SeedServer.....	40

8.1.1 Seed Likage.....	41
8.1.2 UniRef50 Enriched KEGG Orthology (UEKO).....	42
8.2 Algoritmo de LCA.....	43
8.3 Genesis.....	44
8.4 MaLe-PSI-BLAST.....	45
8.5 PathVisio.....	47
8.6 MEGA.....	47
VI. JUSTIFICATIVA.....	49
VII. OBJETIVOS.....	50
1. Objetivo geral.....	50
2. Objetivos específicos.....	50
VIII. METODOLOGIA.....	51
1. Obtenção de sequências e informações.....	51
1.1 Seleção das vias de sinalização.....	51
1.2 Seleção dos identificadores <i>Uniprot</i>	52
2. Recursos computacionais e programas utilizados.....	54
2.1 <i>Seedserver</i>	54
2.2 Análise do último ancestral comum (LCA)	55
2.3 Genesis.....	55
3. MaLe-PSI-Blast.....	56
4. MEGA	57
IX. RESULTADOS.....	58
1. Construção da sinalização associada aos Interferons tipo I.....	58
1.1 Necessidade de inspeção manual e o caso de JAK1.....	67
2. Construção da sinalização associada aos Interferons Tipo II e Tipo III.....	68
2.1 Origem dos Fatores Reguladores de Interferon.....	72
2.2 Busca de um ancestral remoto para receptor de IFN γ	73
3. Construção da sinalização associada ao Complexo TCR.....	74
4. Construção da sinalização associada ao Complexo BCR.....	83
5. Células <i>Natural Killer</i> e receptores semelhantes a Toll.....	89
5.1 Células <i>Natural Killer</i>	89
5.2 Receptores semelhantes a <i>Toll</i>	91

X. DISCUSSÃO.....	95
XI. CONCLUSÕES.....	101
XII. PERSPECTIVAS.....	103
XIII. REFERÊNCIAS.....	104
XIV. PRODUÇÃO CIENTÍFICA.....	113
1. Trabalhos apresentados em congressos.....	113
2. Prêmios.....	114
XV. ANEXOS.....	115
1. Via de sinalização TCR.....	115
2. Via de sinalização TCR	121
3. Via de sinalização IFNs	126
4. Marcadores de NK e TLRs	133

II. ÍNDICE DE FIGURAS

V. INTRODUÇÃO.....	9
Figura 1: Complexo TCR.....	22
Figura 2: Respostas imunológicas mediadas por células.....	23
Figura 3: Desenvolvimento da resposta mediada por CTLs.....	24
Figura 4: Complexo do receptor de linfócito B.....	27
Figura 5: Genes homólogos e suas subdivisões.....	32
Figura 6: Processos usados para organizar e anotar o banco de dados GENE.....	37
Figura 7: Resultado da busca por <i>Homo sapiens</i> no website.....	39
Figura 8: Enriquecimento do KO.....	43
Figura 9: Exemplos de cálculo de LCA.....	44
Figura 10: Interface do programa PathVisio.....	47
VIII. METODOLOGIA.....	51
Figura 1: Busca dos UniProt ID de p38.....	53
Figura 2: Árvore construída pela ferramenta MULTILCA TREE.....	56
IX. RESULTADOS.....	58
Figura 1: Distribuição dos tipos de interferon acordo com seu clado de origem ao longo linhagem evolutiva de <u>Homo sapiens</u>	59
Figura 2: Via de sinalização dos interferons tipo I.....	61
Figura 3: Proteínas do citosol na sinalização IFN tipo I.....	65
Figura 4: Proteínas do núcleo na sinalização IFN tipo I.....	65
Figura 5: Genes transcritos com a ativação da via IFN tipo I.....	66
Figura 6: Transferência lateral do gene JAK1	68
Figura 7: Via de sinalização dos interferons tipo II e III.....	69
Figura 8: Genes transcritos com a ativação da via IFN tipo II.....	71
Figura 9: Distribuição dos fatores de regulação de interferon de acordo com seu clado de origem ao longo da linhagem evolutiva de <u>Homo sapiens</u>	73
Figura 10: Via de sinalização associada ao TCR.....	76
Figura 11: Receptores co-estimulatórios da sinalização TCR.....	78
Figura 12: CaCn.....	78

Figura 13: Proteínas do citosol na sinalização TCR.....	82
Figura 14: Proteínas do núcleo na sinalização TCR.....	82
Figura 15: Componentes não-integrais de membrana da sinalização BCR.....	84
Figura 16: Via de sinalização associada ao BCR.....	86
Figura 17: Proteínas do citosol na sinalização BCR.....	88
Figura 18: Proteínas do núcleo na sinalização BCR.....	88
Figura 19: Marcadores de superfície de NK.....	90
Figura 20: Distribuição dos TLRs de acordo com seu clado de origem ao longo da linhagem evolutiva de <u>Homo sapiens</u>	91
Figura 21: Árvores filogenéticas para os TLR humanos.....	92
X. DISCUSSÃO	95
Figura 1: Padrão de surgimento total de genes.....	96

III. ÍNDICE DE TABELAS

V. INTRODUÇÃO.....	9
Tabela 1: Citocinas do sistema imunológico natural.....	14
Tabela 2: Características do MHC classes I e II.....	18
Tabela 3: Tabela final com os resultados do SeedServer para Interferon gama humano..	41
IX. RESULTADOS.....	58
Tabela 1: Resultado de PSI-BLAST para IFNGR1 e IFNGR2 humano.....;	74
Tabela 2: Distribuição de componentes integrais de membrana de acordo com seu clado de origem em relação a linhagem evolutiva de Homo sapiens.....	84
Tabela 3: Resultado de PSI-Blast para TLR5 humano.....	93

IV. ABREVIACOES E SIGLAS

**em ordem alfabética*

Ag	Antígeno
APC	Célula apresentadora de antígeno
BCR	Receptor de célula B
BLAST	<i>Basic Local Alignment Search Tool</i>
CTL	Linfócitos T citotóxicos
DBJ	<i>Data Bank of Japan</i>
EC	<i>Enzyme commission</i>
HLA	<i>Human Leukocyte Antigens</i>
Igs	Imunoglobulinas
KEGG	<i>Kyoto Encyclopedia of Genes and Genomes</i>
KO	<i>KEGG Orthology System</i>
LB	Linfócito B
LCA	<i>Lowest common ancestor</i>
LT	Linfócito T
MHC	Complexo principal de histocompatibilidade
NCBI	<i>National center for biotechnology information</i>
NK	Célula <i>natural killer</i>
PAMPs	Padrões moleculares associados à patógenos
PID	<i>Process ID</i>
PSI-BLAST	<i>Position-Specific Iterative BLAST</i>
SC	Sistema complemento
taxID	<i>Taxonomy ID</i>
Th	Linfócitos T auxiliares ou <i>helper</i>
TLRs	Receptores semelhantes a <i>Toll</i>
TNF	Fator de necrose tumoral
TReg	Linfócitos T reguladores
UFMG	Universidade Federal de Minas Gerais
UniMES	<i>UniProt Metagenomic and Environmental Sequences</i>

UniParc	UniProt <i>Archive</i>
Uniprot	<i>Universal protein resource</i>
UniProtBK	<i>Protein Knowledgebase</i>
UniRef	UniProt <i>Reference Clusters</i>
IL	Interleucina
UEKO	<i>UniRef50 Enriched KEGG Orthology</i>
SC	Sistema complemento
SL	<i>Seed Linkage</i>
BBH	<i>Bidirectional Best Hit</i>
IFN	Interferons
TLR	Receptores semelhantes a <i>Toll</i>

V. INTRODUÇÃO

1. O sistema imune

Derivado da palavra latina que se refere aos processos sofridos pelos senadores romanos durante seu mandato, *Immunitas*, o termo imunidade hoje se refere ao conjunto de células, tecidos e órgãos que defendem o organismo dos chamados processos infecciosos numa resposta coletiva e altamente regulada. Acredita-se que o termo foi utilizado primeiramente por Tucídides, na Grécia antiga, ao referir-se a uma peste, mas as práticas relacionadas à imunologia são ainda mais antigas, como o hábito chinês de milênios atrás de estimular a imunidade de crianças contra varíola através da inalação de um pó preparado com as secreções cutâneas de pessoas infectadas com a doença [1]. A principal função fisiológica desse sistema é a defesa não só contra microorganismos patogênicos, mas contra todo tipo de macromolécula capaz de deflagrar a resposta imunológica.

A seguir serão discutidos os principais aspectos e componentes do sistema imunológico.

2. Tipos de resposta imunológica

A defesa dos organismos contra microorganismos é basicamente mediada pelas reações da imunidade natural (nativa ou inata) e pelas reações da imunidade adquirida (específica ou adaptativa). A imunidade natural é a chamada primeira linha de defesa do organismo e consiste de mecanismos pré-existentes à infecção. Esse tipo de resposta é rápida e pouco específica, sendo específica apenas para estruturas comuns a grupos de

microrganismos. Neste caso, a resposta acontece de maneira semelhante para diferentes tipos de infecção e não há distinção entre moléculas que apresentam diferenças discretas.

Os principais componentes da resposta imune natural são [2]:

- Barreiras físicas e químicas: epitélio e antibacterianos naturais presentes na superfície do mesmo.
- Células *Natural Killer* (NK).
- Células fagocitárias.
- Proteínas do soro.
- Citocinas.

Outro tipo de resposta imune é a imunidade adquirida. Nesse contexto, há maior especificidade na distinção das macromoléculas e há formação de memória imunológica, ou seja, o sistema se torna mais específico e com respostas ampliadas nas infecções posteriores. Os principais componentes desse sistema são os linfócitos e o mesmo é induzido a partir do contato com o antígeno (Ag), que consiste em todo corpo estranho presente no organismo.

Esses dois sistemas atuam de forma integrada. A imunidade natural não só funciona como primeira tentativa do organismo de combater um corpo estranho, mas também funciona como estimulador da imunidade adquirida. Esse é um passo fundamental, uma vez que inúmeros organismos desenvolveram resistência a esse sistema, fazendo com que sua eliminação precise de uma resposta mais específica. Nesse contexto, é possível inferir que o sistema de defesa mais antigo filogeneticamente é o natural, enquanto o adquirido evoluiu posteriormente [3] e se tornou cada vez mais especializada com a evolução das espécies.

2.1 Resposta imunológica natural

Como mencionado anteriormente, a resposta natural é aquela que rapidamente defende o organismo contra uma invasão de forma pouco específica. O objetivo desta seção é conhecer o sistema imunológico inato destacando seus componentes, funções e interações.

A imunidade natural é formada por componentes mecânicos (epiderme e mucosas), químicos (citocinas e quimiocinas) e celulares. Esse conjunto de componentes faz com que esse sistema seja responsável pela construção da resposta inflamatória [1], onde há aumento da permeabilidade vascular permitindo a defesa contra agentes agressores.

Quando há dano na superfície epitelial, uma cascata de sinais resulta na produção de substâncias que indicam dano ao tecido e são estimuladoras do processo inflamatório. São elas as citocinas, quimiocinas, prostaglandinas e leucotrienos [4]. Tipos celulares, como células dendríticas e macrófagos também interagem com o organismo invasor através de receptores de reconhecimento de padrões especializados, localizados em sua superfície, reconhecendo padrões moleculares associados a patógenos (PAMPs), que são compartilhados por microrganismos. Os receptores semelhantes a Toll (TLRs) são os mais importantes reconhecedores de PAMPs [1, 5].

As PAMPs estão frequentemente relacionadas à sobrevivência do microrganismo e muitos deles desenvolveram resistência a esses mecanismos de defesa inata [6]. Entretanto, esse sistema ainda é fundamental quando se trata de controlar inicialmente uma invasão para que as respostas específicas do sistema adquirido possam atuar. A resposta natural influencia a adquirida e pode torná-la mais eficaz e específica,

ressaltando o que já foi dito anteriormente: a resposta imune ótima acontece a partir da resposta integrada desses dois sistemas.

2.1.1 Componentes do sistema imunológico natural

- **Barreiras epiteliais:** representam a primeira linha de defesa do organismo contra um patógeno e funcionam como uma barreira física que separa o ambiente externo e os tecidos do hospedeiro. Neste conjunto, encontram-se as superfícies mucosas do trato respiratório, gastrointestinal, genitália e a pele. Esses epitélios também produzem antimicrobianos e abrigam LT e LB intra-epiteliais, além de mastócitos [6].
- **Células efectoras:** são células circulantes capazes de migrar para os tecidos. São elas:

(i) Neutrófilos: Se originam na medula óssea a partir de uma linhagem de células comum aos fagócitos mononucleares. São leucócitos polimorfonucleares (núcleo segmentado) abundantemente distribuídos nos vasos, sendo responsáveis por mediar as fases iniciais da resposta inflamatória. Há presença de grânulos no citosol contendo defensinas, fosfolipases, lisozimas, lactoferrinas, dentre outros [7].

(ii) Fagócitos mononucleares: São células que também se originam na medula óssea, tendo como precursores os monoblastos e promonócitos. O primeiro tipo celular que entra no sangue é denominado monócito e ainda não está completamente diferenciado. Há diferenciação completa uma vez que essas células são recrutadas para os tecidos, onde podem assumir diferentes morfologias (por exemplo, as células de Küpffer no fígado). Persistem por mais tempo na inflamação e ainda podem sofrer divisão nos tecidos [8]. Por esses motivos, os macrófagos são as principais células encontradas nos estágios tardios das inflamações e possuem papel central na resposta imune natural. Além disso, também são fundamentais na resposta adquirida.

(iii) Células dendríticas: Também derivadas da medula óssea, as células dendríticas expressam receptores de reconhecimento de padrões e respondem aos microrganismos com a secreção de citocinas, existindo inclusive uma linhagem especializada para a resposta contra vírus. Além disso, apresentam projeções em sua superfície e capacidade fagocitária. São amplamente distribuídas pelo organismo e são fundamentais na conexão

entre os sistemas natural e adquirido, sendo que no segundo funcionam como células apresentadoras de antígenos aos linfócitos T [9].

(iv) **Células NK:** derivadas da medula óssea, são definidas como células citotóxicas não específicas e estão relacionadas aos linfócitos. São capazes de reconhecer células infectadas e células estressadas e respondem a partir da secreção de citocinas e também são ativadoras de macrófagos [1]. Apesar do mecanismo de ação citotóxica semelhante, ao contrário dos CTLs, as NK não são restritas ao MHC e não requerem sensibilização prévia, por isso não são capazes de formar memória imunológica. A atividade das NKs é regulada a partir da ativação e inibição dos receptores tipo imunoglobulinas e dos receptores semelhantes à lectina em sua superfície e isso se dá porque os ligantes desses receptores são as moléculas alteradas durante um processo de infecção ou lesão. Os ligantes ativadores normalmente incluem proteínas virais e induzidas por lesões e a inibição envolve o MHC de classe I, uma vez que células infectadas tem uma baixa na expressão dessas moléculas [10].

- **Sistema complemento (SC) e proteínas circulantes:** Formado por um conjunto de proteínas plasmáticas e de superfície de membrana, o SC é ativado por microrganismos e promove sua destruição, além do processo inflamatório. Sua ativação pode acontecer de duas formas: as vias clássica e alternativa e apesar de cada uma delas ser desencadeada por fatores diferentes, ou seja, o início da ativação é diferente para cada uma, ambas convergem em uma via comum cuja finalidade é a lise do microrganismo [11]. Nos mamíferos, o SC tem um papel importante nos mecanismos de defesa inato e adquirido [1]. O sistema inato ainda apresenta o seu equivalente de resposta humoral composto por outras moléculas reconhecedoras de padrão solúveis e moléculas efetoras. Aqui incluem-se as paraxantinas, ficolinas e colectinas [2].
- **Citocinas:** As citocinas são moléculas secretadas pelo sistema imunológico e que funcionam como mediadoras desse sistema. Sua produção é rápida e sua ação pode ser local ou sistêmica funcionando sempre a partir da ligação com receptores específicos e resultando na expressão de genes relacionados à resposta de interesse [12]. Quando reguladoras da resposta inata, as citocinas trabalham para recrutar leucócitos, aumentar a capacidade de síntese das células

efetoras e estimular o aumento da síntese de proteínas relacionadas à resposta antimicrobiana [4]. Tendo como principal fonte os macrófagos, neutrófilos e NK, no sistema inato as células epiteliais também podem funcionar como fornecedoras de citocinas [1] e dentre elas podem ser citadas as citocinas de acordo com a tabela abaixo:

Tabela 1: Citocinas do sistema imunológico natural.

Citocina	Função
TNF, IL-1, quimiocinas	Mediadores da inflamação
IL-2, IL-15	Estimular proliferação e atividade de NK
IL-6	Aumento da produção de neutrófilos na medula óssea e aumento da síntese de proteínas envolvidas na defesa do hospedeiro.
IL-10	Limitar a ativação de macrófagos
IFN α , IFN β	Controle de infecções virais
IFN γ (derivado de NK)	Ativação de macrófagos

Fontes: 13, 14

2.2 Resposta imunológica adquirida

Como já foi visto, a resposta adquirida é o sistema composto por linfócitos que, a partir do contato com um antígeno, deflagra respostas altamente específicas e que se tornam mais amplas e eficazes com as sucessivas exposições. Existem dois tipos de resposta imune adquirida, cada qual com seus componentes.

- **Imunidade humoral:** é mediada pelos anticorpos, proteínas que estão presentes nas mucosas e no soro. Os anticorpos são produzidos pelas células B e reconhecem e neutralizam antígenos, além de marcá-los para sejam reconhecidos e posteriormente eliminados. Esta é a principal linha de defesa contra microorganismos e toxinas, uma vez que há diferentes tipos de anticorpos capazes de induzir diferentes tipos de respostas [4].
- **Imunidade celular:** mediada pelos linfócitos T, essa respostas esta relacionada ao combate de microorganismos intracelulares através da destruição destes a partir da ativação dos macrófagos pelos linfócitos T auxiliares estimulando a destruição daquilo que foi fagocitado ou através

da destruição direta da célula do hospedeiro infectada a partir da atuação dos linfócitos T citotóxicos [1].

As respostas imunológicas adquiridas apresentam uma série de características fundamentais para o funcionamento do sistema imune, comuns tanto às respostas humoral como celular. Estas propriedades refletem as características dos linfócitos, principais componentes desse sistema [1]. Essas respostas são altamente específicas, podendo distinguir até mesmo diferentes porções de uma mesma macromolécula - os epítopos. Esse alto grau de especificidade acontece em função dos receptores membranares presentes na superfície dos linfócitos e mesmo indivíduos que ainda não tiveram contato com a antígenos apresentam uma grande variedade de receptores capazes de responder à diferentes antígenos, evidenciando a diversidade desse sistema [2].

Outra característica essencial é a expansão clonal que consiste no aumento do número de linfócitos antígeno-específicos quando há infecção, porém, esse sistema também é homeostático, ou seja, quando a infecção é finalmente eliminada, há uma automática redução no número dessas células de defesa. Contudo, mesmo com a contração no número de células, há formação de memória imunológica, ou seja, quando houver uma nova infecção pelo mesmo antígeno, a resposta será mais rápida e intensificada. Por fim, o sistema imune é capaz de evitar lesões no hospedeiro durante as respostas, isto é, há tolerância a antígenos próprios. quando há falhas desse aspecto, resulta-se em disfunções chamadas de doenças auto imunes [15]. O conjunto desses fatores contribui para a formação de um sistema altamente especializado.

2.2.1 Células e tecidos do sistema imunológico adquirido

Células tronco pluripotentes presentes na medula óssea dão origem às células progenitoras mielóides e linfóides e esses progenitores linfóides dão origem aos linfócitos T (LT), linfócitos B (LB) e células *natural killer* (NK) [1], principais componentes do sistema adquirido. As células que vão se diferenciar em LT deixam a medula óssea e migram para o timo, onde termina o processo de maturação. Apenas os LT maduros deixam o timo. Já as células que vão se diferenciar em LB passam por ciclos de proliferação ao mesmo tempo em que as cadeias de seu principal receptor (BCR) são expressas e permanecem na medula até o final de sua maturação [15], após a qual migram para os órgãos linfóides. Quando há falhas no processo, essas células são previamente eliminadas.

Os LT fazem o reconhecimento dos antígenos estruturalmente relacionados aos seus receptores de superfície celular (este mecanismo será descrito posteriormente). Eles reconhecem porções dos antígenos que foram previamente processadas pelas células apresentadoras de antígenos (APC) e estão expostos para reconhecimento ligados ao complexo principal de histocompatibilidade (MHC) que se encontra na superfície de outras células [12]. Há dois conjuntos principais de LT [5]. Dentre elas, três subconjuntos principais que expressam um receptor de antígeno $\alpha\beta$ [16]:

- Linfócitos T auxiliares ou *helper* CD4⁺ (Th): secretores de citocinas, cujas funções são variadas, dentre elas estimular a diferenciação e proliferação celular, assim como ativação e proliferação de outras células, como os LB na imunidade humoral e macrófagos na imunidade celular. O receptor CD4 (T-cell surface glycoprotein CD4) é uma molécula acessória para o reconhecimento de MHC classe II e regula a ativação do LT.
- Linfócitos T citotóxicos CD8⁺ (CTL): destroem células infectadas e células tumorais através da resposta citotóxica. O receptor CD8 (T-cell surface

glycoprotein CD8) é composto por duas cadeias e, como um análogo de CD4, participa no processo de reconhecimento do MHC classe I.

- Linfócitos T reguladores (TReg): inibição da resposta imune através da supressão da função das demais células.

Ainda há o grupo de células que expressam o receptor $\gamma\delta$ e esses LT apresentam funções auxiliares e citotóxicas na imunidade natural.

Os LB, por sua vez, são as únicas células capazes de produzir anticorpos a partir do reconhecimento de antígenos extracelulares. São mediadores da resposta imunológica humoral, como descrito no item 2.2. As principais classes de LB são os LB foliculares, os LB da zona marginal e os LB B-1 e podem ser encontrados em diferentes distribuições e tecidos [15].

3. Mecanismos associados aos linfócitos T

A ativação e proliferação dos linfócitos T são os eventos centrais das respostas imunológicas. A seguir serão tratadas a biologia e bioquímica envolvidas nesses processos.

3.1 Reconhecimento de antígenos - MHC

Como já mencionado, os LT são componentes imprescindíveis quando se trata do reconhecimento do Ag e, portanto, tem papel central na imunidade adquirida. A eliminação do Ag geralmente requer a participação de várias células efetoras, entretanto para que haja reconhecimento, os mesmos devem ser apresentados por um grupo de células específico, as APCs e essa função é desempenhada pelo grupo de proteínas do MHC. Os receptores de Ag das LTs reconhecem tanto o peptídeo antigênico quanto as moléculas do MHC.

Os mais de 200 genes extremamente polimórficos do MHC estão localizados no braço curto do cromossomo 6. Esse polimorfismo é importante na determinação da especificidade de ligação e quando não envolvido na ligação direta com o Ag, formam um determinante antigênico reconhecido pelos LT. Os genes do MHC estão presentes em todos os vertebrados, sendo que na espécie *Homo sapiens* eles recebem a designação HLA (Human *Leukocyte Antigens*). Há quatro classes de HLA [17]:

- Genes da classe I: HLA-B, HLA-C e HLA-A, largamente mais expressos.
- Genes da Classe II: HLA-DP, HLA-DQ e HLA-DR, de menor polimorfismo.
- Genes de proteínas *heat-shock* e citocinas
- Genes do complemento

Existem dois tipos de genes polimórficos do MHC que codificam proteínas estruturalmente distintas, porém homólogas: as moléculas de classe I (MHCI e as moléculas de classe II (MHCII). As principais características dessas duas classes estão listadas na tabela a seguir:

Tabela 2: Características do MHC classes I e II.

Característica	MHC Classe I	MHC Classe II
Tipos celulares	Todas as células nucleadas.	LB, LT ativados, APCs e células epiteliais durante processo inflamatório.
Interação com LT	CTLs CD8+	Th CD4+
Estrutura	Resíduos polimórficos encontrados nos domínios $\alpha 1$ e $\alpha 2$, ligação com CD8 feita pela região $\alpha 3$ e acomodação de 8-11 resíduos de aminoácidos pela fenda de ligação de peptídeo.	Resíduos polimórficos encontrados nos domínios $\alpha 1$ e $\beta 1$, ligação com CD4 feita pela região $\beta 2$ e acomodação de 10-30 resíduos de aminoácidos pela fenda de ligação de peptídeo, ou mais.
Nomeclatura em <i>H. sapiens</i>	HLA-A, HLA-B, HLA-C	HLA-DR, HLA-DQ, HLA-DP
Citocinas estimuladoras	IFN α , IFN β , IFN γ em infecções virais e TNF, linfotóxica em infecções microbianas	O IFN γ é o principal estimulador.
Tipos de Antígenos	Citosólicos, geralmente sintetizados endogenamente.	Extracelulares que tenham sido endocitados em vesículas.

Fontes: 17-19

A expressão das moléculas do MHC é determinante para o reconhecimento de Ags estranhos pelas LTs e sua regulação se dá não somente em nível transcricional, mas também pela ação de citocinas (Tabela 2). O MHCI é expresso em quase todo tipo celular nucleado e isto está relacionado à função efetora das LT CD8+, restritas a essa classe. Uma vez que o objetivo dessas células é eliminar principalmente células infectadas por vírus e dado que os vírus podem infectar qualquer tipo celular, é importante que o máximo de células nucleadas expresse MHCI. De outro modo, os LT CD4+ reconhecem Ags apresentados por um número limitado de células. Como a função principal desses LT é ativar macrófagos e LB, o MHCII é principalmente expresso nesses tipos celulares e fornecem um sistema relacionado a apresentação de peptídeos extracelulares e microbianos [19].

3.2 Processamento e apresentação de antígenos

Para que haja reconhecimento dos Ags pelos LT é necessário que essas moléculas sejam convertidas em peptídeos que se associam às moléculas de MHC, formando complexo peptídeo-MCH presente nas APCs. Os LT podem diferenciar pequenas variações, como apenas um resíduo de aminoácido de diferença e esses peptídeos devem que necessariamente estar ligados ao MHC - restrição ao MHC próprio - sendo que os LTh CD4+ reconhecem os peptídeos ligados ao MHC de classe II, enquanto os CTLs CD8+ reconhecem aqueles ligados ao MHC de classe I. A função do MHC nesse processo indica que os Ags compartilham certas propriedades que os diferenciam daqueles reconhecidos pelos LB, que podem ser tanto peptídicos como polissacarídeos, lipídicos e outras constituídos por substâncias [20].

A participação das APCs é fundamental para que haja respostas dos LT. Elas funcionam não só como processadores dos antígenos em peptídeos, como são co-estimuladores celulares. As APCs podem ser de vários tipos, sendo principalmente células dendríticas, macrófagos e LBs cujas funções principais são iniciar a resposta das LT a Ags protéicos, participar da fase efetora da resposta imune celular e efetuar a apresentação de Ags aos LTh CD4+ em respostas imunes humorais, respectivamente [19].

O processamento dos Ags é feito por organelas e enzimas não exclusivas do processo, ou seja, essas vias evoluíram como adaptações de funções celulares básicas, como reciclagem e degradação de peptídeos, e a associação às moléculas do MHC ocorre antes da exposição na superfície celular. No MHC de classe I, as enzimas responsáveis pela geração de peptídeos são as proteases endossômicas e lisossômicas e o carregamento ao MHC ocorre em um compartimento vesicular no citosol. Em contrapartida, para o MHC de classe II, as enzimas são aquelas encontradas no proteassoma citosólico, e o carregamento do peptídeo ao MHC se faz no retículo endoplasmático [21].

Os dois tipos de MHC constantemente exibem proteínas próprias, além das proteínas estranhas que são relativamente raras. Cabe aos LT buscar por essas proteínas estranhas e não reagir aos Ags próprios - os LT reativos a Ags próprios são eliminados ainda na fase de maturação no timo. Essa especificidade de reconhecimento característico das LTs é essencial para a resposta imunológica.

3.3 Receptor TCR e co-receptores

As respostas dos LT são iniciadas pelo reconhecimento do complexo peptídeo-MHC e para que essas respostas aconteçam são necessários aspectos determinantes,

como especificidade no reconhecimento do Ag, estabilidade da adesão entre as células e liberação de sinais que modularão a resposta imunológica. O receptor que reconhece esse complexo é chamado de receptor de células T (TCR) sendo eles distintos entre si para que haja especificidade de reconhecimento dos diversos peptídeos antigênicos.

Juntamente do TCR, há um grupo de receptores que efetuam a transdução do sinal a partir do reconhecimento do peptídeo antigênico - CD3 e ζ - formando o Complexo TCR, como mostrado na Figura 1. Além disso, há um conjunto de co-receptores que participam da resposta ao Ag como liberadores de sinais ou contribuem para a estabilização da ligação TCR-complexo peptídeo-MHC. Os domínios do TCR específicos para o complexo peptídeo-MHC são regiões variáveis localizadas na porção superior (α e β) semelhantes às cadeias variáveis das imunoglobulinas [22]. As proteínas CD3 e ζ estão associadas de forma não covalente ao heterodímero α e β e elas efetuam a transdução de sinal necessária para a ativação do LT quando há reconhecimento do antígeno [23]. Para que haja pleno funcionamento do receptor TCR é preciso que todos os seus componentes sejam sintetizados. Nos LT maduros, o complexo é montado no retículo endoplasmático e transportado para a membrana plasmática.

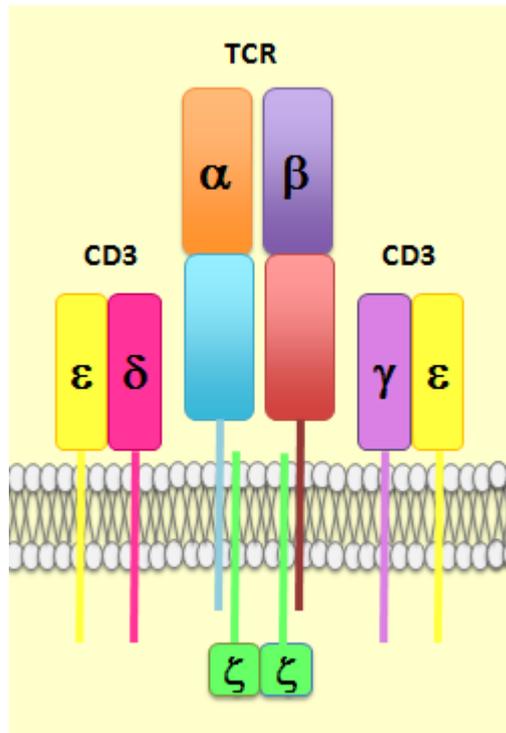


Figura 1: Complexo TCR. O TCR consiste de um heterodímero contendo duas cadeias α e β transmembrana ligadas por pontes dissulfeto e o CD3 é composto pelas proteínas γ , δ e ϵ e são homólogos entre si. O homodímero $\zeta\zeta$ esta covalentemente ligado por pontes dissulfeto.

Os LT também expressam receptores CD4 ou CD8 cuja função principal é efetuar a transdução de sinal quando há reconhecimento do Ag, além de interagirem com as regiões não polimórficas do MHC de classes II e I, respectivamente. Ambos são glicoproteínas, porém o CD4 é um monômero e o CD8 consiste de duas subunidades ligadas por pontes dissulfeto (CD8 como heterodímeros CD8 α e CD8 β) [24].

Além com complexo TCR e dos co-receptores CD4 e CD8, há outros co-receptores e moléculas auxiliares na superfície dos LT que participam na estimulação, ativação, diferenciação e proliferação dessas células.

4. Mecanismos efetores da resposta imune

Uma vez que o processo de reconhecimento do complexo peptídeo-MHC foi finalizado e a célula efetora foi ativada, os chamados mecanismos efetores da resposta imune são necessários para finalizar a resposta imunológica.

4.1 Imunidade mediada por células

A imunidade mediada por células é um mecanismo do sistema imune adquirido regulado pelos LT CD4+ e CD8+, sendo que cada qual, como já descrito anteriormente, tem funções específicas.

Os LT CD4+ efetores, ou Th1, são responsáveis pelo reconhecimento e eliminação dos Ags microbianos. Essas células ativam os fagócitos, principais células de defesa da imunidade inata, para digerir Ags previamente ingeridos. A cooperação entre os sistemas inato e adquirido faz-se necessária, pois inúmeros microrganismos desenvolveram mecanismos capazes de evadir a resposta natural e se multiplicam dentro dos fagócitos [25]. Entretanto, a verdadeira função efetora, isto é, a morte dos microrganismos, é mediada pelos próprios fagócitos. A sequência de eventos que descreve a resposta imunológica mediada por células está descrita na Figura 2.

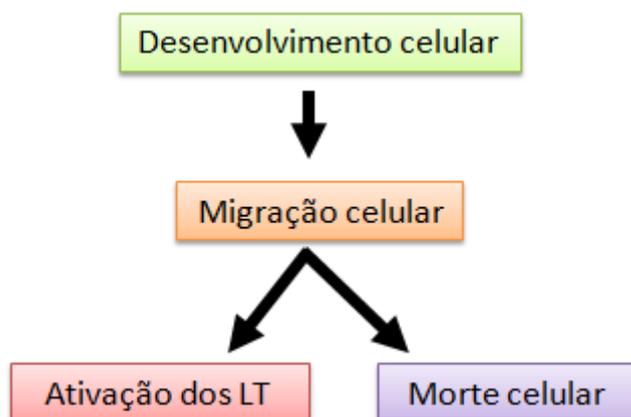


Figura 2: Respostas imunológicas mediadas por células. Inicialmente há o desenvolvimento dos LT nos órgãos linfóides periféricos. Então, segue-se a migração dessas células para os locais de infecção onde os LT podem ser ativados para a eliminação do agente infeccioso, a partir da mediação de citocinas, ou podem atuar na mediação da morte da célula infectada [1].

As Th1 são secretoras de $IFN\gamma$, citocina que favorece a fagocitose e efetua a mediação da ativação dos macrófagos por essas células. Os macrófagos ativados eliminam os microorganismos fagocitados e estimulam a inflamação. Ainda há outro subconjunto de LT CD4+, os Th2, responsáveis pela eliminação de parasitas helmínticos. Estas células são secretoras de IL-4 e IL-5, citocinas que estimulam a produção de IgE e as respostas associadas aos mastócitos e eosinófilos, tipos celulares associados à alergias e eliminação helmintos [26].

Os LT CD8+, ou CTLs, são responsáveis pela eliminação de células que expressam peptídeos de organismos intracelulares associados ao MHC de classe I (Item 3.1). O desenvolvimento da resposta mediado por estas células está descrito na Figura 3.

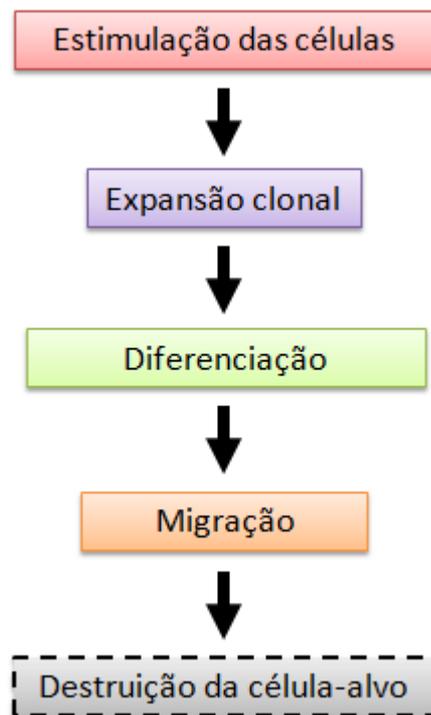


Figura 3: Desenvolvimento da resposta mediada por CTLs. As células são primariamente estimuladas pelo reconhecimento do antígeno ligado ao MHC de classe I nos linfonodos. Então há proliferação dessas células que se diferenciam em células efectoras que migram, já diferenciadas, para os tecidos onde há destruição das células infectadas e liberação dos CTLs [1].

As respostas imunes mediadas por LTs resultam na formação de memória imunológica, a qual pode persistir durante muito tempo pela ação de IL-7 e IL-15. Essas células se instalam em mucosas e respondem aos antígenos [27] futuramente com a secreção de citocinas.

4.1.1 Aspectos da imunidade anti-viral

Com o objetivo de impedir a replicação, infecção e disseminação viral, as células do sistema imune desenvolveram um sistema de defesa que envolve tanto o sistema inato quanto o adaptativo. Uma vez que esses organismos invadem as células do hospedeiro são desencadeados mecanismos que envolvem a indução de citocinas denominadas interferons [28].

Como mencionado no item 2.2.1, os IFNs são um grupo de proteínas heterogêneas que capazes de induzir resistência celular mediante uma infecção viral. Além disso, os IFNs estão envolvidos em outros processos, como diferenciação e crescimento celular.

Essas moléculas são classificadas em três grupos [28]:

- **Tipo I:** esse grupo compreende os IFN α , IFN β , IFN ϵ , IFN κ e IFN ω . Todos eles apresentam homologias em suas estruturas e se ligam num receptor comum chamado de receptor de interferons de tipo I. Este receptor é composto pelas subunidades IFNAR1 e IFNAR2 [29].
- **Tipo II:** Há apenas um tipo, o IFN γ , e ele se liga a um receptor - Receptor de interferon tipo II - também formado por duas subunidades, IFNGR1 e IFNGR2. Esse é um tipo diferenciado de IFN, mas foi originalmente agrupado nesse grupo de citocinas devido à sua capacidade de interferir nas infecções virais [29].
- **Tipo III:** São três tipos diferentes de IFN codificados por diferentes genes, sendo eles IL-28A, IL-28B e IL-29. Possuem atividades antivirais, embora distintas dos tipo I e II. Essas moléculas se ligam em receptores

heterodiméricos formados pelas cadeias, IL28RA e IL10RB, sendo a última uma das subunidades do receptor da família IL-10. Embora o tipo de resposta desencadeada por esses IFNs seja semelhante àquela induzida por $INF\alpha$, a expressão desses genes é restrita a tecidos de origem epitelial [30].

A infecção viral induz a produção de IFNs tipo I e após sua interação com o receptor acontece a indução de um conjunto de genes que inibem a replicação do vírus e aumentam o potencial de ação das NK [29]. Os receptores desses IFNs estão associados a uma transdução de sinal do tipo JAK/STAT. Essa resposta chega ao núcleo onde é iniciada a transcrição a partir da ligação em sítios específicos nos promotores, os ISGs (IFN-Stimulated Genes). Dentro desses promotores também está presente a IFN-Gamma Activated Site (GAS), sendo este complexo induzido por $IFN\gamma$ [31]. Além disso, acontece a modulação da resposta adaptativa a partir do aumento da expressão de MHC classe I, aumento da sobrevivência dos LT e maturação das células dendríticas.

4.2 Imunidade humoral: ativação das células B e produção de anticorpos

Como mostrado no item 2.2, a imunidade humoral é mediada por anticorpos, moléculas responsáveis pela defesa do organismo contra Ags solúveis extracelulares, microrganismos extracelulares e toxinas. Essas moléculas se ligam aos Ags e agem neutralizando e eliminando-os. Os tipos e quantidades de anticorpos podem variar de acordo com inúmeros fatores, como o tipo de Ag e se já houve exposição prévia.

A resposta humoral se inicia com o reconhecimento antigênico pelos LB específicos, a partir de sua ligação com receptores de Ig na superfície de membrana. Essa ativação pode resultar na expansão clonal desses LB ou sua diferenciação em plasmócitos secretores de anticorpos, também conhecidos como imunoglobulinas (Igs).

A ativação dos LB se inicia com o agrupamento de IgM e IgD de membrana que reconhecem os Ags em conjunto com as moléculas Ig α e Ig β . Esse conjunto de moléculas, conforme ilustrado na Figura 4, formam o complexo do receptor antigênico de linfócitos B (BCR) [32]. Esse receptor atua na endocitose do Ag e na transdução de sinal e os primeiros eventos celulares induzidos pela ligação do Ag são a proliferação e diferenciação dos LB [33].

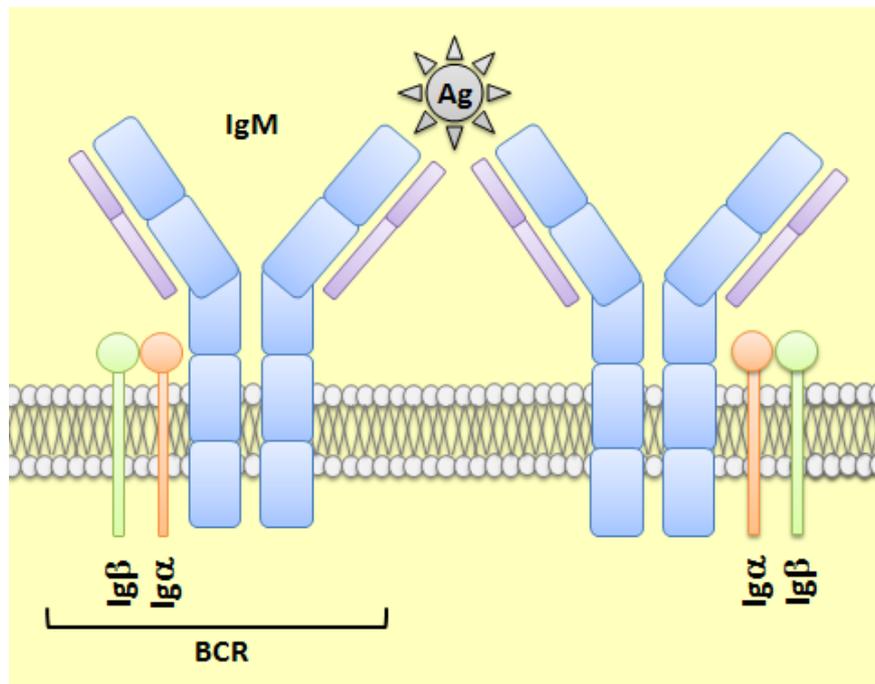


Figura 4: Complexo do receptor de linfócito B.

A ativação dos LB pode acontecer de forma dependente de LTs ou de forma independente. No primeiro caso, as respostas de Igs a Ags protéicos estão associadas aos LT CD4+ que trabalham no reconhecimento do Ag e na ativação dos LBs. Os Ags são reconhecidos nos órgãos linfóides periféricos onde são capturados por APCs, então são apresentados aos LTh e os LB são ativados a partir dessa interação com os LT [34]. No segundo caso, por sua vez, acontece na presença de Ags não protéicos e a ativação do LB se dá pelo contato do Ag com Igs de membrana.

As Igs compõe um grupo de proteínas produzidas pelos LB que podem tanto estar ligadas à membrana celular onde funcionam como receptores como podem também estar livres no soro, sendo esta a forma que atua como mediadora da imunidade humoral [12]. Essas moléculas são estruturalmente semelhantes, contendo duas cadeias pesadas e duas cadeias leves ligadas covalentemente entre si, entretanto as regiões de ligação de Ag são altamente variáveis e como uma especificidade diferente.

As Igs são classificadas em diferentes subtipos, cada qual com propriedades diferentes desencadeadas a partir da ligação com o Ag. As principais classes são: IgA, IgD, IgE, IgG e IgM [35]. Além disso, as Igs podem se ligar em até dois Ags, com exceção da IgM que pode se ligar em até 10. Essas ligações são altamente específicas e podem ocorrer modificações na estrutura das Igs durante a resposta imune [36].

Uma vez produzidas na medula óssea e nos órgãos linfoides, as Igs são secretadas e entram na circulação. Então, são transportados para o lúmen de órgãos onde exercem suas inúmeras funções, sendo a principal delas a neutralização de microrganismos patogênicos e toxinas.

5. Aspectos evolutivos do sistema imune

A história da vida é certamente moldada pelos crescentes riscos imunológicos associados a faunas parasitárias em seus ambientes e a competência imunológica pode ser determinante no sucesso reprodutivo e *fitness* de várias espécies [37]. Determinar a importância relativa dos elementos do sistema imune é fundamental para estudos comparativos que podem apontar elementos chave e sua origem filogenética. Contudo, essas análises comparativas se tornam complicadas devido aos seguintes fatores: (i) coevolução do sistema imune com outros sistemas do organismo, ou

seja, a evolução desse sistema não deve ser considerada em separado, (ii) visão incompleta da filogenia dos metazoários, e (iii) a multiplicidade do sistema imune faz com que este seja uma unidade de evolução complexa e plástica, onde alguns elementos podem evoluir em poucas gerações [38].

Os metazoários, possuem mecanismos de reconhecimento entre *self* e *non-self* e são capazes de se proteger da invasão de microrganismos, parasitas e células exógenas, porém, os mecanismos utilizados não são necessariamente homólogos dentre os filos [39]. Em poríferos, por exemplo, há atividade fagocitária exercida pelos arqueócitos, como um mecanismo para eliminar as partículas *non-self* e, de forma mais complexa, há um mecanismo de encapsulamento do estranho dentro de um agregado celular para uma posterior eliminação [40].

A resposta imune adaptativa surgiu com os ancestrais dos vertebrados mandibulados (Gnathostomata) há aproximadamente 500 milhões de anos. Homólogos de Igs, TCR, MHC classes I e II já foram identificados em todos os mandibulados existentes, entretanto, nenhum desses genes foi identificado em vertebrados não-mandibulados (agnatos) e invertebrados [41]. Os Gnathostomata geraram um repertório diverso de receptores de antígeno tanto nos LT quanto nos LB através da recombinação gênica (recombinação VDJ), enquanto peixes agnatos montaram seus receptores de linfócitos variáveis através da recombinação de regiões ricas em leucinas, como acontece com os TLRs. Ainda que plantas, deuterostomados e cordados invertebrados apresentem grande variação desses receptores de reconhecimento, apenas alguns poucos são necessários para a sobrevivência de nematóides, insetos e vertebrados [42].

Muitas famílias de diferentes genes de receptor com função predita foram identificadas em grupos além dos Gnathostomata. Entretanto, a identidade do receptor

"primordial" foi interrompida pelos diversos mecanismos de recombinação sofridos por esses genes nesse táxon [41]. O aparecimento desse sistema de recombinação dos receptores de linfócitos em vertebrados pode ter sido impulsionado por uma necessidade de facilitar a plasticidade morfológica juntamente com a possibilidade de reconhecimento de uma gama maior de antígenos [42].

Inúmeras abordagens, incluindo análises de genoma e proteoma, estão sendo aplicadas a agnatos, protocordados e outros vertebrados deuterostomados. Esses estudos podem produzir as provas definitivas quanto a presença ou ausência de homólogos da imunidade adaptativa em espécies onde não há sistema imune adaptativo [41].

6. Homologias e conceitos

O sequenciamento de genomas resulta numa representação linear a qual sucintamente resume a estrutura da molécula sequenciada [43] e é fundamental na tentativa de elucidar a cadeia de eventos contínuos ocorridos durante o processo evolutivo. Assim como elementos genômicos estruturais e funcionais, os genes apresentam similaridades em suas sequências e estas são comumente referidas como homologias. Elas podem ocorrer em grau variado e decorrem de processos evolutivos os quais deram origem às sequências envolvidas.

Há vários eventos que determinam o surgimento de sequências homólogas, como especiação, duplicação seguida de divergência, transferência horizontal e mutações. Além disso, assume-se que sequências homólogas tem uma origem evolutiva em comum e elas podem ainda desempenhar a mesma função biológica, por exemplo, quando a

similaridade das cadeias de aminoácidos é alta ela é tida como forte indicativo da manutenção da função.

Por esse motivo, técnicas de agrupamento que correlacionam proteínas de função desconhecida a outras são valiosas e muito utilizadas para inferência e propagação da informação biológica, bem como para estudos de evolução gênica e da genômica comparativa [12].

De acordo com Altenhoff e Dessimoz, a homologia é a relação entre um par de genes que compartilham um ancestral comum e estes podem ser diferenciados em duas classes [45]:

- **Parálogos:** são pares de genes que divergiram a partir da duplicação de um gene ancestral e que ainda podem ser subdivididos em:

- (i) In-Parálogos: envolve um par de genes e um evento de especiação como referência. Ocorrem quando há duplicação após a especiação, ou seja, a duplicação está restrita a uma linhagem.

- (ii) Out-Parálogos: este caso se dá a partir de um evento de duplicação que antecede a especiação. A especiação aumenta a divergência entre essas cópias, que acabam por constituir famílias gênicas distintas.

- **Ortólogos:** referidos como pares de genes que divergiram a partir de um evento de especiação biológica e, uma vez que surgem por especiação, qualquer conjunto desses genes tem a mesma história evolutiva. A Figura 5 representa esquematicamente a organização espacial e temporal desses genes ao longo de um cladograma. Cada espécie está representada por uma caixa colorida e os genes pelas letras dentro das caixas. O último ancestral comum está representado por uma estrela e ele inicialmente passa por um processo de especiação (Esp1) estando presente na espécie azul e na laranja (nesta, em cópia única, o gene "E"). Então, acontece um evento de duplicação (Dup) do gene (são as cópias "A" e "C"). Nesse contexto, assumindo-se que A e B ou A e E, por exemplo, mantiveram a

mesma composição gênica, eles são ditos ortólogos. Já os pares A e C são exemplos de genes parálogos, originados por duplicação na espécie A. Nota-se ainda que o par A e B são originados por duplicação após Esp1, então esse par é um caso de in-parálogos. Um exemplo de out-parálogos seria o par B e D que sofreram duplicação antes da especiação Esp2. Diferentemente da linhagem A, onde o gene C está mantido por ter fixado sua nova função na espécie azul, se muitas especiações como Esp2 ocorrem, além da que gera a espécie amarela, a tendência será dos genes B e D divergirem a ponto de formarem genes distintos, ainda que participantes de uma mesma família gênica. Essa maior divergência por especiação de genes previamente duplicados os definiria como out-parálogos. Um grande desafio para um software é agrupar como homólogos os ortólogos e in-parálogos, abrindo uma nova entrada para o grupo de out-parálogos.

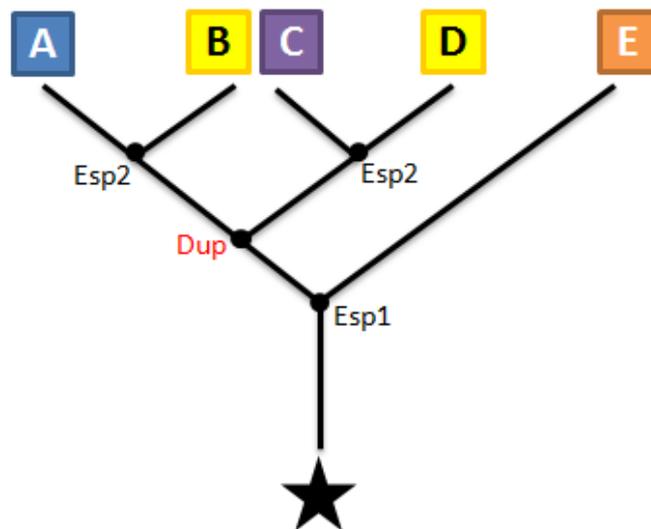


Figura 5: Genes homólogos e suas subdivisões. O ancestral está representado por uma estrela; diferentes espécies estão representadas por quadrados coloridos e genes por letras. Esp1: primeiro evento de especiação. Esp2: segundo evento de especiação. Dup: duplicação.

Ainda é importante ressaltar que a transferência horizontal de genes funciona como agente no processo evolutivo. Quando uma espécie adquire um gene de um outro organismo de forma independente de descendência, através da transferência horizontal, esses genes ortólogos, original e adquirido, são ditos genes xenólogos [44].

Nos dias de hoje, entretanto, a principal motivação para a distinção entre ortólogos e parálogos se faz para o estudo e predição da função gênica. É de comum acordo que os ortólogos são susceptíveis a ter função biológica similar. Por outro lado, acredita-se que os parálogos diferem muitas vezes em função, suavemente em caso de in-parálogos, pois sub-funcionalização ou neo-funcionalização são rapidamente fixadas, ou fortemente no caso de out-parálogos, pois especiações subseqüentes os expõem a mudanças funcionais grandes e estão comumente associados ao aparecimento de novos genes. Conseqüentemente, os ortólogos são de interesse não só para inferir (e propagar) computacionalmente a função gênica, como para traçar a história evolutiva dos mesmos [45].

Os parâmetros usados para que duas sequências sejam consideradas homólogas são arbitrários, mas geralmente, durante o agrupamento, visa-se a inclusão de genes ortólogos - aqueles que eram o mesmo gene no último ancestral comum das espécies envolvidas - e os in-parálogos. É desejada a exclusão dos out-parálogos, uma vez que eles provavelmente tiveram sua função alterada e podem pertencer à mesma família gênica, mas são já genes distintos [46].

6.1 Métodos de agrupamento de sequências homólogas

Dada a importância dos estudos em genômica comparativa e a utilidade de anotações funcionais e estudos evolutivos, diferentes metodologias para o estudo e agrupamento de sequências homólogas foram desenvolvidas. De um modo geral, esses métodos exigem a inclusão de apenas genomas completos.

A comparação de diversos genomas, par a par, extraíndo o melhor alinhamento recíproco (*Bidirectional Best Hit* - BBH) comparando-se todas as proteínas utilizando o

algoritmo BLAST [47] é uma forma simples e eficiente de inferir a homologia, quando um número pequeno de genomas é considerado, como no algoritmo InParanoid [48]. Ainda que se limite a pares de organismos, este é um algoritmo capaz de fazer a distinção entre *in*-parálogos e *out*-parálogos, dados alguns parâmetros, como não agrupar genes de um mesmo organismo se o *score* for inferior ao do alinhamento com um gene de outro organismo.

A comparação par a par também pode ser feita de forma que envolva mais de dois organismos. Isso é possível quando há formação de melhores cruzamentos triangulares [49], mas esses métodos são mais sensíveis a erro em caso de deleções gênicas e a proteínas com múltiplos domínios. Há também técnicas como as aplicadas na base de dados do OrthoMCL [50] que fazem uso de modelos de Markov para refinar o agrupamento final.

Outros métodos constroem árvores filogenéticas para distribuição de ortólogos e parálogos a partir de homólogos pré-computados. Entretanto, para que haja melhor detecção de especiações e duplicações gênicas, métodos que envolvem similaridade de sequências e árvores filogenéticas foram desenvolvidos. O Ensembl Compara [51], mesmo contendo poucos genomas eucariotos, é uma fonte de grupos de homólogos pouco influenciáveis por deleções, enquanto os grupos PHOGs [52] lidam bem com proteínas multi-domínios. No entanto, quando comparados entre si, métodos baseados em filogenia mostram-se menos eficientes do que os algoritmos de similaridade de sequência [53].

Com a crescente disponibilidade de sequenciamento de genomas, manter atualizações dos grandes bancos de dados de grupos de homólogos deve se tornar superior. Nesse sentido, o Laboratório de Biodados (UFMG) desenvolveu uma ferramenta

capaz de criar grupos de homólogos sob demanda a partir de sequências proteicas de interesse, com a grande vantagem de incluir informações provenientes de genomas incompletos [54].

7. Bancos de dados biológicos

Devido ao grande volume de informações fornecidas atualmente, os bancos de dados biológicos são fundamentais para disponibilizar e facilitar o acesso à informação. Eles representam uma das principais ferramentas de suporte para os estudiosos, principalmente aqueles relacionados a dados de sequenciamento de genes e proteínas. Para sua utilização, é importante que eles respeitem determinados critérios como frequência de atualizações, facilidade do acesso, qualidade e quantidade de informação disponível e confiabilidade dos dados.

Nos tópicos dessa seção serão abordados os bancos de dados utilizados como referência no presente trabalho.

7.1 *Universal protein resource (UniProt)*

O Uniprot é um repositório central de sequências de proteínas formado a partir da combinação dos bancos *Swiss-Prot*, TrEMBL e PIR-PSD. Ele apresenta uma interface simples que permite fácil acesso a informações funcionais e estruturais de proteínas.. Nele são encontradas informações de todos os organismos envolvidos em projetos de sequenciamento. Os dados podem ser encontrados nos formatos FASTA, GFF, RDF, XML.

UniProt é composto por diferentes componentes, sendo que cada um deles cabe à diferentes usos [55]:

- **Protein Knowledgebase (UniProtKB):** é o ponto de acesso central para o enorme volume de informações derivadas da anotação de proteínas (e suas referências), como função e classificação. Ele é composto pelo (i) UniProtKB/Swiss-Prot onde se encontram as proteínas anotadas manualmente e pelo (ii) UniProtKB/TrEMBL, constituído de proteínas anotadas automaticamente.
- **UniProt Archive (UniParc):** é um repositório abrangente não redundante que engloba dados de diversos bancos como o UniProtKB e o GenBank [56].
- **UniProt Reference Clusters (UniRef):** gerado com base nos dados dos dois bancos citados anteriormente, utiliza o algoritmo CD-HIT (Ref) para criar agrupamentos (*clusters*) de sequências de acordo com a porcentagem de identidade e de sobreposição das sequências. Está distribuído em três bancos: UniRef100, UniRef90 e UniRef50. O UniRef ainda permite que as coleções de sequências estejam sempre atualizadas e as redundâncias sejam escondidas.
- **UniProt Metagenomic and Environmental Sequences (UniMES):** exclusivo para dados de metagenômica, fornece também versões com agrupamento (UniMES cluster) [57].

7.2 Kyoto Encyclopedia of Genes and Genomes (KEGG)

Desenvolvido pelo *DNA Data Bank of Japan* [58], o KEGG vem sendo diariamente atualizado utilizando-se análises computacionais associadas à curadoria manual desde sua criação, em 1995 [59]. É um completo recurso que integra informações genômicas, químicas e sistêmicas, e consiste num conjunto de bancos de dados. Dentre eles é importante ressaltar:

- **KEGG PATHWAYS:** é uma coleção de diagramas que esquematizam vias de sinalização que incluem genes, proteínas, RNAs e demais compostos químicos,

bem como as relações desses componentes entre si dentro da via. Os grupos de vias são classificados em diferentes seções e por este banco ser limitado à interações moleculares, ele é incrementado pelo KEGG BRITE, um bando de dados de ontologia, onde há informações sobre classificações hierárquicas de genes, proteínas, organismos, etc [60].

- **KEGG GENES:** é uma coleção de catálogos de genes para os genomas completamente sequenciados. Para cada um desses genes é dada uma anotação onde se torna possível estabelecer correspondências para a construção de vias no KEGG PATHWAYS e hierarquias no KEGG BRITE, além de permitir conexões com os demais bancos de dados presentes no KEGG (Figura 6). Para a construção de tais correspondências é usado o conceito de genes ortólogos (seção x - genes homologos e conceitos).

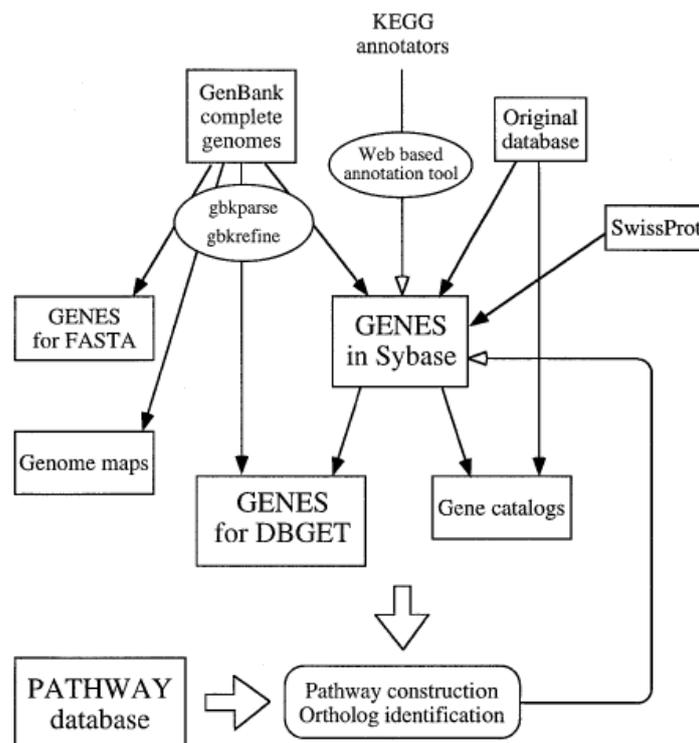


Figura 6: Processos usados para organizar e anotar o banco de dados KEGG Genes [59].

- **KEGG Orthology System (KO):** O KO é uma coleção de grupos de ortólogos definidos manualmente relacionados às vias do KEGG PATHWAYS e às ontologias do BRITE. Formado por milhares de grupos de homólogos construídos manual e automaticamente e provenientes dos genomas completos de archaeas, procariotos e eucariotos, cada um desses grupos representa as variantes de um

mesmo gene e são identificados por um conjunto de números precedido da letra K, por exemplo K04687 (Interferon Gama). Quando é uma enzima de função conhecida, há também a associação a um número - EC (*Enzyme commission*) [46]. Uma vez associados a genes, os grupos KO são utilizados como nós na construção de vias metabólicas PATHWAY.

7.3 NCBI *Taxonomy*

O banco de dados *Taxonomy* do NCBI existe para fornecer informações sobre nomenclatura e classificação biológica dos diversos organismos cujos genomas estão depositados em diferentes bancos de dados biológicos, assim como dados da literatura (atualizações são automaticamente incorporadas no banco) [61].

Quanto à classificação filogenética, há um esforço em manter apenas os grupos monofiléticos - aquele que inclui todas as espécies ancestrais de um ancestral [62] - de acordo com o consenso atual da sistemática biológica. Também é importante lembrar que os dados de classificação e nomenclatura utilizados são aqueles encontrados na primeira literatura publicada, como de praxe.

Os dados são organizados como uma árvore, sendo que cada nó representa um táxon e cada um desses táxons é identificado por um número identificador: o "taxonomy id" [63]. Associado à esse número segue-se uma série de outras informações como o "rank", onde é mostrada a classificação completa daquele organismo (Figura 7).

The screenshot shows the NCBI Taxonomy Browser interface. At the top, there are navigation tabs for Entrez, PubMed, Nucleotide, Protein, Genome, Structure, and PMC. Below these is a search bar with the text 'as' and a dropdown menu set to 'complete name'. There are 'Go' and 'Clear' buttons. Below the search bar, it says 'Display 3 levels using filter: none'. The main content area is titled 'Homo sapiens' and lists the following information: Taxonomy ID: 9606, Genbank common name: human, Inherited blast name: primates, Rank: species, Genetic code: Translation table 1 (Standard), Mitochondrial genetic code: Translation table 2 (Vertebrate Mitochondrial), Other names: common name: man, authority: Homo sapiens Linnaeus, 1758. Below this is a 'Lineage (full)' section with a list of taxonomic ranks: cellular organisms; Eukaryota; Opisthokonta; Metazoa; Eumetazoa; Bilateria; Deuterostomia; Chordata; Craniata; Vertebrata; Gnathostomata; Teleostomi; Euteleostomi; Sarcopterygii; Dipnotetrapodomorpha; Tetrapoda; Amniota; Mammalia; Theria; Eutheria; Boreoeutheria; Euarchontoglires; Primates; Haplorrhini; Simiiformes; Catarrhini; Hominoidea; Hominidae; Homininae; Homo.

Figura 7: Resultado da busca por *Homo sapiens* no website NCBI Taxonomy.

O Taxonomy contém dados de um número crescente de espécies disponibilizados em um servidor constantemente atualizado. Uma vez que essa informação é carregada em bancos locais, é possível realizar consultas para descobrir, por exemplo, os ancestrais diretos de um determinado táxon, como realizado no presente trabalho. Além disso, ainda há a função “Common tree” que permite a adição de diferentes organismos, a partir de seus “taxonomy IDs”, e estes são agrupados em uma árvore filogenética e a visualização gráfica é permitida.

7.4 Wikipathways

É um recurso aberto que contribui para a manutenção e atualização de conteúdo dedicado à vias biológicas. Uma vez registrado, o usuário pode contribuir sob monitoramento de um grupo de administradores [64]. Entretanto, a curadoria das vias e sua manutenção são de responsabilidade da comunidade de usuários.

O *WikiPathways* é formado por três componentes principais:

- **MediaWiki:** Desenvolvido pela Fundação WikiMedia, está escrito em PHP e usa um banco de dados back-end. Este é um componente usado em diversas *wiki* e esta disponível desde 2002 [65].

- PathVisio: É um programa para análise, edição e construção de vias biológicas.
- BridgeDb: uma plataforma que abrange dados de genes, proteínas e metabólitos [66].

O *WikiPathways* inclui muitos tipos de vias, sendo elas de metabolismo, regulação ou sinalização. Estão incluídas vias de animais, bactérias e plantas e com o uso de um recurso de busca, essas vias podem ser encontradas pelo nome, por seus componentes e/ou com base em categorias de ontologia. Cada via vem acompanhada de um texto de descrição, referências, dados da versão e uma lista dos componentes.

8. Ferramentas e serviços

Neste ítem serão listadas e detalhadas as ferramentas e serviços utilizados durante o desenvolvimento do presente trabalho.

8.1 SeedServer

Desenvolvido ao longo do doutorado de Rafael Guedes (Laboratório de Biodados/UFMG), o *SeedServer* é um programa de livre acesso construído em linguagem PERL que, em linhas gerais, consiste numa ferramenta de busca por homólogos a partir do carregamento de identificadores UniProt (*Seeds*) de interesse. Em seguida, carrega-se a taxonomia (reino ou filo) onde deverá ocorrer a busca e também são ajustados os parâmetros do *Seed Linkage* (SL), programa que executa as buscas por homólogos, e caso a proteína de interesse participe de um grupo KO enriquecido no laboratório (UEKO) é consultando este banco de dados e os demais membros do grupo são recrutados.

As sequências recrutadas são armazenadas numa tabela *MySQL*. Na coluna referente ao SL são mostrados os números 2 para a *Seed*, 1 para sequências recrutadas pelo SL e 0 para não recrutadas por este programa. Na coluna referente ao UEKO, o

número 1 se refere às sequências presentes no KO, 2 para aquelas que estão presentes na porção enriquecida do UEKO e 0 para ausentes. Quando a *Seed* não pertence a KO ou UEKO (número 0), seu UniRef50 é consultado e é exigido um alinhamento BLAST superior a 50% em relação ao tamanho da *Seed* para que haja recrutamento, metodologia denominada *UE-Seed*, já que é a mesma utilizada para enriquecer os grupos KO. Neste caso, a coluna relativa ao UEKO mostra o número 3 [46]. Nas demais colunas são encontradas informações sobre número de resíduos de aminoácidos, identificador taxonômico do organismo, se há curadoria manual *SwissProt*, o filo e a descrição proteica (Tabela 3).

O próximo passo consiste na validação das proteínas recrutadas pelo SL e UEKO feito através de buscas PSI-BLAST. Este procedimento foi desenvolvido ao longo da tese de Henrique Assis e, em linhas gerais, consiste num método onde é evitada a deterioração da matriz com um simples procedimento, evitar que o *auto-score* da própria sequência *Seed*, ao longo das iterações, cai abaixo de um determinado percentual, que foi calibrado para 70% como *default*.

Tabela 3: Tabela final com os resultados do *SeedServer* para Interferon gama humano.

UniProtID	Cluster	SL	KO	Tamanho	TaxID	Clado	Validação	PSI	Descrição
P01579	1	2	1	166	9606	1	1	1	IFNG_HUMAN Interferon gamma
G3QT96	1	1	0	166	9696	1	1	1	G3QT96_GORGO Interferon gamma
G1QTG5	1	1	2	166	61853	1	1	1	G1QTG5_NOMLE Interferon gamma
P63311	1	1	2	165	9545	1	1	1	IFNG_MACNE Interferon gamma
P63310	1	1	1	165	9544	1	1	1	IFNG_MACMU Interferon gamma

8.1.1 *Seed Linkage*

Desenvolvido por Barbosa-Silva e colaboradores, o *Seed Linkage* é um programa de livre acesso capaz de criar grupos de proteínas homólogas a partir de uma única

proteína (*query*). O programa faz buscas exaustivas de similaridade de sequência com uso de BBH e é também capaz de usar genomas incompletos em suas buscas.

O algoritmo inicia alinhando a sequência proteica de um organismo de interesse com outras sequências de diversos genomas presentes no banco de dados. As relações de BBH estabelecidas com outros genomas servem de parâmetro para as respectivas buscas de *in*-parálogos enquanto a melhor relação BBH obtida é utilizada para busca de *in*-parálogos no organismo de interesse [54].

8.1.2 UniRef50 Enriched KEGG Orthology (UEKO)

Desenvolvido por Gabriel Fernandes no Laboratório de Biodados (UFMG), o banco de dados UEKO é uma tabela local no MySQL que consiste no enriquecimento dos grupos de genes ortólogos KO a partir das sequências presentes no banco de dados UniRef50. Grupos UniRef50 ao qual cada proteína em um determinado KO pertence são consultados e o enriquecimento realizado seguindo filtros de qualidade descritos por Fernandes *et al.* [67], uma metodologia inicialmente utilizada para enriquecimento de agrupamentos proteicos do *Clusters of Orthologous Groups* (COG) [68].

O algoritmo que compõe o UEKO faz o recrutamento de sequências para um dado grupo KO a partir de alinhamentos BLAST entre as sequências originais do KO e sequências presentes no UniRef50 de cada uma das originais. Em seguida, há uma verificação do tamanho do alinhamento dividido pelo tamanho da sequência original (*coverage*) onde se exige um mínimo de 0.5. Note-se que o UEKO agrega todos os genes do grupo KO mais os genes de agrupamentos UniRef50 que satisfaçam o limiar de recrutamento das entradas do UniProt.

Esse enriquecimento agrega sequências derivadas de genomas incompletos aos grupos KO e aumenta o número de proteínas em aproximadamente 300%, aumentando assim a precisão na identificação da origem dos genes (Figura 8).

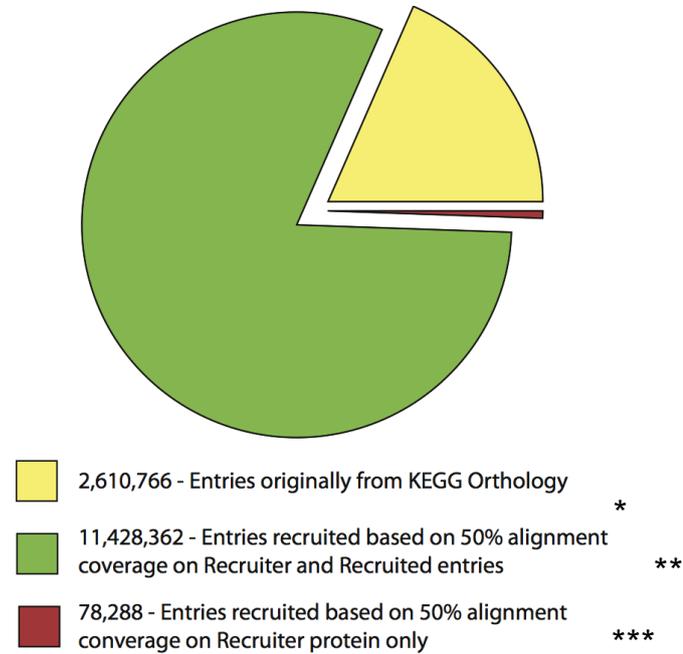


Figura 8: Enriquecimento do KO. Em amarelo, o número de sequências no bando de dados original e em verde, a contribuição do enriquecimento na construção do UEKO. [69]

8.2 Algoritmo de LCA

Desenvolvido ao longo do projeto de doutorado de Henrique Velloso no Laboratório de Biodados (UFMG), o algoritmo LCA mostra, em uma árvore filogenética, o ancestral comum mais recente de dois ou mais clados (Figura 9). Esse algoritmo funciona como um módulo que opera sobre árvores taxonômicas e retorna o táxon ancestral comum a uma lista de TaxIDs de interesse, usados como input, logo, esse conceito somente pode ser aplicado em árvores que apresentam raiz [63]. Os TaxIDs podem ser encontrados como descrito no item 7.3.

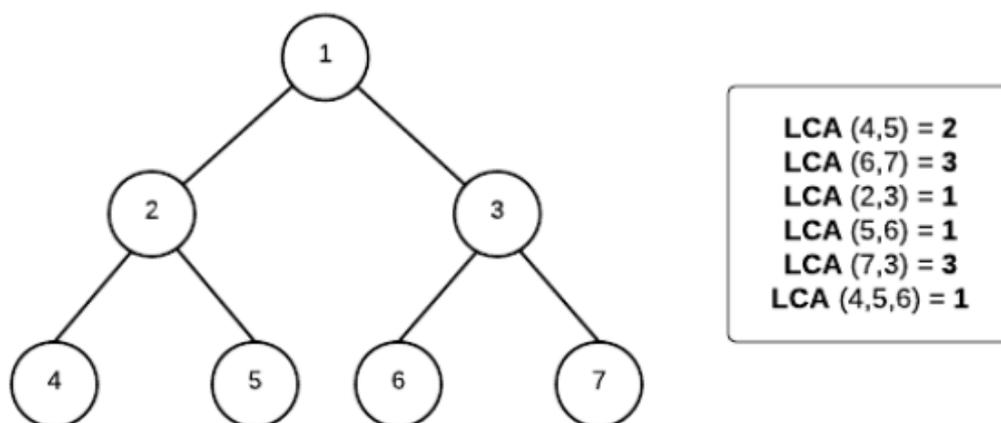


Figura 9: Exemplos de cálculo de LCA

A utilização do LCA esta disponível via linha de comando e por meio de *Web Service*. Para acesso via linha de comando, existe o script *LCARunner*, que lê um arquivo contendo uma lista de TaxIDs e escreve na saída padrão o tax_id e nome científico do LCA calculado [63]. Já o uso do LCA via *Web Service* pode ser encontrado em biodados.icb.ufmg.br/services/.

8.3 Genesis

O *website* Genesis possui três ferramentas para inferência da múltipla origem cladística de genes. A primeira, denominada *KO origin*, permite que o pesquisador explore as múltiplas origens de um gene específico. A segunda, chamada de *Species origin*, realiza a análise de ancestralidade gênica em um âmbito de espécies. A última ferramenta permite que o pesquisador realize a análise de origem múltipla de genes em um conjunto personalizado de organismos de entrada.

Este *website* foi construído com o *framework* Grails versão 2.3.2. Esse *framework* propicia grande aumento na produtividade por meio de vários recursos que evitam tarefas repetitivas de codificação. A linguagem de programação utilizada em Grails é Groovy [63].

8.4 MaLe-PSI-BLAST

Desenvolvido durante o doutorado de Henrique de Assis Lopes Ribeiro, no Laboratório de Biodados (UFMG), o MaLe-PSI-BLAST sugere anotação de proteínas hipotéticas. Para o presente trabalho, ele foi utilizado no sentido de encontrar grupos de proteínas relacionadas a uma *query* de interesse cuja função é similar.

Disponível em www.bows.com.br, o programa é simples e de fácil utilização. O mesmo deve ser preenchido com a sequência FASTA de interesse antes de iniciar um processo que pode ser recuperado em qualquer outro momento a partir do número "*process ID*" (PID) gerado uma vez que o processo é iniciado.

É importante notar que no cabeçalho da sequência FASTA (primeira linha) está adicionado *taxonomy ID* ou *taxid* do organismo de interesse. É importante adicionar essa informação para que o cálculo do LCA não seja comprometido. Para os demais dados, não é necessário este adicional.

Uma vez submetida a sequência FASTA de interesse, o programa executa as seguintes funcionalidades:

- BLASTp: O BLAST (*Basic Local Alignment Search Tool*) é um algoritmo que trabalha o alinhamento local de sequências biológicas primárias, tais como aminoácidos ou nucleotídeos, e matrizes de pontuação. Pontuam-se os *matches* e *mismatches* conforme o valor atribuído para a substituição dos aminoácidos. Aqui, comumente são usadas as matrizes BLOSUM e PAM. Também há pontuação de abertura de *gaps*.
- PSI-BLAST (*Position-Specific Iterative BLAST*): Algoritmo indicado para a busca de relações distantes de uma proteína. Inicialmente é criada uma lista com as proteínas proximalmente relacionadas para que elas possam ser combinadas e seja gerado um perfil resumindo as principais características presentes nessas sequências. Este perfil é então utilizado

para consulta em banco de dados, onde um grupo maior de proteínas é encontrado. Com este novo grupo formado é construído um novo perfil e assim sucessivamente [70]. No MaLe-PSI-Blast, os PSI-BLASTs são realizados com cinco iterações.

- Cálculo do índice de confiança: tanto para os dados de BLASTp quanto PSI-BLAST, um *ensemble* de modelos de redes neurais e *random-forest* calcula a confiança de que cada *subject* de BLASTp/PSI-BLAST seja realmente relacionado à *query*. Para tal, estes algoritmos de aprendizado de máquina foram treinados em grupos PANTHER e utilizam as pontuações atribuídas aos *subjects* em cada iteração do PSI-BLAST bem como o tamanho da *query*. O BLASTp pode ser considerado um caso especial de PSI-BLAST onde o número de iterações é igual a 1 [71].
- Cálculo do LCA: Para se calcular o LCA, os *taxonomy* IDs de todas as proteínas presentes na tabela de resultados são obtidos a partir da tabela de *Idmapping* (consulta em SQL). A consulta retorna uma lista com 18 valores em que cada valor representa um táxon e as listas são percorridas até que se encontre o primeiro valor em comum. Este valor será o LCA.
- Construção da árvore filogenética e cálculo das distâncias: Para calcular as distâncias filogenéticas é utilizado o MUSCLE (alinhamento múltiplo) e o resultado é submetido no programa FastTree v.2.1.7 que utiliza métodos heurísticos para gerar uma árvore que atenda à máxima da verossimilhança. A árvore gerada é então processada por um parser em PERL que calcula o comprimento dos ramos entre a *query* e as demais sequências amostradas [70].
- Mineração de Anotação: Esta técnica as anotações das proteínas encontradas pelo MaLe-PSI-Blast são tratadas como um grupo (*cluster*) e então ordenadas pela sua proximidade com o centro deste grupo. No entanto para o cálculo deste centroide as anotações foram ponderadas conforme o índice de confiança da proteína (*subject*). Dessa forma, proteínas com maior índice de confiança tem maior peso e contribuem mais para a posição do centroide e proteínas com menor confiança (possivelmente falsos positivos) contribuem menos [71].

8.5 PathVisio

O PathVisio é um programa de livre acesso para análise e construção de vias biológicas. O programa apresenta uma interface simples, como mostrado na Figura 10, e é de fácil utilização, além de permitir visualização prática e fluida de vias no formato gml. O download do programa pode ser feito acessando www.pathvisio.org/, assim como recursos adicionais ao programa disponíveis em plug-in e tutoriais de uso.

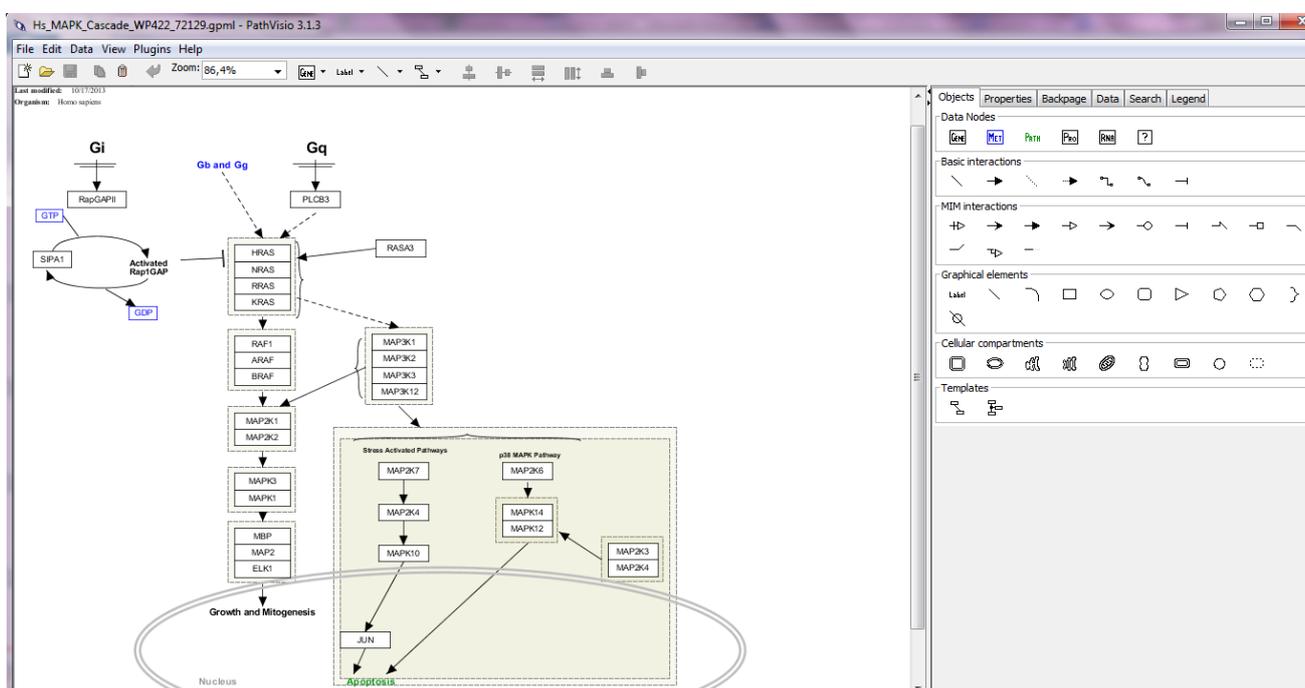


Figura 10: Interface gráfica do programa PathVisio.

8.6 MEGA

O programa MEGA é uma ferramenta integrada que realiza alinhamento de sequências, inferência de árvores filogenéticas, estimativa dos tempos de divergência, busca nas bases de dados online, estimativa das taxas de evolução molecular, inferência de sequencia ancestral e realiza testes de hipóteses nos estudos evolutivos. O MEGA é usado por um grande número de pesquisadores no sentido de reconstruir as histórias

evolutivas das espécies e inferir a extensão e natureza das forças seletivas que moldam a evolução dos genes e espécies de interesse [72].

O presente trabalho irá mostrar um estudo sobre a origem dos genes componentes de importantes subsistemas do sistema imune, sendo eles as vias de sinalização de IFNs, TCR e BCR. O mesmo foi feito para marcadores de NK e todos os receptores semelhantes a *Toll*.

VI. JUSTIFICATIVA

Grupos de proteínas homólogas são importantes fontes para propagação da informação biológica, bem como inferência da origem de genes. Há um interesse crescente no estudo de mecanismos de origem gênica e na origem de conjuntos gênicos associados a vias biológicas. O sistema imunológico se mostra um sistema adequado para esse tipo de estudo, uma vez que é provável a existência de genes de origem recente envolvidos nas sinalizações. Além disso, essa abordagem permitiria visualizar, pela primeira vez, como um sistema bem descrito e constantemente presente na literatura se formou ao longo da evolução. Quem seriam os primeiros interferons? Como foi sendo elaborado o mecanismo de sinalização para receptores das principais células do sistema imune? Os componentes da resposta inata aparecem antes ou depois da resposta adaptativa? Enfim, o presente trabalho se propõe a reproduzir o retrato da origem do sistema imune.

VII. OBJETIVOS

1. Objetivo geral

Ao longo da evolução dos sistemas, novos genes surgiram na natureza e para demonstrar a origem dos genes, as vias de sinalização do sistema imune revelam-se um modelo adequado. Assim, o presente trabalho visa delinear a construção do sistema imunológico ao longo da evolução, a partir das sequências de aminoácidos dos representantes humanos que serão usadas como *Seeds* para procurar homólogos.

2. Objetivos específicos

- Mostrar a construção da via de sinalização associada ao TCR.
- Mostrar a construção da via de sinalização associada ao BCR.
- Mostrar a construção da via de sinalização associada aos três tipos de IFN e respectivos produtos da ativação dessas vias.
- Estudar a distribuição filogenética dos TLRs humanos.
- Mostrar a evolução dos principais marcadores de superfície de NK.
- Retratar aspectos da origem do sistema imune.

VIII. METODOLOGIA

1. Obtenção de sequências e informações

Esta sessão descreve como foram buscadas e construídas as vias utilizadas nesse trabalho e, também, como foram buscados e verificados os identificadores UniProt referentes a cada membro componente destas vias.

1.1 Seleção das vias de sinalização

Para sistemas em que as vias de sinalização estão bem descritas, como aquelas utilizadas no presente trabalho, os dados podem ser encontrados em bancos de dados como Kegg Pathways (disponível em: <http://www.genome.jp/kegg/pathway.html>) e *Wikipathways* (disponível em: <http://www.wikipathways.org/index.php/WikiPathways>). As buscas devem ser feitas a partir das palavras-chave e do organismo de interesse, se houver.

Em muitos casos, o conteúdo das vias foi manipulado para a adição de informação. No presente trabalho, isso foi feito a partir da sobreposição da informação presente nas mesmas vias dos diferentes bancos e dados da literatura e esse processo foi realizado com a utilização do programa PathVisio versão 3.1.3, disponível para download em <http://www.pathvisio.org/downloads/>. O programa lê arquivos na extensão GPML e exporta o conteúdo em TXT, PWF, GPML, XML, MAPP, PDF, PNG, SVG e TIFF. Além disso, seu funcionamento requer Java 1.6 ou 1.7 e não é compatível com Java 8.

1.2 Seleção dos identificadores UniProt

Os genes e proteínas utilizados ao longo deste trabalho foram obtidos por busca manual através da consulta de diversas fontes. Foram retirados dados da literatura (livro texto, artigos e revisões) e de bancos de dados de imagens (KEGG *Pathways* e *Wikipathways*).

Uma vez encontrado o gene ou proteína de interesse é feita manualmente a busca pelo referente identificador UniProt. Em alguns casos, a molécula está associada a alguns dados, sendo eles:

- *Gene symbol*: correspondente à abreviação do nome de um gene, este dado será o único disponível em grande parte dos casos. Para tal, a busca pelo identificador *UniProt* deve ser feita diretamente na página desse banco de dados (<http://www.uniprot.org/>) e alguns cuidados devem ser tomados. A Figura 1 descreve os detalhes que devem ser consultado na busca pela proteína correta. Na barra superior faz-se a busca pelo *Gene symbol* ou nome do complexo, em alguns casos. Se há um organismo de interesse, este também pode ser adicionado à busca. Em seguida, o *UniProt* retorna uma lista com não só com a proteína de interesse, mas com aquelas relacionadas e de função similar, como mostrado pela caixa vermelha na figura. No caso de complexos, ele retorna os membros do mesmo e cabe ao usuário filtrar manualmente as proteínas de interesse. Para tanto, alguns itens disponíveis no perfil de cada gene/proteína devem ser cuidadosamente verificados, sendo eles: *Names & Taxonomy*, *Function* e *Sequence status*. Como nos passos seguintes serão feitas buscas por ortólogos. Somente entradas com *status Complete* (e não *Fragment*) foram utilizadas. Outras inúmeras informações relevantes estão disponíveis para o gene/proteína corretos sejam selecionados. Para o presente trabalho foram utilizadas apenas as sequências *SwissProt*. Estas sequências estão indicadas na figura pela seta amarela.

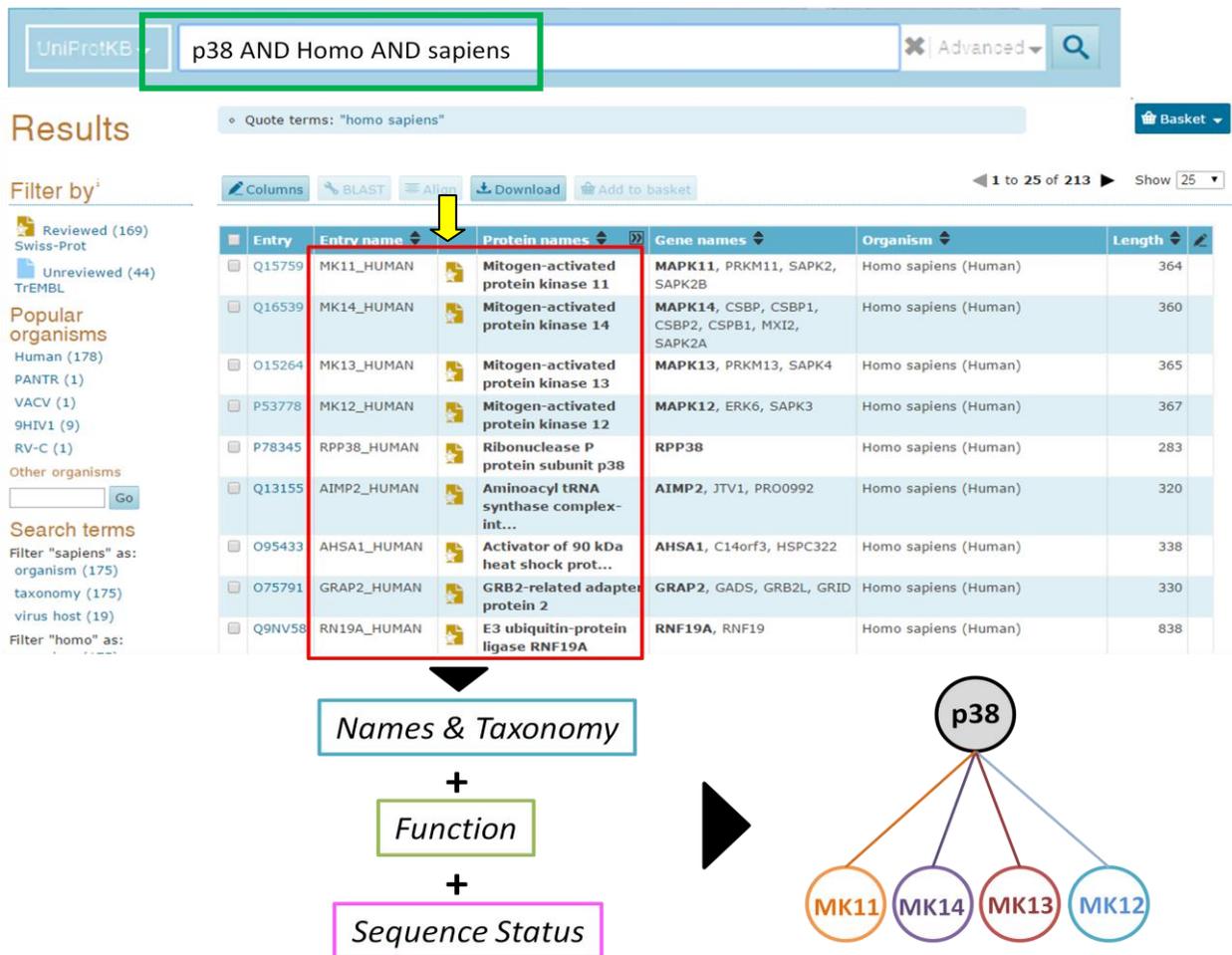


Figura 1: Busca dos UniProt ID de p38. Por ser um complexo, a busca por p38, indicada na caixa verde, retorna não só as proteínas componentes do complexo, mas também aquelas relacionadas ao mesmo, como indicado pela caixa vermelha. Verificando-se características de cada uma dessas proteínas, como itens indicados nas caixas azul, verde claro e rosa, é possível encontrar os membros de p38, indicados pelos círculos.

- Gene ID: sequência numérica que identifica um gene cujos dados podem ser acessados em <http://www.ncbi.nlm.nih.gov/gene>. Nessa página, o UniProt referente ao gene da busca se encontra no final da página, na sessão *Related sequences*, item *Protein Accession*.
- KO: conjunto de números precedido da letra K. A consulta deve ser feita a partir do endereço <http://www.genome.jp/kegg/kegg3.html> e na tabela contendo as informações do gene, na sessão *Genes* estão disponíveis os Gene ID por espécie e um link para os identificadores UniProt. Acessando a primeira opção, o Uniprot será encontrado como descrito acima e o acesso a segunda opção permite que os UniProt sejam inspecionados diretamente. Em alguns casos, um gene estará associado a mais de um KO.

Nesse caso, todos devem ser checados até que o UniProt adequado seja encontrado.

2. Recursos computacionais e programas utilizados

Para realização deste trabalho foram utilizados recursos computacionais do Laboratório Biodados localizado no Instituto de Ciências Biológicas, UFMG. Estes recursos incluem computadores *quad-core* com 4-8 GB de memória RAM e conexão com internet de 100MB/s. Também foi utilizada a máquina Pinguim da Faculdade de Medicina de Ribeirão Preto (USP) que contem 24 *cores* e 64 GB de memória RAM e 1 TB de HD. Todos estes computadores tinham sistema operacional Linux distribuição *CentOS*.

2.1 *Seedserver*

O agrupamento de gene homólogos foi feito utilizando a ferramenta *SeedServer*, cujo *download* pode ser feito através do link *Submit Job* disponível no servidor <http://maxixe.icb.ufmg.br:8080/BOWSWeb/appDetails.jsp?app=SeedServer> (cliente). As *Seeds* a serem usadas como *input* nesse programa devem ser sempre identificadores UniProt e a submissão dessas *Seeds* é feita a partir da seguinte linha de comando:

```
java -jar SeedServerSubmitJob.jar -UniprotID identificador UniProt
```

Uma vez submetido o processo, o programa retorna um número de controle (PID) para posterior recuperação do resultado que pode ser feita através da linha de comando

```
java -jar getresults.jar -pid PID
```

ou a partir do download de um programa Java (*Get Results GUI*) cuja interface gráfica permite a recuperação dos resultados através do PID e visualização facilitada fora do terminal. Para o funcionamento deste, é necessário que o Java esteja atualizado.

Na realização deste trabalho foram usado os parâmetros *default* do programa, como descrito por [73] e [74].

2.2 Análise do último ancestral comum (LCA)

A busca pelo LCA de dois ou mais organismos pode ser feito através do algoritmo disponível no REST *webservice* LCAWS que pode ser acessado pelo endereço <http://merengue.icb.ufmg.br:8080/BioToolsService/lca/9606+10096>. O *input* requer um conjunto de taxIDs e o *output* contém a informação taxonômica relacionada ao LCA referente aos taxIDs usados como *input*. A primeira resposta pode retornar um clado "*no rank*", mas também adiciona o próximo clado classificado na linhagem de acordo com a classificação de Lineu. Para o presente trabalho, foram usados os clados "*no rank*".

2.3 Genesis

O algoritmo de LCA aplicado a grupos de ortólogos é bastante preciso na inferência da origem da maioria dos genes. No entanto, esse algoritmo é ineficaz para casos em que genes foram transferidos para organismos em outros ramos da árvore da vida pelo fenômeno conhecido como transferência horizontal (ou lateral). Espera-se que esse evento seja muito mais frequente em genes originados em clados mais recentes de procariotos (como restritos a famílias e transferidos para gêneros distantes). Porém, análises manuais detectaram alguns casos em eucariotos. O algoritmo denominado LCA Múltiplo, além de aumentar a precisão do algoritmo de LCA original, é também capaz de detectar transferências horizontais e deleções gênicas.

Com objetivo de desmembrar a sub-árvore originada no nodo resultante do LCA simples em sub-árvores menores e indicar os clados onde um gene aparece pela primeira

vez em um ramo da árvore (Figura 2), seja pelo surgimento *de novo*, seja por transferências horizontais, o LCA múltiplo foi utilizado. Esta ferramenta, denominada MULTILCA TREE está disponível no *website* <http://biodados.icb.ufmg.br:8080/Genesis/multiLCATree/index>. O programa exige duas listas como entrada: a lista de espécies que possui o gene (positiva) e a lista de espécies que comprovadamente não possui o gene (negativa). No presente trabalho as listas positivas compreendem os taxIDs obtidos do *output* do *Seedserver* e a negativa, uma lista de genomas completos presente no próprio Genesis.

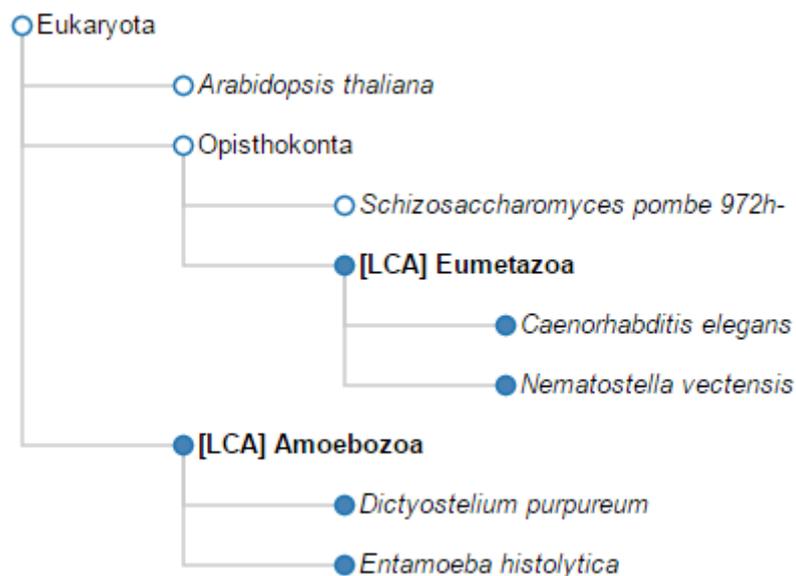


Figura 2: Árvore construída pela ferramenta MULTILCA TREE. Para este exemplo foi utilizado o KO K00923 (Phosphatidylinositol-4-phosphate 3-kinase).

3. MaLe-PSI-BLAST

Este é outro recurso que foi utilizado para inspecionar possíveis ancestrais de proteínas em casos especiais. O PSI-BLAST é uma poderosa ferramenta para a captura de homólogos remotos. No entanto depois de várias iterações a *Position Specific Substitution Matrix* (PSSM) construída pelo programa pode deteriorar, sendo que a própria *query*

pode não constar da lista de *hits*. O MaLe-PSI-BLAST usa ferramentas de aprendizado de máquina para evitar esse fator, melhorando a busca por homólogos remotos. O download deste serviço pode ser feito através do link *Submit Job* disponível em <http://maxixe.icb.ufmg.br:8080/BOWSWeb/appDetails.jsp?app=malepsiblast>.

Sequências FASTA devem ser usadas como *input* nesse programa e a submissão de um processo é feita a partir da seguinte linha de comando:

```
java -jar malepsiblastSubmitJob.jar -fasta nome do arquivo
```

Assim como no *SeedServer*, o programa retorna um número de processo (PID) para posterior recuperação do resultado que pode ser feita através da linha de comando

```
java -jar getresults.jar -pid PID
```

ou através da janela *Get Results* GUI.

Na realização deste trabalho foram usado os parâmetros *default* do programa, conforme descrito na tese do Dr. Henrique de Assis Lopes Ribeiro.

4. MEGA

As sequências completas de TLRs humanas foram analisadas em separado utilizando o programa MEGA 6 [75], disponível para download em <http://www.megasoftware.net/>. O alinhamento das sequências foi realizado utilizando o programa MUSCLE [76] implementado no MEGA, utilizando parâmetros padrão do programa (*gap opening* = -2,9; *gap extend* = 0; *hydrophobicity multiplier* = 1,2). Árvores filogenéticas foram construídas com os métodos *Neighbor Joining* (bootstrap: 10000) e *Maximum Likelihood* (bootstrap: 100) também implementados no MEGA.

IX. RESULTADOS

Examinando a distribuição taxonômica dos genes homólogos, foi possível delinear os clados da linhagem humana onde cada proteína componente das vias escolhidas para serem caracterizadas neste trabalho ocorreu pela primeira vez. Para tanto, como descrito no item “Material e Métodos”, as vias foram inicialmente estruturadas e a partir do nome oficial do Gene (símbolo gênico), foi feita a busca pelo identificador UniProt da proteína e com ele o sistema denominado *SeedServer*, produzido no Laboratório de Biodados, foi executado para encontrar os homólogos e conseqüentemente listar os organismos que os contêm.

A seguir serão descritas as vias analisadas para retratar a evolução da resposta imune: as sinalizações associadas ao complexo do receptor de linfócito T, ao complexo receptor linfócito B e aos três tipos de interferons. E ainda serão analisados os receptores semelhantes a *Toll* e os marcadores de células *Natural killer*.

1. Construção da sinalização associada aos Interferons tipo I

Para combater as infecções virais, as células do sistema imunológico desenvolveram um eficiente sistema integrado de defesa que associa a imunidade inata à adaptativa. Uma vez que o vírus invade a célula, esse sistema é ativado a partir da indução de citocinas chamadas interferons [77]. Como descrito no item 4.1.1 (em Introdução), os IFNs constituem um grupo heterogêneo de proteínas e são conhecidos pela sua habilidade de induzir resistência celular a invasões virais. Entretanto, os IFNs ainda são capazes de interferir em outras funções, como o crescimento celular. Utilizando a abordagem preclara, mapeamos a origem dos diversos IFNs.

A Figura 1 mostra a distribuição dessas citocinas, de acordo com seu suposto clado de origem. Esse grupo de moléculas é dividido em três grupos, IFNs do tipo I, II e III. Dentre os IFNs do tipo I, verificamos que o mais ancestral é o IFN α , originado em Euteleostomi. Isso significa que genes homólogos ao humano foram encontrados em vertebrados incluindo os peixes, pois o clado Euteleostomi é o último (mais recente) que é comum entre as linhagens evolutivas do homem e do peixe-zebra, por exemplo. Na mesma época, Euteleostomi, também encontramos evidência sugerindo a origem do único IFN do tipo II, o IFN γ . Os demais IFNs do tipo I têm origem mais recente, IFN β , IFN ϵ e IFN κ Amniota e IFN ω Eutheria. Isso significa que os primeiros podem ser encontrados também em aves, que compartilham a linhagem evolutiva com o homem até o clado Amniota, enquanto o IFN ω por sua vez é restrito a organismos placentários (Eutheria).

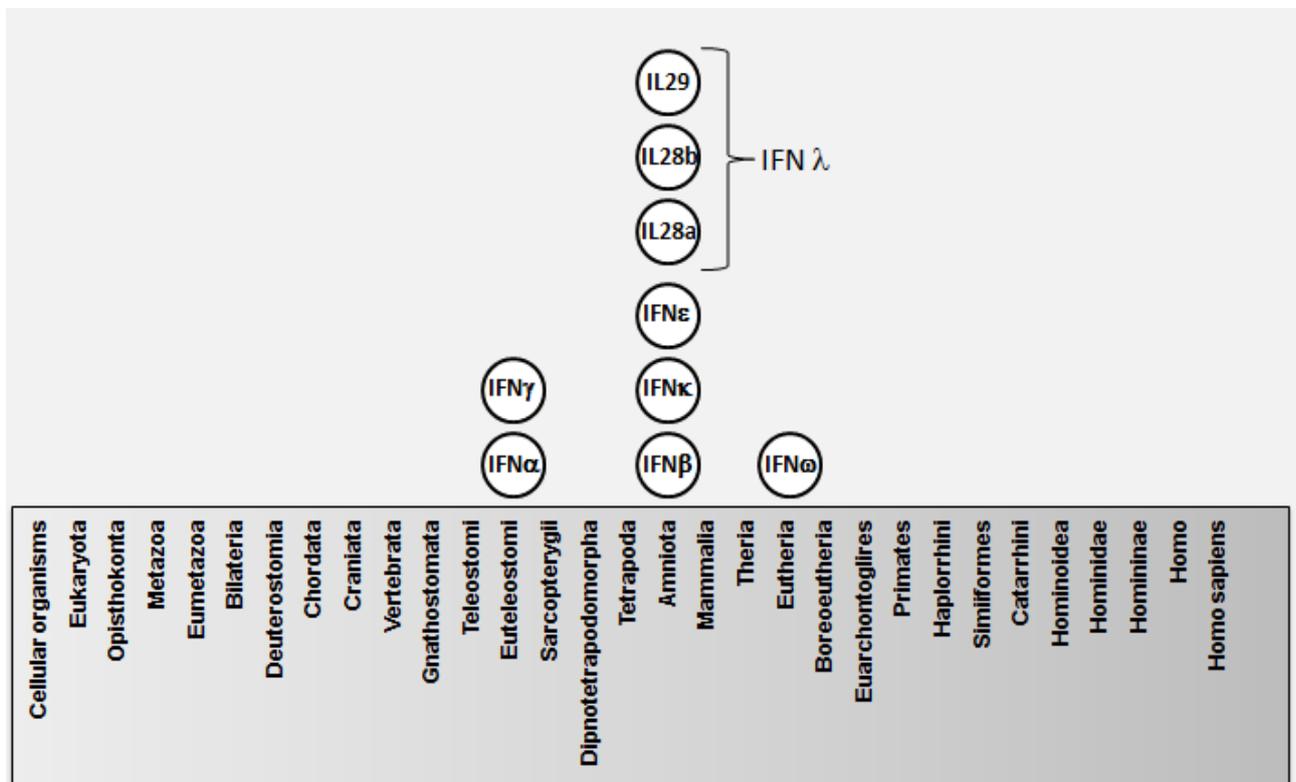


Figura 1: Distribuição dos tipos de interferon de acordo com seu clado de origem ao longo da linhagem evolutiva de Homo sapiens.

O quadro fica mais completo quando abordamos toda a sinalização efetuada pelos IFNs. A Figura 2 mostra a via de sinalização dos IFNs tipo I: a família $IFN\alpha$, $IFN\beta$, $IFN\epsilon$, $IFN\kappa$ e $IFN\omega$ foram representados por hexágonos, com suas ancestralidades sinalizadas por círculos preenchidos e assinalados. A infecção viral induz a produção de IFNs tipo I e, a partir da interação com o receptor, induz a estimulação de uma série de genes que inibem a replicação viral e aumentam o potencial lítico das células NK. Além disso, os IFNs tipo I modulam a resposta adaptativa através do aumento da expressão de MHC I [77].

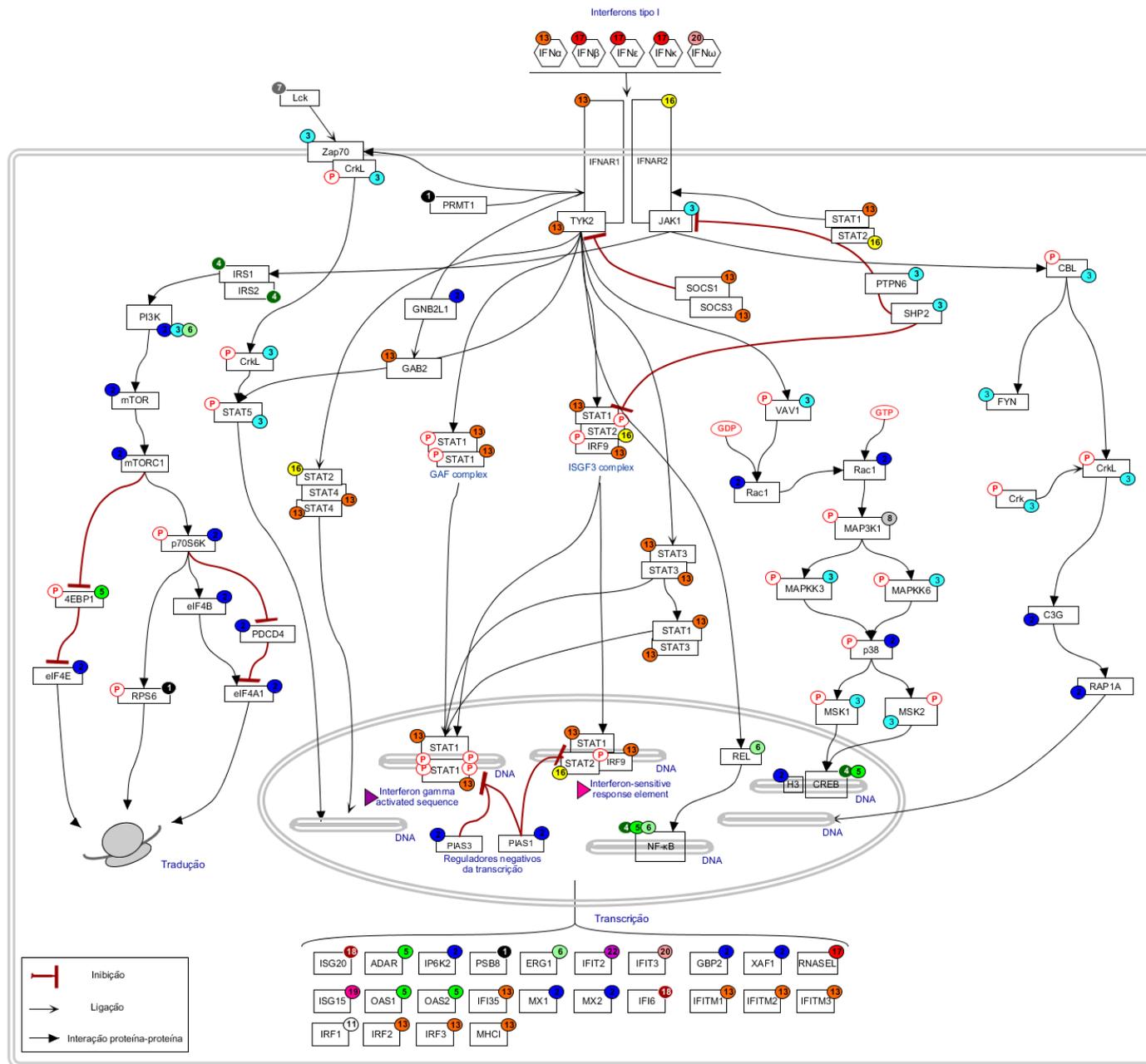


Figura 2: Via de sinalização dos interferons tipo I. O LCA dos genes utilizados na construção dessa via estão distribuídos e identificados pelos clados *Cellular organisms*, clado 1; *Eukaryota*, clado 2; *Opisthokonta*, clado 3; *Metazoa*, clado 4; *Eumetazoa*, clado 5; *Bilateria*, clado 6; *Deuterostomia*, clado 7; *Chordata*, clado 8; *Gnathostomata* (em branco), clado 11; *Euteleostomi*, clado 13; *Tetrapoda*, clado 16; *Amniota*, clado 17; *Mammalia*, clado 18; *Theria*, clado 19; *Eutheria*, clado 20; *Euarchontoglires*, clado 22. **Nota:** alguns clados intermediários não tiveram ocorrência (ex: *Sarcopterygii*, clado 12).

Como representado na Figura 2, os diversos IFNs do tipo I se ligam a um único receptor heterodimérico [78] e cada uma de suas subunidades interage com um membro da família JAK. O receptor dos IFN tipo I tem sua subunidade IFNAR1 associada a TYK2 e sua subunidade principal IFNAR2 associada a JAK1 [79]. Um marco importante na evolução dessa via é a origem dos Euteleostomi e Tetrapoda, visto que estes são os clados onde é observada a origem das duas subunidades dos receptores (na figura estão representados por laranja, clado 13 e amarelo, clado 16, respectivamente). O fato de que a subunidade principal IFNAR2 tem sua origem alguns passos evolutivos mais tarde, em Tetrapoda, indica um possível aperfeiçoamento do sistema nesta época. Nos organismos Euteleostomi que não evoluem para Tetrapoda, como o peixe-zebra, e que não possuem essa cadeia, ou a ligação ocorre em um receptor homodimérico ou uma outra cadeia diferente forma o receptor heterodimérico.

A etapa inicial da sinalização dessas duas vias consiste na ativação do receptor associado às JAKs que ocorre em função do estímulo pelos IFNs. Em seguida, essas JAKs sofrem auto-fosforilação e são ativadas [79]. A ligação dos IFNs tipo I ao receptor resulta na rápida ativação do receptor associado a JAK e TYK2, então JAK1 regula a fosforilação das STATs STAT1, STAT2, STAT3 e STAT5 [80]. As STATs 4 e 6 também podem ser ativadas, mas esses tipos estão restritos a apenas alguns tipos celulares, como células endoteliais [81]. Sendo os componentes centrais da transdução de sinal, é importante ressaltar a ancestralidade dessas proteínas. Em sua maioria são originadas conjuntamente com a origem dos ossos, em Euteleostimi (em laranja, clado 13), como mostra a Figura 2. Em contraste, JAK1 e STAT5 mostram-se como componentes mais ancestrais, ao ter como LCA o clado Opisthokonta (em azul claro, clado 3). Este clado reúne o homem e fungos.

Após a fosforilação e ativação das STATs por JAK, essas proteínas se translocam para o núcleo onde se ligam a sítios específicos nos promotores de “genes estimulados por IFNs” (ISGs) [82]. Um dos principais complexos transcricionais é o ISGF3, composto pelas formas ativas de STAT1, STAT2 e IRF9 [80]. Esse complexo é o único que se liga a elementos específicos chamados *Interferon sensitive response elements* - ISREs (na Figura 2, assinalado pelo triângulo rosa) - que estão presentes nos promotores de alguns ISGs, assim iniciando a sua transcrição. Outros complexos de STATs são induzidos por IFNs tipo I, sendo alguns deles os homodímeros de STATs 1, 3, 4 e 5 e dímeros como STAT1-STAT1 (complexo GAF) e STAT1-STAT2 [83]. Os complexos induzidos por IFNs também ativam elementos do tipo GAS (*IFN-Activated site*), também presentes em ISGs [80] e na figura representados pelo triângulo roxo. Assim, a transdução de sinal principal é originada à época dos primeiros IFNs, com a origem de TYK2, STAT1, STAT3, STAT4 e IRF9 em Euteleostomi (em laranja, clado 13), sendo JAK1 mais antiga e STAT2 aparecendo (ou atingindo a composição similar à atual humana) posteriormente, em Tetrapoda.

Outro membro da via regulado pelos IFNs é a via das MAPKs [84]. JAK regula a fosforilação de VAV, resultando na ativação de Rac1 que, por sua vez, regula a sinalização via p38. Uma cadeia de MAPK kinase kinase (MAPKKK) também é ativada: MAPKKK é ativada e regula MAPKK3 e MAPKK6, que fosforilam p38, resultando em sua ativação. A p38 ativada regula a atividade de MSK1 e MSK2 [80, 84]. As proteínas transcrição reguladas por p38 incluem CREB e Histona H3 ativada (na figura, estão localizados no núcleo que está representado pela elipse em linha dupla). Em geral, esta parte da sinalização é antiga, provavelmente tendo sido recrutada na origem da sinalização mais específica.

TYK2 e JAK1 regulam a fosforilação de IRS1 e IRS2, o que promove sítios de ligação para a subunidade p85 de PI3K. Uma vez ativadas, esse conjunto de moléculas regulam a ativação de mTOR, sendo este último responsável pela iniciação da tradução de mRNAs e pela fosforilação de p70S6K. Então, RPS6 é fosforilada resultando também na iniciação da tradução. mTOR também regula a fosforilação do repressor da tradução 4EBP1 [85], que gera a desativação e consequente dissociação de eIF4E, permitindo a iniciação da tradução dependente de CAP. Como observado à esquerda figura, esse grupo de proteínas tem origem ancestral, sendo, em sua maioria, originadas em Eukaryota (em azul escuro, clado 2).

Observando a distribuição de cores na figura, se nota que a transdução de sinal está dividida em dois componentes: a sinalização JAK-STAT formada em grande parte por componentes cuja origem se deu em Euteleostomi (em laranja, clado 13) e as sinalizações adjacentes compostas majoritariamente por membros de origem ancestral, se destacando Eukaryota e Opisthokonta (em azul escuro, clado 2 e em azul claro, clado 3, respectivamente). A Figura 3 representa a distribuição dos componentes que podem ser considerados citoplasmáticos por clado, sendo estes compostos por um total de 70 proteínas distribuídas em 10 diferentes clados e tendo 80% deles se distribuído entre *Cellular organisms* e Bilateria. Considerando-se que em Euteleostomi é originada a sinalização mais específica, verifica-se que grande parte dos componentes se origina contrariamente em clados bem ancestrais como Eukaryota e Opisthokonta.

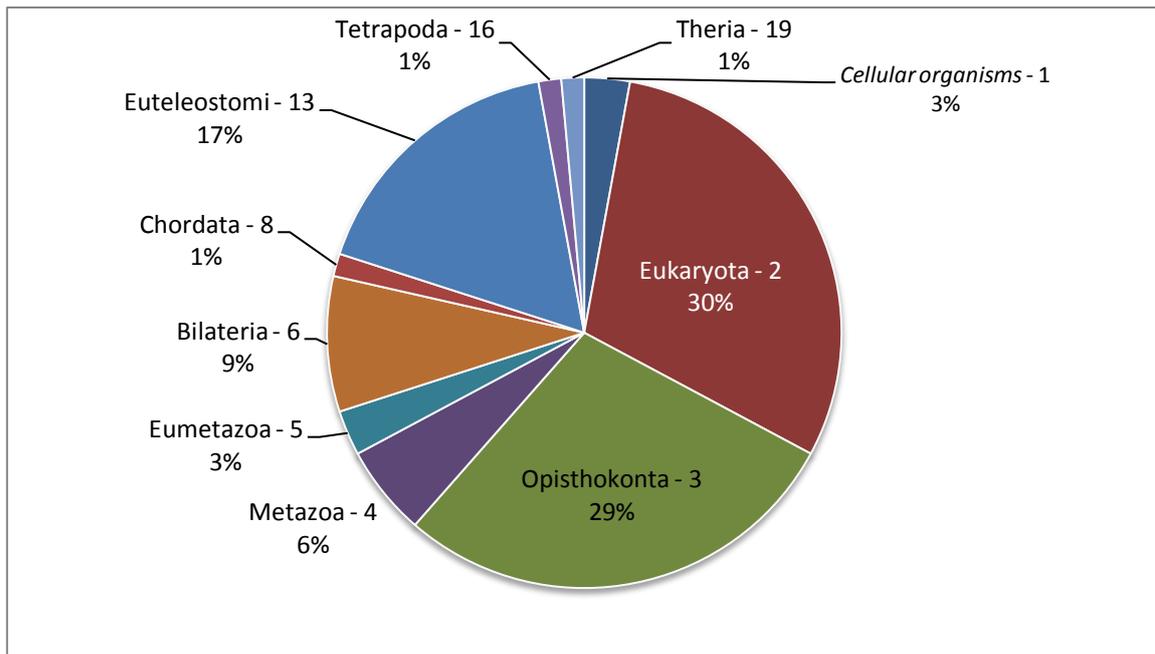


Figura 3: Proteínas do citosol na sinalização IFN tipo I. Distribuição de componentes citoplasmáticos da transdução de sinal de acordo com seu clado de origem ao longo da linhagem evolutiva de *Homo sapiens*. A ordem evolutiva segue em sentido horário a partir de *Cellular organisms*. **Nota:** as cores na área do gráfico foram escolhidas de forma arbitrária.

No núcleo, a distribuição da origem das 11 proteínas acontece entre os clados Eukaryota e Bilateria. Sendo assim, esse compartimento celular mostra-se o mais ancestral quando comparado aos demais compartimentos. A Figura 4 representa graficamente a distribuição dessas proteínas.

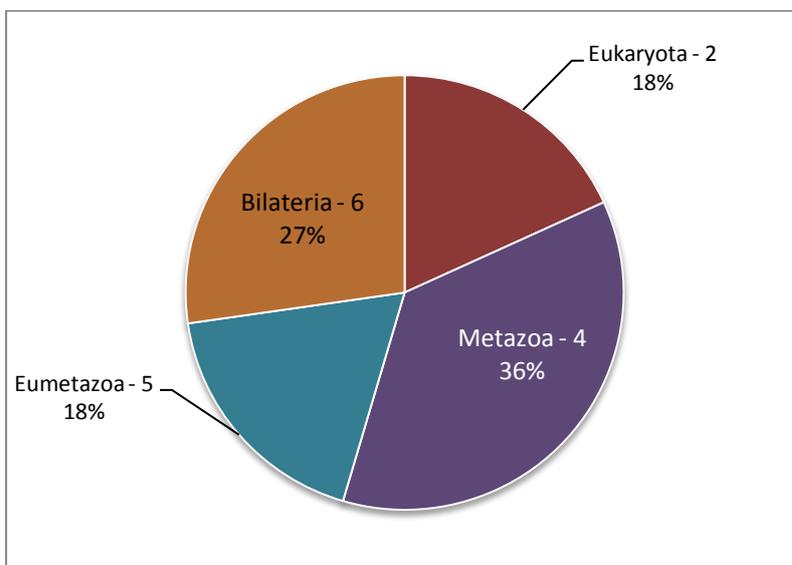


Figura 4: Proteínas do núcleo na sinalização IFN tipo I. Distribuição de componentes nucleares de acordo com seu clado de origem ao longo da linhagem evolutiva de *Homo sapiens*. Foram contabilizados apenas componentes exclusivamente nucleares. A ordem evolutiva segue em sentido horário a partir de Eukaryota. **Nota:** as cores na área do gráfico foram escolhidas de forma arbitrária.

De um modo geral, os componentes dos compartimentos celulares citosol e núcleo são mais ancestrais em relação àqueles presentes na membrana. Entretanto, a principal sinalização, JAK-STAT, só está completa nos moldes como se conhece atualmente em Tetrapoda. Isso acontece devido a STAT2 ter sua origem nesse clado e essa proteína é imprescindível para a formação do complexo ISGF3, sendo este, por sua vez, necessário para a ativação da transcrição. Sugere-se, então, que aos componentes ancestrais da transdução de sinal, foi associada a resposta JAK-STAT, sendo esta a principal e mais adequada sinalizadora entre o estímulo dado pela citocina e a expressão gênica. O mesmo é válido para os componentes nucleares.

A via induzida por IFNs tipo I ativa a síntese de genes associados à resposta imune. Os principais produtos da ativação por essa citocina estão organizados por uma chave na parte inferior da Figura 2 e estão indicados quantitativamente na Figura 5. Neste caso, nota-se uma contribuição importante de clados recentes.

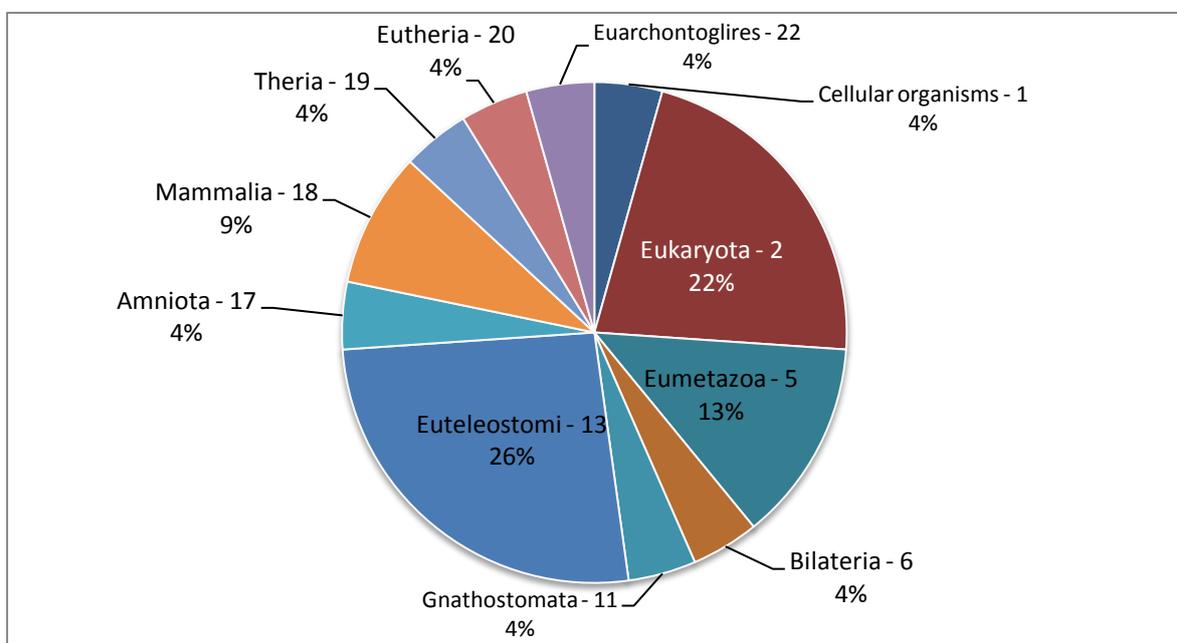


Figura 5: Genes transcritos com a ativação da via IFN tipo I. Distribuição dos principais produtos da ativação por IFN α e IFN β de acordo com seu clado de origem ao longo da linhagem evolutiva de *Homo sapiens*. A ordem evolutiva segue em sentido horário a partir de *Cellular organisms*. Nota: as cores na área do gráfico foram escolhidas de forma arbitrária.

A ativação dessa via ainda leva a ativação de MHC classe I. Todas as moléculas componentes de MHC classe I apresentam seu LCA em Euteleostomi. Observa-se que 54% dos genes ativados por esses IFNs tem sua origem mais ancestral que a origem do IFN α e essa porcentagem sobe para 80% considerando-se o IFN β . Isso sugere que os promotores desses genes se tornaram responsivos ao estímulo por essas citocinas posteriormente à sua origem, em Euteleostomi e Amniota, respectivamente.

1.1 Necessidade de inspeção manual e o caso de JAK1

Para fins de refinar as buscas pelos LCAs, os resultados no relatório *taxonomy report* de todas as proteínas componentes das vias foram verificados com cuidado manual. Nesse contexto, um caso especial foi observado, sendo ele a proteína *Janus kinase 1* (JAK1). Neste, a presença de uma proteína do coanoflagelado *Monisiga ovata* no *cluster* de genes homólogos atrai o LCA da mesma para um clado bem ancestral: Opisthokonta. Então, foi feita uma análise de LCA múltiplo no sentido de refinar a busca pelo LCA de JAK1. Esta ferramenta foi desenvolvida recentemente e com ela a determinação do LCA não é afetada por possíveis transferências laterais, já que em caso de transferências são reportados o LCA que envolve as transferências verticais associadas ao homem e, em separado, um segundo LCA para o nó onde ocorre a possível transferência, ou convergência, ou o agrupamento por artefato de uma proteína sem relação de homologia. O algoritmo do LCA múltiplo considera organismos com genoma completo que não possuem o gene para interromper a busca e propor múltiplos LCAs. Isso pode ser observado na árvore mostrada na Figura 6, resultante do uso da ferramenta *Multiple LCA tree* do Genesis, descrito no item 2.3, em Metodologia.

Num contexto onde a sinalização JAK-STAT se constrói entre os clados Euteleostomi e Tetrapoda, como observado na Figura 2 e na Figura 8 (adiante), pode-se sugerir que a presença de *M. ovata* dá-se a partir da transferência lateral do gene JAK1. Como uma tentativa de buscar o LCA real dessa proteína, *M. ovata* foi excluído do *cluster*. O novo resultado pode ser observado na Figura 6B.

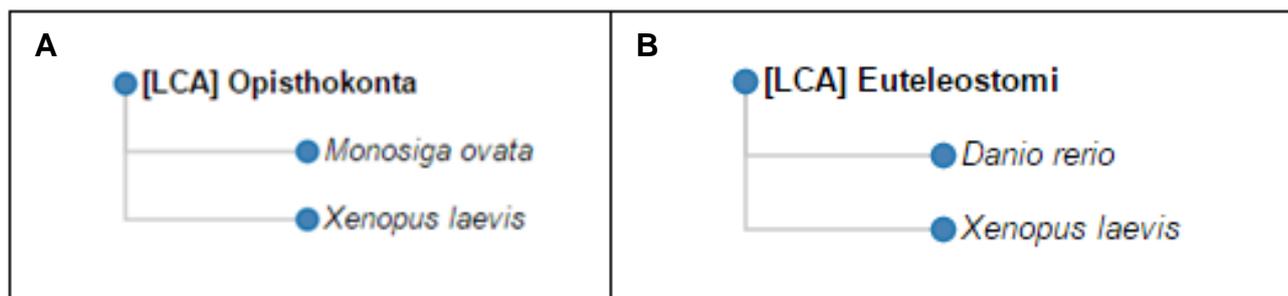


Figura 6: Transferência lateral do gene JAK1. 6A: árvore de LCA múltiplo construída com todos os organismos do *cluster* de homólogos indicando LCA em Opisthokonta. 6B: árvore de LCA múltiplo construída com todos os organismos do *clusters* de homólogos, exceto *M. ovata*, indicando LCA em Euteleostomi.

2. Construção da sinalização associada aos Interferons Tipo II e Tipo III

O IFN tipo II é formado apenas pelo IFN γ . Essa citocina apresenta inúmeras atividades imuno-regulatórias e em humanos se liga a um receptor próprio formado por duas subunidades homodiméricas: a subunidade IFNGR1 que associa-se a JAK1 e IFNGR2 a JAK2 [82]. A Figura 7 mostra o LCA de IFN γ em Euteleostomi (hexágono, em laranja, clado 13), em contrapartida, as duas subunidades do receptor tem sua origem em Tetrapoda (em amarelo, clado 16). Nota-se que a citocina é um pouco mais ancestral que seu receptor. Esse evento será discutido posteriormente no item 4.2, onde buscamos possíveis genes ancestrais das cadeias do receptor utilizando uma metodologia apropriada para encontrar homólogos distantes.

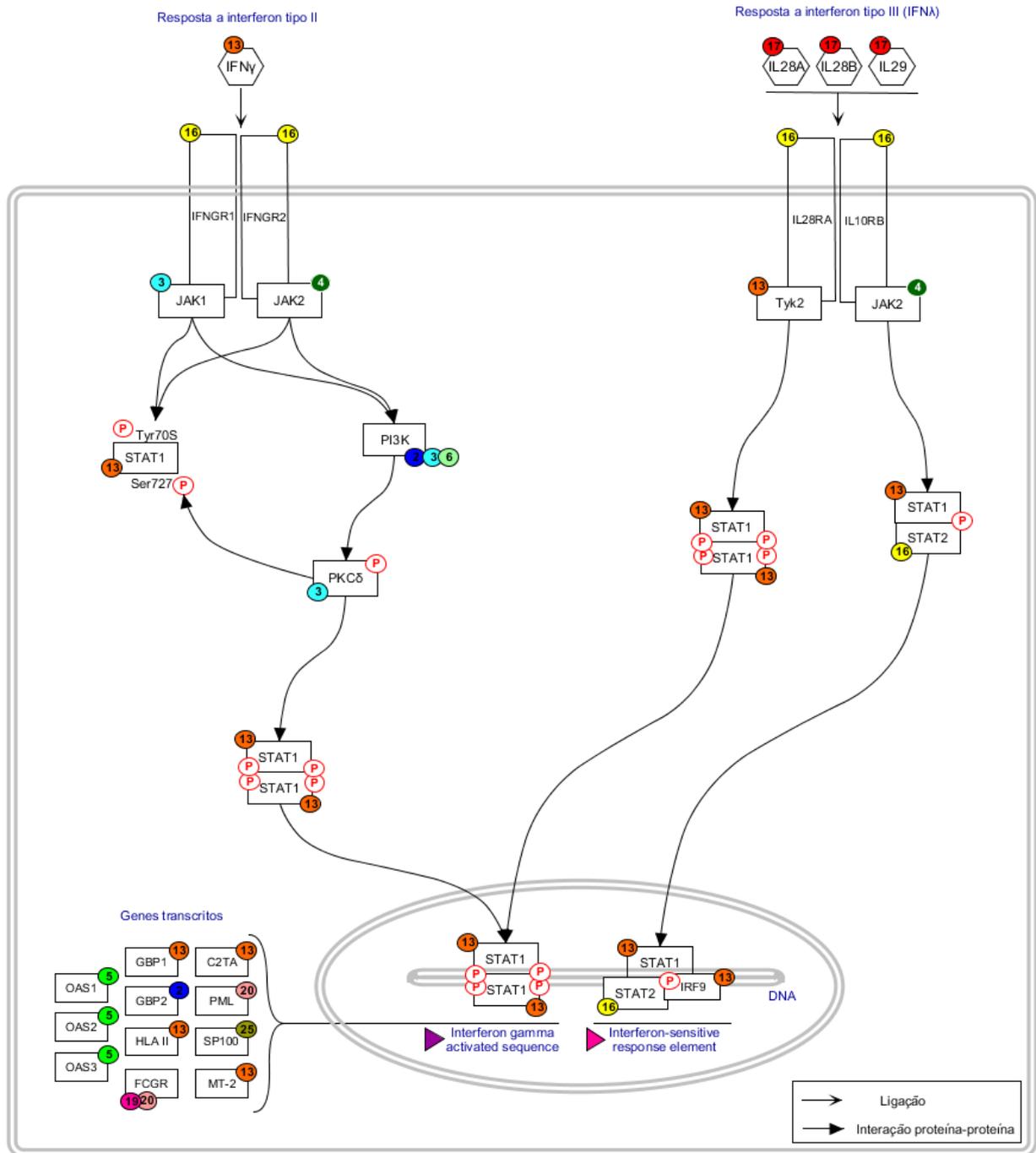


Figura 7: Via de sinalização dos interferons tipo II e III. O LCA dos genes utilizados na construção dessa via estão distribuídos e identificados pelos clados: **Eukaryota**, clado 2; **Opisthokonta**, clado 3; **Metazoa**, clado 4; **Eumetazoa**, clado 5; **Bilateria**, clado 6; **Euteleostomi**, clado 13; **Tetrapoda**, clado 16; **Amniota**, clado 17; **Theria**, clado 19; **Eutheria**, clado 20; **Simiiformes**, clado 25. **Nota:** alguns clados intermediários não tiveram ocorrência (ex: Sarcopterygii, clado 12).

O estímulo de IFN γ ao seu receptor gera a ativação de JAK1 e JAK2 e estas regulam a fosforilação de STAT1 no resíduo de tirosina na posição Tyr701 [80]. Essa forma fosforilada de STAT1 forma homodímeros e transloca-se para o núcleo, onde se liga em elementos GAS (na figura, assinalado pelo triângulo roxo), presentes nos promotores dos genes regulados por IFN γ . Uma vez ativadas, as JAKs também regulam a ativação da subunidade p110 de PI3K. Esse evento resulta na ativação de PKC δ , que então regula a fosforilação de STAT1 em Ser727 [86]. Apesar da fosforilação do resíduo de serina não ser essencial para a translocação ao núcleo ou ligação às regiões promotoras, ela é necessária para a ativação completa da transcrição.

Quando fosforilada, STAT1 forma um homodímero que se transloca para o núcleo, onde se liga ao elemento GAS [87]. Diferente de como acontece a sinalização dos IFNs tipo I, o IFN γ não inclui a formação de ISGF3 e por isso não é capaz de induzir os genes que tem apenas ISREs (na figura, assinalado pelo triângulo roxo) em seus promotores. A origem das STATs em períodos aproximados a do IFN γ e seu receptor reforça o que foi observado para a sinalização pelos interferons IFN α .

A ativação por IFN γ induz a transcrição de proteínas envolvidas nas reações imunológicas e participantes de outros processos celulares, como diferenciação e proliferação. Os principais produtos da ativação por essa citocina estão organizados por uma chave na parte inferior da Figura 7 e estão indicados quantitativamente na Figura 8.

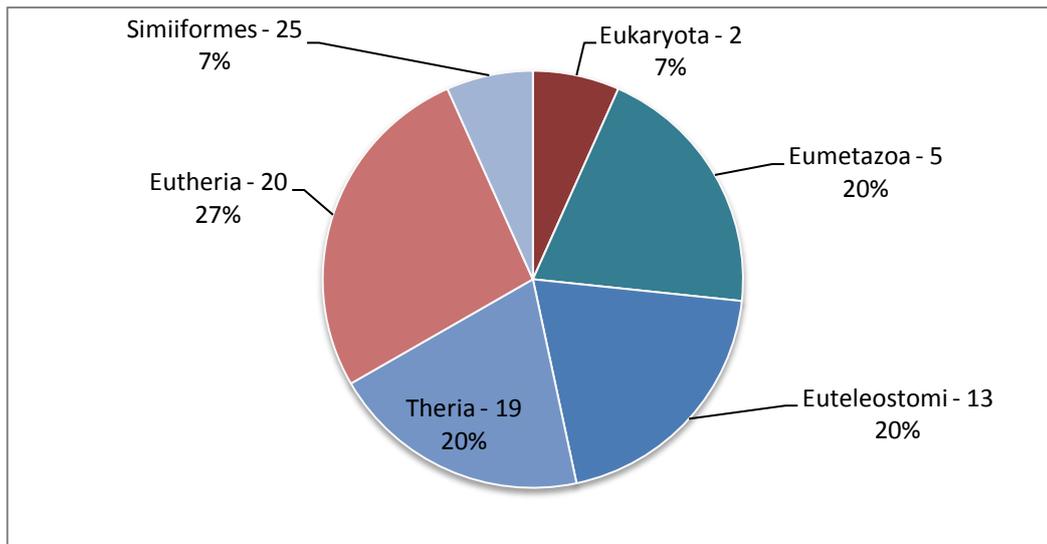


Figura 8: Genes transcritos com a ativação da via IFN tipo II. Distribuição dos principais produtos da ativação por IFN γ de acordo com seu clado de origem ao longo da linhagem evolutiva de Homo sapiens. A ordem evolutiva segue em sentido horário a partir de Eukaryota. Nota: as cores na área do gráfico foram escolhidas de forma arbitrária.

Ainda, a partir da ativação dessa via, há a indução da síntese de MHC classe II (não contabilizado na Figura 8). Todas as moléculas componentes de MHC de classe II, sendo um total de 28 membros, apresentam seu LCA em Euteleostomi. Em concordância com o que ocorre com a sinalização a IFN α , reforça-se o conceito de que um passo evolutivo importante na construção de ambas as respostas é a evolução dos Tetrapoda. Nesse passo evolutivo, toda a sinalização específica já está formada (JAK-STAT) e o fato de alguns genes induzidos terem sua origem anterior, nos clados Eukaryota e Eumetazoa, sugere que os promotores desses genes passaram a ser responsivos à atividade do IFN γ .

Há ainda uma terceira classe de IFNs, os interferons do tipo III. Também conhecidos como IFN λ (IL28A, IL28B e IL29) [88], essas moléculas também possuem atividade antiviral e se ligam há um tipo diferente de receptor. A sinalização gerada pelo estímulo dessas citocinas também está descrita na Figura 7. Esses IFNs tem sua origem em Amniota (hexágono, em vermelho, clado 17) se ligam a um receptor heterodimérico, formado por IL28RA e IL10RB. As duas subunidades do receptor tem sua origem anterior,

em Tetrapoda (em amarelo, clado 16). IL10RB também funciona como uma das subunidades do receptor de IL10 [89]. A ativação desse receptor leva a ativação de STAT1 e STAT2 e conseqüentemente à formação de complexos ISGF3 e GAF. Por esse motivo são capazes de ativar genes que apresentam ISRE e GAS, indicados pelos triângulos rosa e roxo, respectivamente, em seus promotores.

Esses IFNs são similares aos IFNs tipo I, uma vez que sua ativação é iniciada pela infecção viral ou presença de RNA dupla fita. Eles aumentam a resistência da célula à infecção e ativam uma via de transdução de sinal similar aos IFNs tipo I [90]. Essa citocina representa uma conexão evolutiva entre IL10 e IFNs tipo I, contudo são estruturalmente mais similares aos IFNs. Essa molécula também ativa STAT3 e STAT5 em outras sinalizações, que estão caracteristicamente mais associadas à sinalização de IL10 [91]. A origem das moléculas que são reunidas sobre a denominação de $INF\lambda$ em Amniota indica que trata-se de uma adaptação evolutiva bem mais recente, contemporânea da origem do $INF\beta$.

2.1 Origem dos Fatores Reguladores de Interferon

Ainda foram analisados todos os fatores reguladores de interferon (IRFs), mesmo aqueles que não compõem as vias acima estudadas. A distribuição desses IRFs pode ser observada na Figura 9. Notavelmente, o sistema regulatório tem uma evolução mais ancestral, sendo apenas IRF5 mais recente que a maioria das proteínas chave na resposta como os próprios IFNs, os receptores e as STATs.

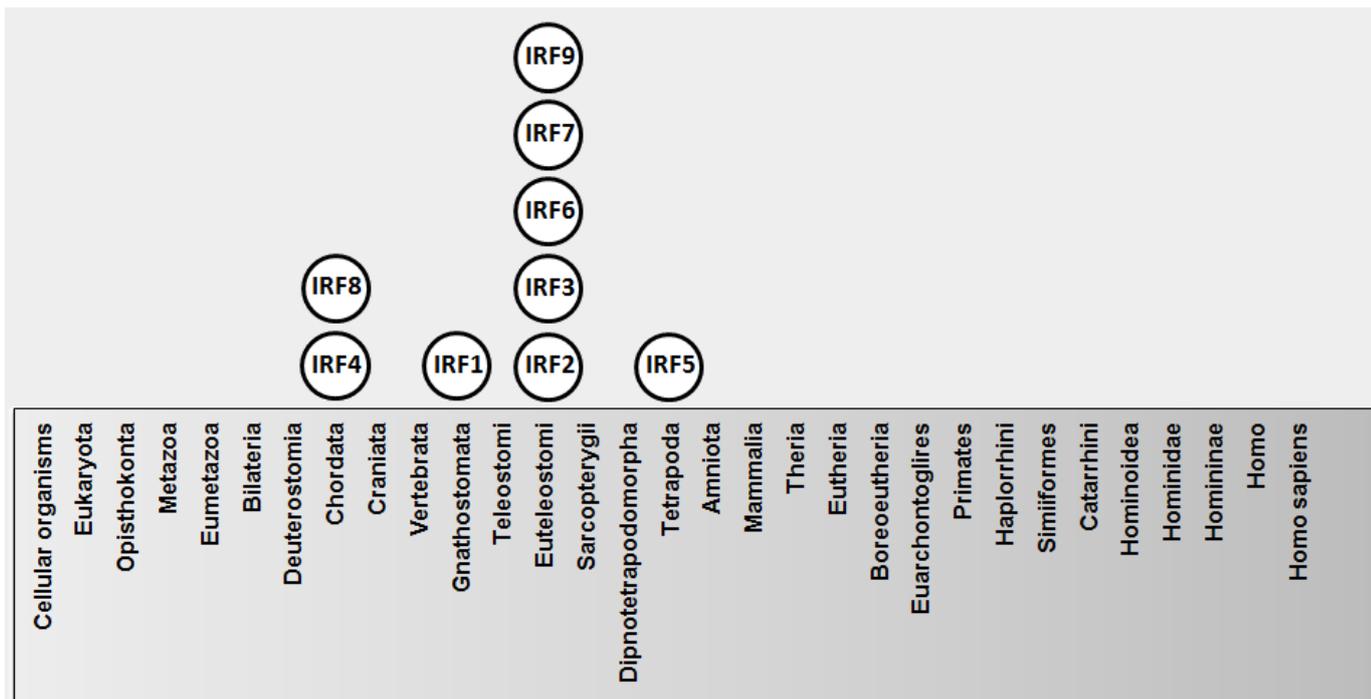


Figura 9: Distribuição dos fatores de regulação de interferon de acordo com seu clado de origem ao longo da linhagem evolutiva de Homo sapiens.

2.2 Busca de um ancestral remoto para receptor de IFN γ

A partir da Figura 7 observa-se que as duas subunidades do receptor de IFN γ têm sua origem em Tetrapoda enquanto o próprio IFN γ origina-se em Euteleostomi. Sabendo-se que este é um clado ancestral em relação a Tetrapoda e que, de acordo com a literatura, IFN γ é capaz de interagir exclusivamente com seu receptor, ficam as perguntas: (i) estaria IFN γ funcionando como molécula secundária antes do advento de seu receptor? (ii) seria IFN γ produto da modificação de outras citocinas mais ancestrais?

Em uma tentativa de buscar possíveis ancestrais para as cadeias do receptor, foram feitas análises com PSI-BLAST. Utilizando matrizes de substituição construídas com os resultados da primeira busca, o PSI-BLAST consegue detectar famílias gênicas próximas. Foi utilizada uma ferramenta construída no Laboratório de Biodados (UFMG) onde inteligência artificial classifica os candidatos da busca e atribui um valor de

confiança à saída (além dos parâmetros de identidade, vindo do próprio BLAST, e *SelfScore*, que é o *score* da própria *query* na iteração de PSI-BLAST utilizada para a classificação). Surpreendentemente, tanto para IFNGR1 quanto para IFNGR2, os mesmos candidatos a ancestral foram encontrados. Estes resultados estão descritos na Tabela 1.

Foi possível identificar similaridade com receptores de IL10 de peixes, sugerindo, assim, que o receptor de IL10 possa fazer o papel de receptor do homólogo de IFN γ encontrado em peixes. Nesses Euteleostomi que não evoluíram para Tetrapoda, é plausível que IFN γ atue similarmente a IL10, enquanto que em Tetrapoda a via tornou-se muito específica para IFN γ .

Tabela 1: Resultado de PSI-BLAST para IFNGR1 e IFNGR2 humano.

Descrição	Identidade	SelfScore	Confiança	Organismo	
B5X9M6_SALSA Interleukin-10 receptor beta chain	18.27	0.8068	0.5304	<u>Salmo salar</u>	} IFNGR1
Q800G2_TETNG Helical cytokine receptor CRFB4	20.83	0.8068	0.5496	<u>Tetraodon nigroviridis</u>	
C0H9L6_SALSA Interleukin-10 receptor beta chain	17.56	0.6362	0.4513	<u>Salmo salar</u>	
C0H9L6_SALSA Interleukin-10 receptor beta chain	22.04	0.7244	0.7991	<u>Salmo salar</u>	} IFNGR2
Q800E8_TETNG Cytokine receptor CRFB4	21.7	0.7244	0.7957	<u>Tetraodon nigroviridis</u>	
Q800G2_TETNG Helical cytokine receptor CRFB4	21.88	0.7244	0.7972	<u>Tetraodon nigroviridis</u>	

O mesmo teste foi realizado para IFN γ , entretanto, a molécula mais ancestral encontrada como resultado é o próprio IFN γ de Salmo salar.

3. Construção da sinalização associada ao Complexo TCR

Como mencionado no item 2.2.1 (em Introdução), os LT desempenham várias funções ao longo da resposta imunológica. O TCR é um complexo com proteínas integrais

de membrana e ele participa na ativação dessas células a partir do reconhecimento de um Ag a ele apresentado.

A resposta se inicia com a estimulação do TCR pelas moléculas do MHC presentes na membrana das APCs. Essas moléculas estão associadas ao Ag e funcionam como um adaptador para o reconhecimento dos mesmos. Os genes do MHC se encontram agrupados numa grande família gênica [92] e todas as proteínas que o compõe, tanto MHCI quanto MHCII, tiveram sua origem junto aos Euteleostomi. O reconhecimento antigênico induz uma série de sinais intracelulares, descritos na Figura 10, que vão resultar em proliferação e diferenciação celular, produção de citocinas, apoptose, dentre outros.

Figura 10: Via de sinalização associada ao TCR. O LCA dos genes utilizados na construção dessa via estão distribuídos e identificados pelos clados *Eukaryota*, clado 2; *Opisthokonta*, clado 3; *Metazoa*, clado 4; *Eumetazoa*, clado 5; *Bilateria*, clado 6; Deuterostomia, clado 7; Chordata, clado 8; Gnathostomata (em branco), clado 11; *Euteleostomi*, clado 13; *Amniota*, clado 17; *Mammalia*, clado 18; *Theria*, clado 19; *Eutheria*, clado 20. **Nota:** alguns clados intermediários não tiveram ocorrência (ex: *Sarcopterygii*, clado 12).

O TCR é composto por seis diferentes cadeias de proteínas. As moléculas CD3 (CD3 γ , CD3 δ , CD3 ϵ e CD247), estão localizadas lateralmente ao receptor principal e possuem um característico motivo de fosforilação de tirosina chamado *Immunoreceptor Tyrosine-based Activation Motifs* (ITAMs) [92]. O componente CD247 tem, assim como as moléculas do MHC, sua origem em Euteleostomi (em laranja, clado 13), entretanto os componentes CD3 γ , CD3 δ , CD3 ϵ se originaram em Amniota (em vermelho, clado 17), sugerindo um aperfeiçoamento do sistema na época do advento desses organismos. As cadeias TCR tem caudas curtas que atravessam a membrana sendo elas compostas por dois domínios Ig. Um deles é altamente variável e se conecta com o Ag. Durante o desenvolvimento do LT, há expressão do *Pre T-cell antigen receptor* [93], com origem em Eutheria (não mostrado na figura), entretanto essa molécula não está presente nos LT maduros. Isso sugere que a resposta associada ao estímulo desse complexo só se tornou efetiva, nos modos como se conhece atualmente, nesse clado, uma vez que é necessária a constituição completa do complexo para que haja estímulo e consequente resposta intracelular. A especificidade da ligação é adicionada alguns passos evolutivos depois, em Amniota, com a adição de CD4 e CD8 ao sistema.

A ativação do TCR acontece simultaneamente com a participação de uma série de receptores co-estimulatórios. A Figura 11 mostra quantitativamente o conteúdo integral de membrana em relação aos clados taxonômicos de Homo sapiens.

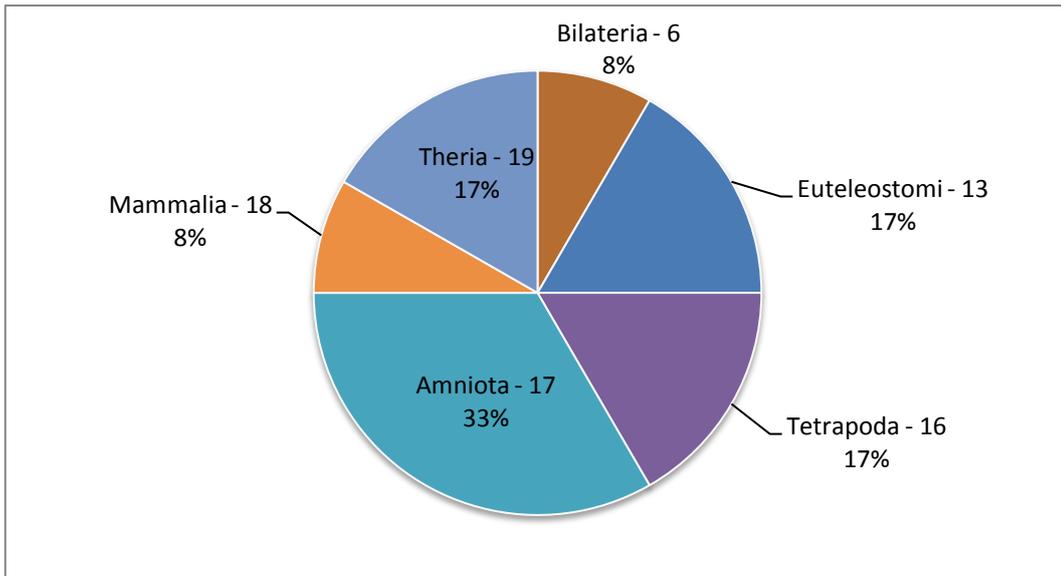


Figura 11: Receptores co-estimulatórios da sinalização TCR. Distribuição dos receptores co-estimulatórios de acordo com seu clado de origem ao longo da linhagem evolutiva de Homo sapiens. A ordem evolutiva segue em sentido horário a partir de Bilateria. **Nota:** As cores na área do gráfico foram escolhidas de forma arbitrária.

Separadamente foram analisados componentes de canais de cálcio devido as suas inúmeras subunidades (Figura 12, onde verifica-se intenso aparecimento de genes em Euteleostomi) e das proteínas não integrais de membrana. Esse último caso trata-se das proteínas Ras, conjunto de 6 proteínas que se encontram localizadas na superfície interna da membrana [94]. Cinco delas (HRAS, KRAS, NRAS, MRAS e RRAS2) tiveram sua origem juntamente com os organismos eucariotos e uma delas (RRAS) originou-se em Bilateria.

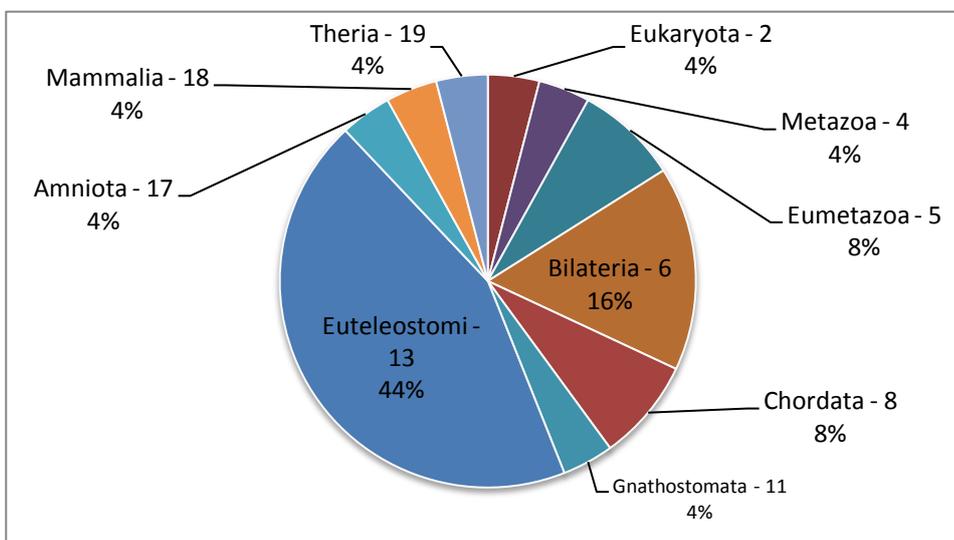


Figura 12: CaCn. Distribuição dos componentes de diferentes canais de cálcio de acordo com seu clado de origem ao longo da linhagem evolutiva de Homo sapiens. A ordem evolutiva segue em sentido horário a partir de Eukaryota. **Nota:** as cores na área do gráfico foram escolhidas de forma arbitrária.

A ativação do complexo TCR desencadeia a fosforilação das ITAMs. Sequencialmente são fosforiladas e assim ativadas Lck (ativada pela interação com o MHC, CD4 ou CD8 e CD28) e Fyn, proteínas responsáveis por dar início a sinalização [95]. Em seguida, ZAP70 funciona amplificando o sinal e esse conjunto de proteínas induzem a fosforilação de tirosinas em vários sítios das proteínas LAT e TRIM, levando a ativação da via ERK, JNK, NF- κ B e NFAT que terminam por induzir funções efetoras dos LT [96].

O receptor CD28 é um co-estimulador da ativação dos LT com origem em Tetrapoda (em amarelo, clado 16) e sua ação acarreta o aumento da produção de citocinas, proliferação e previne a morte celular. Ele pode se ligar a B7-1 ou B7-2 [97, 98], ambos presentes na membrana da APC e com origem em Amniota (em vermelho, clado 17), possibilitando a adesão inicial entre LT e APC. Por outro lado, o receptor CD45, o mais ancestral deles como será mostrado na Figura 10, modula a fosforilação das tirosina kinases Lck e Fyn.

Uma vez ativada, Lck ativa a fosforilação do complexo TCR e, conseqüentemente, Fyn e ZAP70 [96]. Esta última permanece ligada ao CD247 no complexo e induz a ativação de LAT, um adaptador integral de membrana que se liga a GADS, SLP76, ITK e Fyb estimulando PLC γ 1 que é responsável pela produção dos segundos mensageiros DAG e IP3 (representados como círculos identificados em vermelho) a partir da clivagem de PIP2 na membrana plasmática, e TRIM que contribui para a sinalização PI3K-Akt [99]. Os segundos mensageiros atuam das seguintes maneiras:

- DAG: ativa proteínas como PKC δ e as proteínas Ras [94].
- IP3: se liga no receptor IP3R presente na superfície do retículo endoplasmático e libera cálcio. Esse evento leva a ativação dos canais de cálcio que permite o influxo de desse íon para dentro da célula. O aumento nos níveis de cálcio intracelular ativam a Calcineurina (em azul escuro, clado 2), uma fosfatase que interrompe os efeitos inibitórios da Calmodulina (em azul escuro, clado 2) e cuja ativação leva a

desfosforilação de NFAT, permitindo assim a entrada da sinalização no núcleo [100].

Além disso, LAT, com LCA em Eutheria (em rosa, clado 19) também se liga a GRB2 e GADS para facilitar a montagem de complexos de sinalização macromoleculares que são necessários para a ativação do LT. A interação de LAT fosforilada com GRB2 promove um mecanismo onde esta proteína recruta SOS para a membrana e ativa Ras. Em consequência disso, uma série de serina/treonina kinases são ativadas [101], sendo elas: Raf1, MEK, ERK1 e ERK2 (sendo as duas últimas enzimas originadas em Eukaryota e responsáveis pela ativação das MAPKs). A ativação das ERKs, entretanto, podem ser inibidas por VHR em sua forma fosforilada por ZAP70 [102], que também ativa p38 (Eukaryota, em azul escuro, clado 2) e JNK (Metazoa, em verde escuro, clado 4).

Em outro ponto central da via, Vav, de origem em Opisthokonta (em azul claro, clado 3), atua fosforilando e ativando Rac, uma proteína que se liga a GTP e efetua a mediação da ativação de MEKK1, e também participa como ativadora de p38 através da ativação de MKK3 e MKK6. JNK, por sua vez, é ativada através da ativação de MKK4 e MKK7. De um modo geral, esses são membros ancestrais da via, com LCAs que variam de Eukaryota à Chordata. Essas MAPKs fosforilam os fatores de transcrição envolvidos na formação do complexo Jun-Fos, assim como NFAT, localizados no núcleo (representado pelo círculo cinza de linha dupla). As ERKs, JNK e p38 fosforilam os fatores de transcrição Elk1, c-Jun e ATF2, também localizados no núcleo.

Outro fator de transcrição importante é o NF- κ B, juntamente com IL-2. Os membros de NF- κ B controlam a expressão de vários genes envolvidos na reação inflamatória, apoptose a resposta imunológica e sua ativação depende da estimulação do TCR e CD28 [103]. NF- κ B tem seus componentes com LCAs em Metazoa, Eumetazoa e

Bilateria (em diferentes tons de verde, clados 4, 5, e 6, respectivamente) e um conjunto de proteínas regula a sua ação: CARD11, MALT1 e BCL10. Essas proteínas agem conjuntamente na regulação da atividade de NF- κ B através da ativação das IKKs, complexo que regula a atividade do mesmo a partir da fosforilação de I κ B [104], o que leva a sua ubiquitinação. Livre da associação com I κ B NF- κ B pode se mover para o núcleo e ativar a transcrição. Ainda é importante ressaltar que a ativação do complexo TCR também leva ao rearranjo do citoesqueleto, como indicado à esquerda na figura, através da ativação de proteínas que se ligam a GTP: Rac, PAK e CDC42.

A regulação negativa da sinalização é necessária para que a hiperativação do sistema imune não ocorra e para que o sistema retorne a um estado ótimo pós ativação. Quando SIT interage com SHP2 funciona como regulador negativo da sinalização [104]. Em LT em repouso, PAG, um adaptador transmembrana é fosforilado e em associação com CSK resultando também na inibição de TCR. Esse tipo de inibição é rapidamente suprimido assim que há estímulo do receptor pelo complexo antígeno-MHC [105]. Ainda há a ação inibitória promovida por CTCL4 (em vermelho, clado 17). Essa molécula tem grande afinidade por B7-1 e B7-2, tendo inclusive o mesmo LCA Amniota, comparando-se com o receptor Tetrapoda CD28 e fica na superfície da membrana onde interage com SHP2 impedindo a fosforilação de TCR, mas quando há necessidade de rápida ativação do LT, CTCL4 é rapidamente endocitada [106]. Na Figura 10, o endossomo está representado por um círculo cinza de linha tracejada.

A Figura 10 mostra que os componentes dos compartimentos celulares citosol e núcleo são mais ancestrais em relação àqueles presentes na membrana. No citosol, ainda há algumas adições recentes em Euteleostomi e Amniota, ainda assim 89% dessas proteínas tiveram sua origem entre Eukaryota e Bilateria (Figura 13). Isso pode ser

justificado quando nota-se que parte dessas proteínas participam de outros processos celulares. Sugere-se, então, que a transdução de sinal foi adequada e associada à resposta gerada pela ativação do complexo TCR durante sua construção ao longo do processo evolutivo.

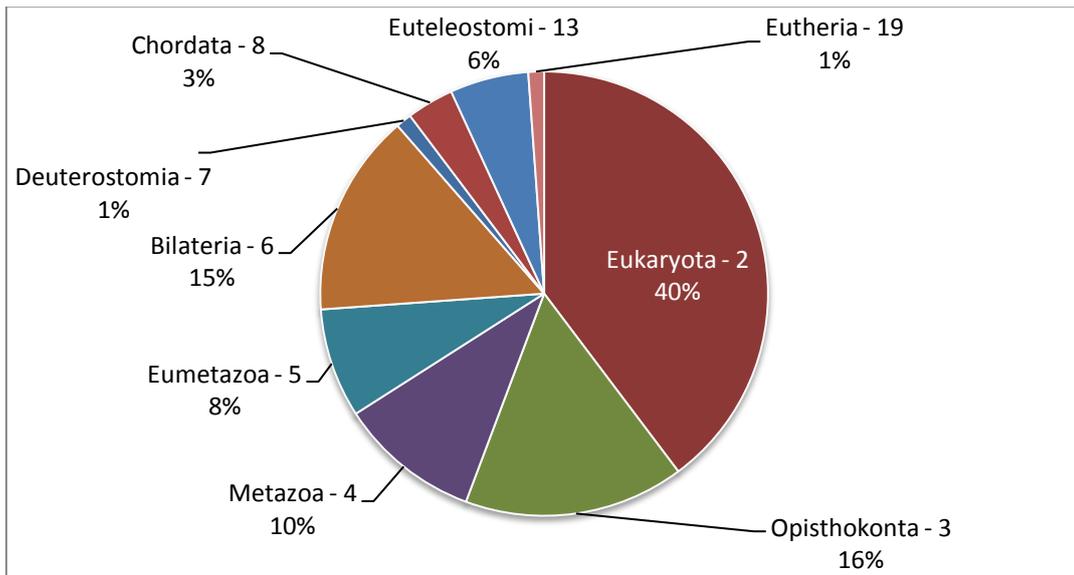


Figura 13: Proteínas do citosol na sinalização TCR. Distribuição componentes citoplasmáticos da transdução de sinal de acordo com seu clado de origem ao longo da linhagem evolutiva de Homo sapiens. A ordem evolutiva segue em sentido horário a partir de Eukaryota. **Nota:** As cores na área do gráfico foram escolhidas de forma arbitrária.

Os componentes nucleares se distribuem em três clados, sendo eles Metazoa, Eumetazoa e Bilateria, como indicado na Figura 14.

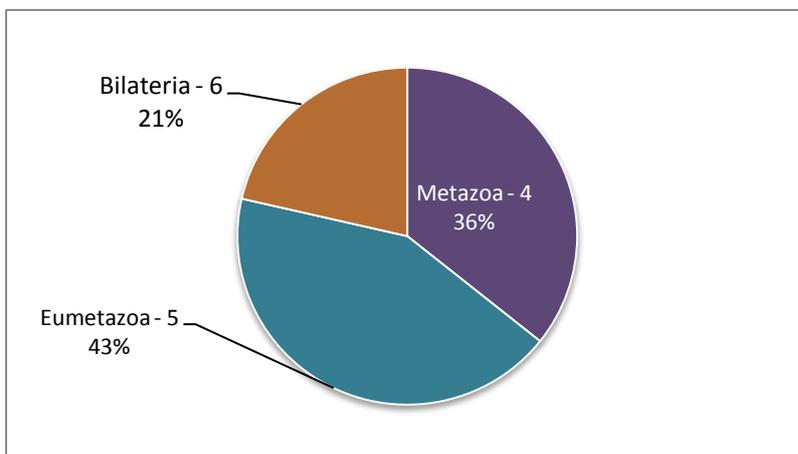


Figura 14: Proteínas do núcleo na sinalização TCR. Distribuição componentes nucleares de acordo com seu clado de origem ao longo da linhagem evolutiva de Homo sapiens. A ordem evolutiva segue em sentido horário a partir de Metazoa. **Nota:** As cores na área do gráfico foram escolhidas de forma arbitrária.

Por fim, analisando-se simultaneamente a composição proteica por compartimento celular, é possível notar um pico de genes em Amniota para membrana, uma explosão de genes entre os clados Eukaryota e Bilateria para citosol, com poucas adoções em Eutelostomi e, por fim, todas as adições nucleares acontecendo nos clados Metazoa, Eumetazoa e Bilateria. Isso sugere que a transdução de sinal de associou a estimulação do complexo TCR que foi construído tardiamente ao longo do processo evolutivo e os promotores de genes ancestrais relativos à esse tipo de resposta se tornaram responsivos a esse estímulo.

4. Construção da sinalização associada ao Complexo BCR

Outro importante cenário que buscou-se retratar, ao lado da ativação do LT, é aquele que produz a sinalização associada ao complexo do receptor da célula B (BCR). O BCR é fundamental para o funcionamento dos LB, tanto para seu desenvolvimento como para sua atividade. Esse receptor é composto por imunoglobulinas de membrana, como mencionado no item 4.2 (em Introdução), e pelas moléculas associadas CD79A e CD79B (Figura 15), que efetuam a transdução do sinal e têm seu LCA em Euteleostomi (em laranja, clado 13).

Assim como o TCR, este receptor precisa de receptores auxiliares que modulam a resposta. CD45 regula a transdução de sinal através da fosforilação e PCL γ 2 é responsável pelo aumento da atividade do receptor [107]. Esses receptores tem sua origem em Bilateria e Eutheria, respectivamente, portanto aparentemente são recrutados a participar de um mecanismo mais recente. Ainda há o complexo de co-receptores CD19, envolvido na transdução do sinal, CD21, ligante de partículas antigênicas opsonizadas (ambos posteriores à época dos Euteleostomi) e CD81 (este mais ancestral), de função

pouco clara [108]. A Tabela 2 representa a distribuição desses receptores por clado taxonômico. Esta parte deste sistema concentra as proteínas mais recentes (mostradas em tabela para melhor clareza).

Tabela 2: Distribuição de componentes integrais de membrana de acordo com seu clado de origem em relação a linhagem evolutiva de Homo sapiens.

Receptor	Abreviação	LCA	Level
B-cell antigen receptor complex-associated protein alpha chain	CD79a	Euteleostomi	13
B-cell antigen receptor complex-associated protein beta chain	CD79b	Euteleostomi	13
B-lymphocyte antigen CD19	CD19	Eutheria	20
Complement receptor type 2	CD21	Amniota	17
B-cell receptor CD22	CD22	Euteleostomi	13
Receptor-type tyrosine-protein phosphatase C	CD45	Bilateria	6
Low affinity immunoglobulin gamma Fc region receptor II-b	FcγRIIB	Eutheria	20

A Figura 15 representa por sua vez os componentes não integrais de membrana por clado taxonômico. Nesse caso, proteínas mais ancestrais predominam.

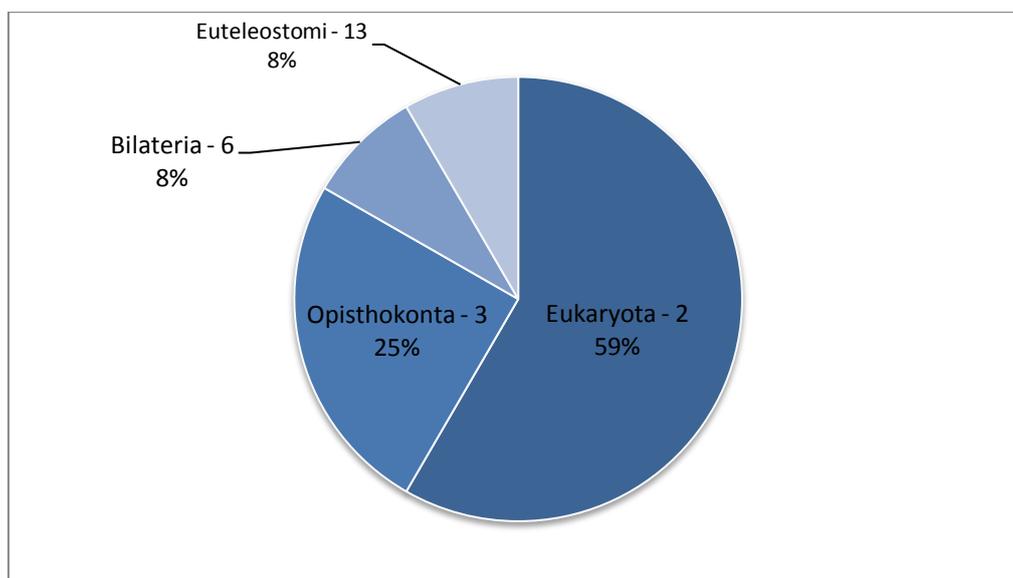


Figura 15: Componentes não-integrais de membrana da sinalização BCR. Distribuição de componentes não-integrais de membrana de acordo com seu clado de origem ao longo da linhagem evolutiva de Homo sapiens. A ordem evolutiva segue em sentido horário a partir de Eukaryota. Nota: as cores na área do gráfico foram escolhidas de forma arbitrária.

Uma vez que BCR é estimulado, como mostrado na Figura 16, três tipos de tirosina cinase são ativadas: SYK, BLK e Fyn. Uma vez ativada, SYK fosforila GAB e Bap31 e ambas vão atuar na ativação de PI3K. BLK e Fyn atuam indiretamente na ativação de Ras, Vav e MAPks [109]. A transdução de sinal é iniciada com a ativação de Fyn, resultando na fosforilação de CD79a e CD79b (ambos mais recentes, LCA em Euteleostomi), PI3K, Vav, e PCL γ 2. PCL γ 2 é importante na produção de IP3 [110] e na regulação do cálcio intracelular. Sendo responsável pelo aumento da atividade celular, essa molécula permite a conversão de PIP2 em IP3 e DAG (círculos assinalados em vermelho) através do adaptador hDAPP1. Assim, PKC é ativada e fosforila INK via CARD10, MALT1 e BCL10, resultando na fosforilação de I κ B e translocação de NF- κ B para o núcleo. IP3 se liga ao IP3R, localizado na membrana do retículo endoplasmático (na figura representado pela forma em linha cinza tracejada) e estimula a liberação de cálcio intracelular [111]. SLP65 está envolvida na regulação do estado fosforilado de PCL γ 2 e sua localização celular. Uma vez que BCR é estimulado, essa molécula é fosforilada por Syk. Então, BTK é recrutada e ativa PCL γ 2, levando à produção de IP3 e à mobilização de cálcio. A elevação do cálcio intracelular resulta no aumento de NFAT e facilita sua translocação para o núcleo através de sua desfosforilação pela Calcineurina [112].

Figura 16: Via de sinalização associada ao BCR. O LCA dos genes utilizados na construção dessa via estão distribuídos e identificados pelos clados **Eukaryota**, clado 2; **Opisthokonta**, clado 3; **Metazoa**, clado 4; **Eumetazoa**, clado 5; **Bilateria**, clado 6; **Chordata**, clado 8; **Gnathostomata** (em branco), clado 11; **Euteleostomi**, clado 13; **Amniota**, clado 17; **Eutheria**, clado 20. **Nota:** alguns clados intermediários não tiveram ocorrência (ex: **Sarcopterygii**, clado 12).

A ligação de PI3K a CD19 ainda ativa PIP3, e Vav ativa Rac, que em seguida ativa CDC42 (proteína de origem recente, à esquerda, em vermelho, clado 17) que é ativadora de MEKK e que fosforila e ativa JNK. Então, essa molécula entra no núcleo e ativa o fator de transcrição c-Jun. Além disso, quando ativada, PI3K ainda leva à ativação de PDK e consequentemente Akt. Esta última, por sua vez, fosforila caspases e transportadores de glicose, além de mTOR, IKK, NF- κ B e GSK3 [113] promovendo um efeito anti-apoptótico.

A ativação de BCR também resulta na indução da sinalização do complexo Ras (via SHC, GRB2 e SOS) e Rap1, GTPases que levam à ativação das MAPK cinases, incluindo ERK, JNK e p38 [114]. O efeito da produção de segundos mensageiros e ativação de proteínas sinalizadoras combinam a regulação de vários fatores de transcrição, que vão mediar a transcrição no LB.

A sinalização derivada da ativação do BRC conduz a proliferação, diferenciação e apoptose. O tipo de resposta será determinado pelo estado do LB, natureza do Ag, afinidade da interação com o Ag e as condições do ambiente celular.

Com respeito à ancestralidade dessa resposta, um total de 92 componentes da transdução de sinal no citosol (77%) se origina entre os clados Eukaryota e Metazoa, e esse valor aumenta para 89% adicionando o clado Bilateria (Figura 17). Enquanto isso, no núcleo, a totalidade dos componentes evoluiu entre os clados Metazoa e Bilateria (Figura 18).

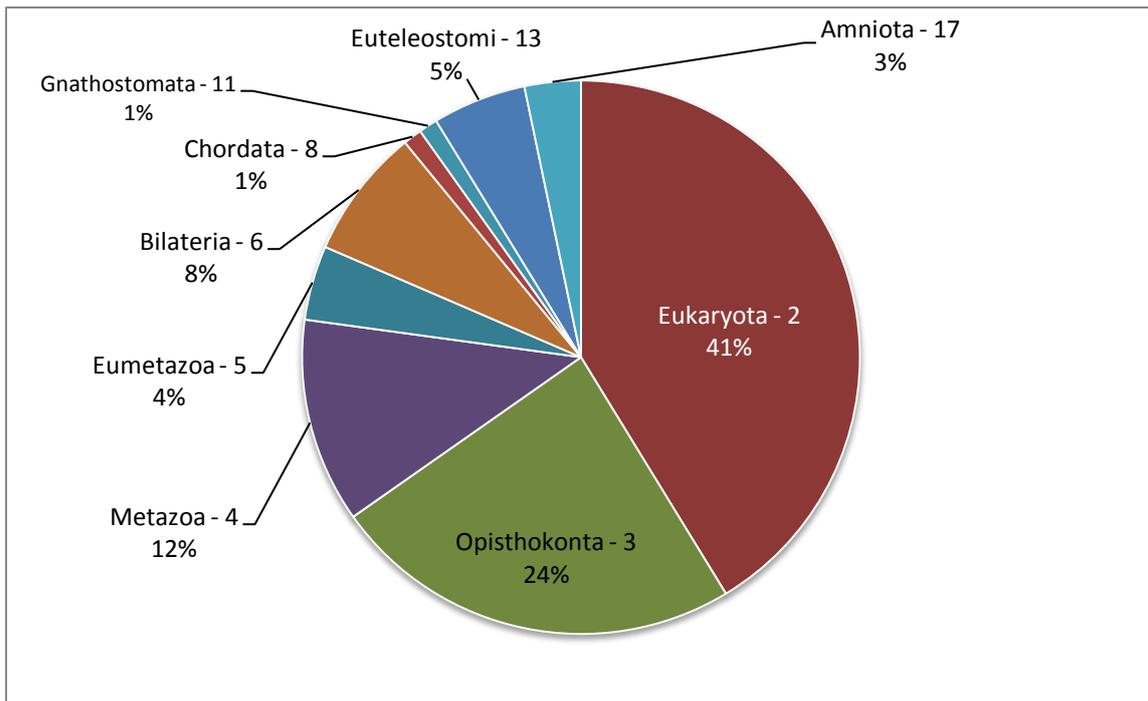


Figura 17: Proteínas do citosol na sinalização BCR. Distribuição dos componentes citoplasmáticos da transdução de sinal de acordo com seu clado de origem ao longo da linhagem evolutiva de Homo sapiens. A ordem evolutiva segue em sentido horário a partir de Eukaryota. Nota: as cores na área do gráfico foram escolhidas de forma arbitrária.

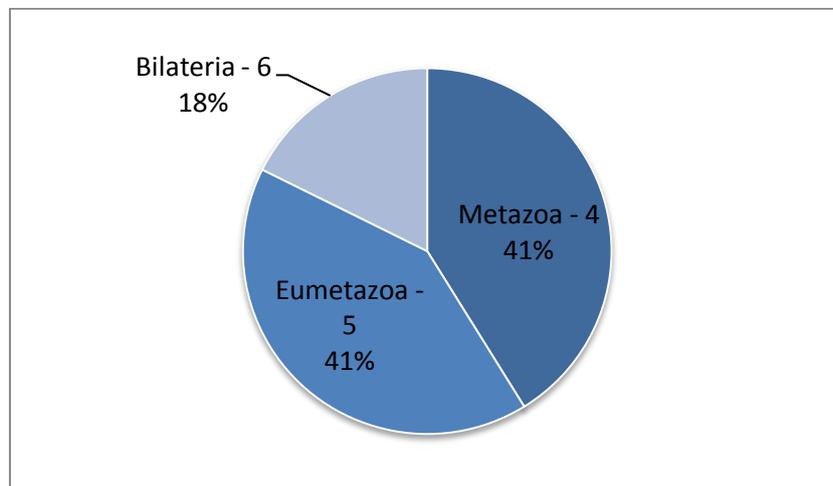


Figura 18: Proteínas do núcleo na sinalização BCR. Distribuição dos componentes nucleares de acordo com seu clado de origem ao longo da linhagem evolutiva de Homo sapiens. Foram contabilizados apenas componentes exclusivamente nucleares. A ordem evolutiva segue em sentido horário a partir de Eukaryota. Nota: as cores na área do gráfico foram escolhidas de forma arbitrária.

A partir disso, torna-se possível sugerir que a construção da transdução de sinal associada ao BCR incorporou muitos elementos pré-existentes ao construir-se a resposta à ativação do BCR, ao longo do processo evolutivo.

Em resumo, é possível notar na membrana uma grande adição de genes em Amniota, além das adições de CD79A e CD79B em Euteleostomi, uma explosão de genes entre os clados Eukaryota e Bilateria para citosol, com poucas adoções em Eutelostomi, indicando um aperfeiçoamento do sistema e, por fim, todas as adições nucleares acontecendo entre os clados Metazoa e Bilateria.

5. Células *Natural Killer* e receptores semelhantes a *Toll*

Dois outros aspectos relevantes para retratar o cenário da evolução do sistema imune são a origem dos principais marcadores de células NK e dos receptores semelhantes a *Toll*. Os marcadores de NK são uma fonte de informação para revelar a origem mais específica da célula e mesmo por inferência deveria compreender proteínas mais recentes, pois não focamos em uma rede de sinalização como nos casos já expostos. Por outro lado, como receptores semelhantes a *Toll* são encontrados em muitos organismos, o estudo de época de origem da forma similar à humana pode nos dar uma ideia de quando os mecanismos dos quais participam vão se estabelecendo. Ambos são aspectos do sistema imune inato.

5.1 Células *Natural Killer*

As células Natural Killer (NK) são um tipo de linfócito componente majoritariamente do sistema imune inato. Essas células desempenham uma importante função no combate a infecções virais e destruição de células tumorais e são ativadas por

IFN tipo I e outras citocinas produzidas por macrófagos. Além disso, as NK funcionam como uma ponte entre as imunidades inata e adaptativa quando secretam citocinas que estimulam e guiam a resposta de outros agentes do sistema imune e dos linfócitos da resposta imune adaptativa.

Intencionando inferir o momento na evolução onde a resposta efetiva, nos moldes como conhecemos atualmente, foi formada, foi feita a busca pelos identificadores *UniProt* dos marcadores de NK para o agrupamento de seus genes homólogos através da ferramenta *SeedServer*. Com a distribuição taxonômica dos genes homólogos foi encontrado o LCA dessas proteínas e o resultado para os 22 marcadores escolhidos de acordo com a literatura está representado na figura abaixo:

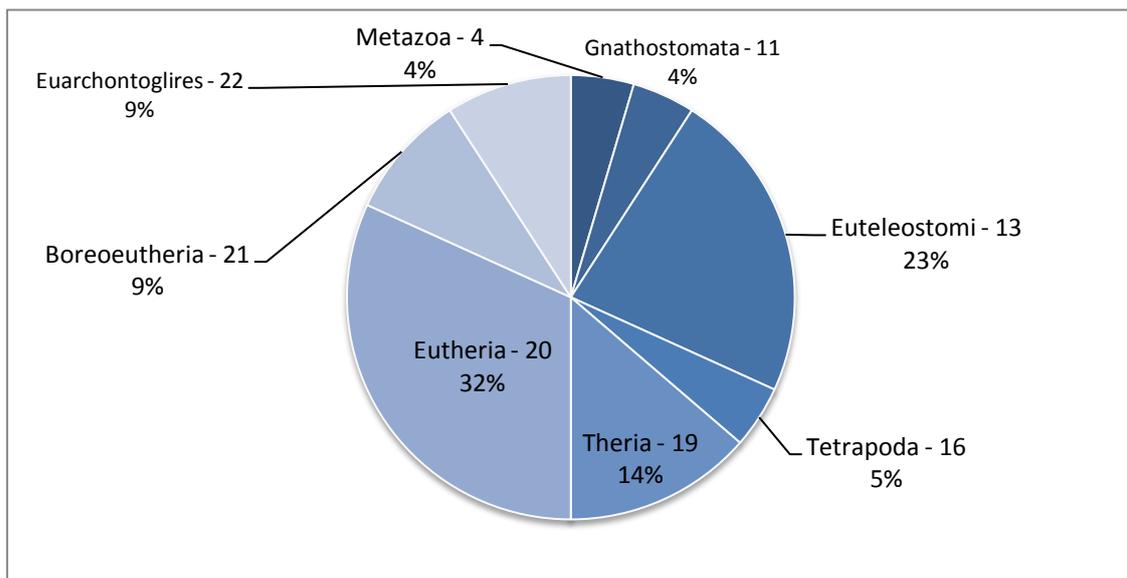


Figura 19: Marcadores de superfície de NK. Distribuição dos marcadores de célula NK de acordo com seu clado de origem ao longo da linhagem evolutiva de *Homo sapiens*. A ordem evolutiva segue em sentido horário a partir de Metazoa. Nota: as cores na área do gráfico foram escolhidas de forma arbitrária.

A distribuição da origem dos marcadores de NK ao longo dos cladogramas taxonômicos humanos mostra um total de 92% dos mesmos se originando a partir da origem dos ossos, em Euteleostomi. Isso sugere que esse tipo de resposta mediada pelas NK são uma

adição recente ao sistema imune e funcionam como um aperfeiçoamento não só do sistema imune inato, mas como da integração entre este com a imunidade adaptativa.

5.2 Receptores semelhantes a *Toll*

Os receptores semelhantes a Toll (TLR - *Toll-like receptor*) são um grupo de proteínas transmembrana expressas pelas células do sistema imune inato. São capazes de reconhecer padrões que detectam os padrões PAMPs [116] e ativam vias de sinalização que levam às respostas imunes e inflamatórias.

Em humanos, essa família é formada por 10 diferentes receptores formados por: uma cauda citoplasmática homóloga a IL1R, o domínio TIR, uma região transmembrana e um domínio extracelular rico em leucinas [117]. A distribuição dos TLRs humanos de acordo com sua origem ao longo da linhagem está representada na Figura 20.

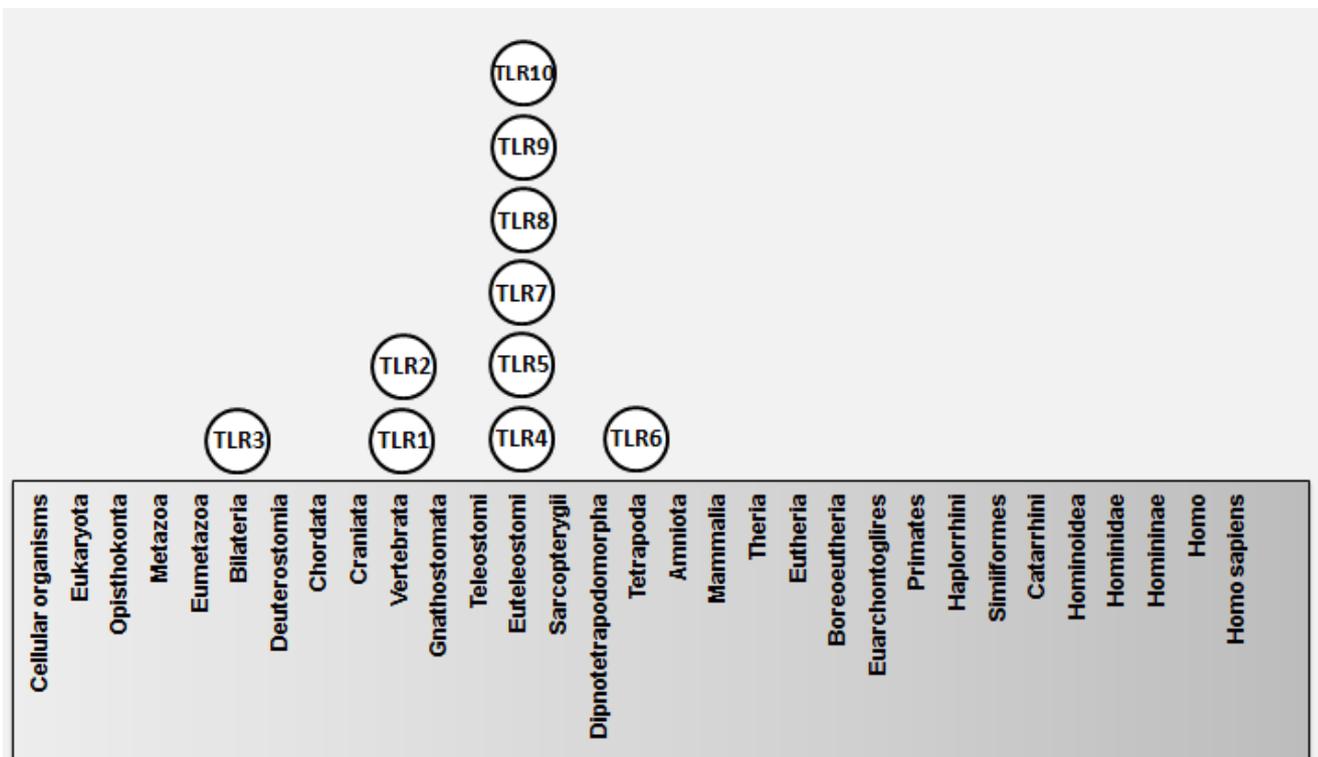


Figura 20: Distribuição dos TLRs de acordo com seu clado de origem ao longo da linhagem evolutiva de Homo sapiens.

De um modo geral, a distribuição principal desses TLRs ocorre em clados recentes da evolução de *H. sapiens*. Contudo, o LCA ancestral de TLR3 se destaca como o mais ancestral e, no sentido de refinar a reconstrução da história evolutiva dessas moléculas, estimando as taxas de evolução molecular, uma árvore filogenética foi construída com *Neighbor-Joining* (*bootstrap*: 10000), representada na Figura 21A, e uma com *Maximum Likelihood* (*bootstrap*: 100), mostrada na Figura 21B. Para a construção de ambas foram utilizadas as sequências completas dos 10 receptores e IL1R foi utilizado como grupo externo. Ambas concordam na distribuição aproximada dos ramos.

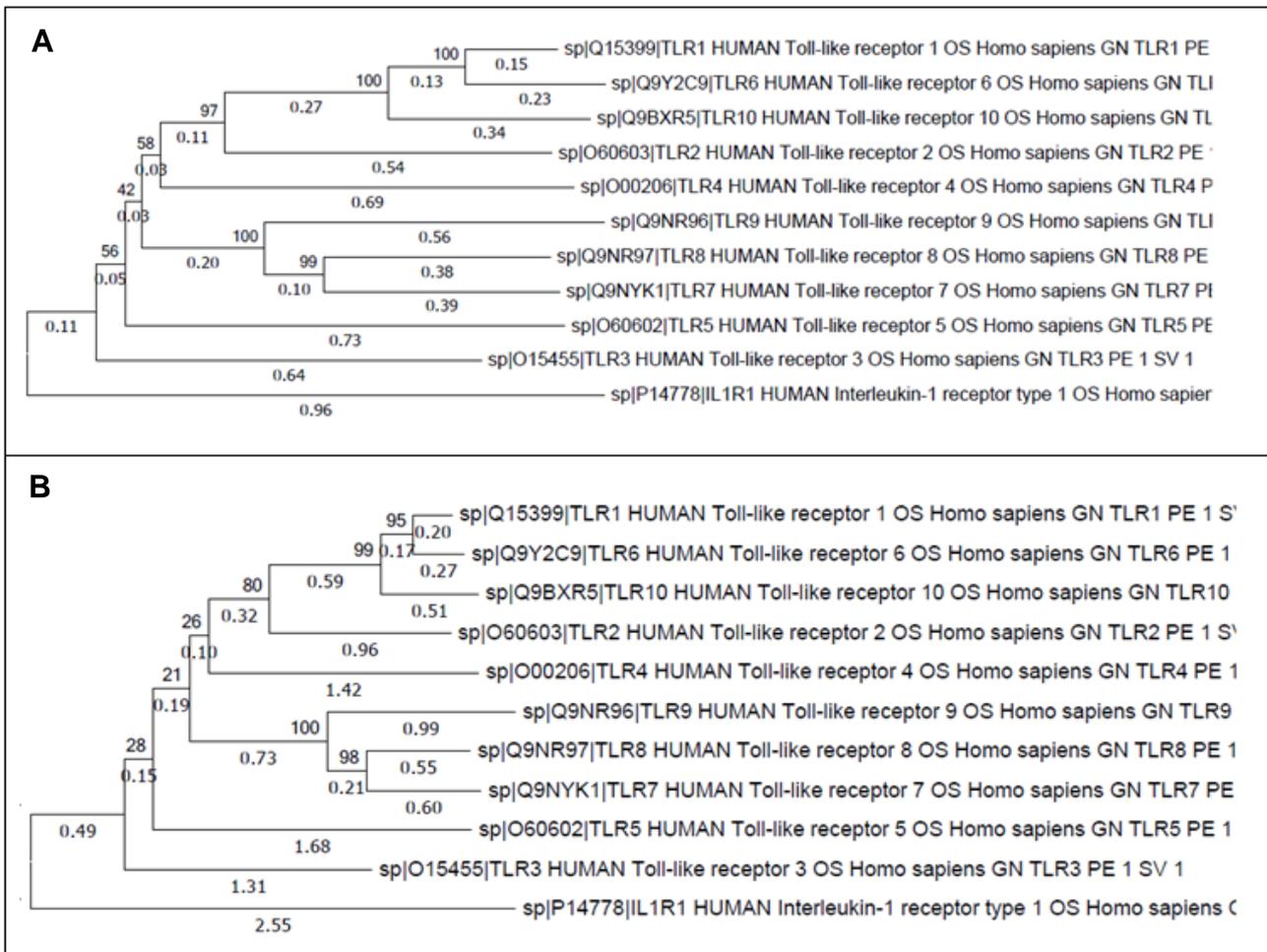


Figura 21: Árvores filogenéticas para os TLR humanos. O número sobre cada ramo representa o *bootstrap* e o número sob cada ramo representa o comprimento desse ramo. **A:** Árvore construída com *Neighbor-Joining*. **B:** Árvore construída com *Maximum Likelihood*.

As árvores mostram um agrupamento de TLR7, TLR8 e TLR9 que pode ser justificado pelo fato de todas essas serem proteínas endossomais e ainda com o mesmo LCA, em Euteleostomi. Ainda, há um bom agrupamento de TLR1, TLR6, TLR10 e TLR2. Os LCAs estão distribuídos entre os clados Vertebrata, Euteleostomi e Tetrapoda, sugerindo que a primeira expansão teria originado TRL1 e TRL2 que deram origem a esse grupo (TRL6 e TRL10 aparecem mais próximos de TRL1). Essas quatro moléculas apresentam características em comum relacionadas às vias de transdução de sinal e os genes ativados pelas mesmas.

Os demais receptores estão relativamente isolados e TLR3 aparece, como esperado, na base da árvore. Entretanto, o mesmo acontece com TLR 5. Sabendo-se que o LCA de TLR5 ocorre em Euteleostomi, foi feita uma análise de PSI-BLAST no sentido de buscar por possíveis ancestrais TLR3. O resultado para esta busca está descrito na tabela abaixo:

Tabela 3: Resultado de PSI-Blast para TLR5 humano.

Descrição	Identidade	SelfScore	Confiança	Organismo
E6ZJE2_DICLA Toll-like receptor 3	24.51	0.9564	0.8113	<u>Dicentrarchus labrax</u>
B6ZNC8_ROULE Toll-like receptor 3	26.29	0.9564	0.8093	<u>Rousettus leschenaultii</u>
Q7TNI8_RAT Toll-like receptor 3	25.14	0.9564	0.8060	<u>Rattus norvegicus</u>
F1CKK3_BUBBU Toll-like receptor 3	26.8	0.9564	0.7991	<u>Bubalus bubalis</u>
A2SWM4_LOXAF Toll-like receptor 3	26.58	0.9564	0.7955	<u>Loxodonta africana</u>
Q76CT9_PAROL Toll-like receptor 3	25.08	0.9564	0.7952	<u>Paralichthys olivaceus</u>

O resultado do PSI-BLAST sugere fortemente que TLR5 pode ter se originado a partir de TLR3 nos Euteleostomi, já que os homólogos distantes mais próximos são de peixes, como o D. labrax. Essa ideia ainda é reforçada devido a presença de outros

mamíferos na tabela, como R. leschenaultii e R. norvegicus. A hipótese alternativa é que o TLR5 teria se perdido em organismos Bilateria que não evoluíram até Vertebrata, embora menos provável. Nesse caso, e observando-se que o ramo que leva a TLR5 é longo, é notável que genes possivelmente originados em clados mais recentes pareçam mais ancestrais, talvez pela menor quantidade de divergências que ocorreram até que obtivesse a função fixada.

A ativação dos TLRs ocorre após a ligação de seu ligante cognato e duas vias principais são ativadas: (i) a via principal que leva a ativação do fator de transcrição NF- κ B e (ii) a via MAPKs-p38-JNK [117]. Estas vias estão associadas às sinalizações discutidas anteriormente, então se sabe que a transdução de sinal, também para os diferentes TLRs, é bastante ancestral em termos gerais.

X. DISCUSSÃO

Ao longo da evolução dos sistemas, novos genes surgiram na natureza e para demonstrar a origem dos genes, as vias de sinalização do sistema imune revelam-se um modelo adequado. O sistema imune evoluiu de modo a formar uma rede de células, tecidos e órgãos que protegem o organismo contra microrganismos patogênicos, células tumorais e de outros antígenos. A origem evolutiva dos subsistemas que compõe a imunidade envolve o advento de receptores, elementos de transdução de sinal e também efetores nucleares.

Além disso, há importantes aspectos na evolução que tangem o sistema imune, como os recentes desenvolvimentos de mecanismos de resistência a infecções virais. Assim, com o objetivo de delinear a construção deste sistema, o presente trabalho buscou a partir de curadoria manual as sequências de aminoácidos dos representantes humanos e estas foram usadas como *Seeds* para procurar homólogos. Ao examinar a distribuição taxonômica desses homólogos, através do algoritmo LCA, foi possível retratar os clados, na linhagem humana, onde cada proteína ocorreu pela primeira vez.

O último ancestral comum de 385 genes envolvidos em diversos subsistemas do sistema imune foi inferido. Estes genes foram escolhidos devido a sua importância central nas respostas imunológicas. Um detalhe importante mostra que a distribuição total dos genes do sistema imune de acordo com seu LCA concorda com a mesma distribuição dos genes totais de *H. sapiens*, como mostrado na Figura 1.

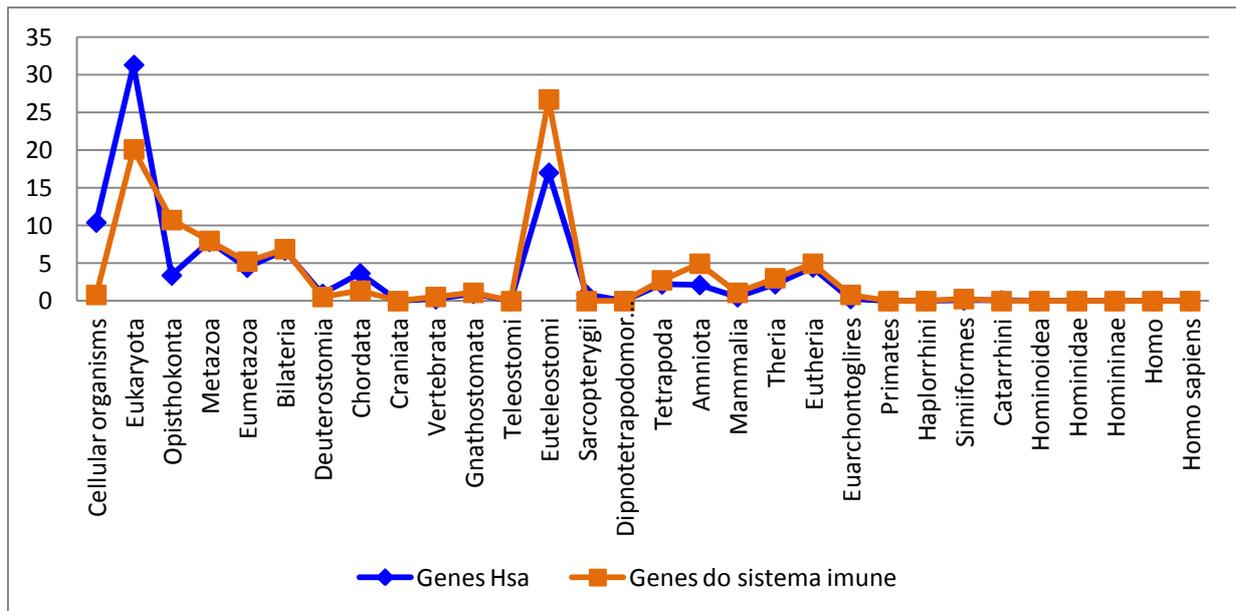


Figura 1: Padrão de surgimento total de genes. Distribuição total dos genes humanos (curva em azul) e dos genes do sistema imune estudados (curva em laranja) de acordo com seu clado de origem ao longo da linhagem evolutiva de Homo sapiens.

Sendo assim, o presente trabalho permitiu a visualização da formação de alguns sistemas centrais nas respostas imunes. Os IFNs estão distribuídos entre os clados Euteleostomi e Eutheria, sendo que todos os membros da família do IFN α (parálogos) se originam em Euteleostomi e todos os membros componentes do grupo de IFNs tipo III em Amniota. O IFN ω tem a origem mais tardia, em Eutheria, sugerindo que mesmo estabelecido, o sistema continuou sendo aperfeiçoado. Os receptores cognatos dessas citocinas são distribuídos entre os clados Euteleostomi e Tetrapoda. As cadeias de IFNs tipo I tem ancestralidade diferente, sendo IFNAR1 em Euteleostomi e IFNAR2 em Tetrapoda. Vale lembrar que o LCA de IFNAR2 pode ser mais ancestral, devido ao número defasado de genomas completos disponíveis, como, por exemplo, Sarcopterygii ter apenas um representante e mal anotado (celacanto). Assim, esta subunidade foi mostrada como tendo sua origem em Tetrapoda, mas é possível que seja encontrada em Sarcopterygii. Não obstante, essa cadeia não foi detectada nos vários genomas de peixe

disponíveis e o clado Tetrapoda está apenas a dois passos evolutivos de Euteleostomi. Considerando que os dois clados que se interpõe são Sarcopterygii e Dipnotetrapodomorpha, vale reforçar a defasagem de genomas completos disponíveis para espécies taxonomicamente agrupadas nesses clados. Nossos dados sugerem que as cadeias realmente se originam em épocas distintas, todavia podendo ser mais próximas do que já indicado.

As cadeias do receptor estão diretamente relacionadas com uma sinalização JAK-STAT. Essas proteínas trabalham ligadas ao receptor e levam o sinal ao núcleo. Mostrou-se que a maior parte dessas proteínas tem sua origem em congruência com a origem do sinal na membrana, diferentemente dos demais tipos de sinalização que não tem essa característica de conectar diretamente a membrana ao núcleo, como é feito pelas STATs. É notável como a transdução de sinal em células B não apresenta muitas proteínas da época da origem dos principais elementos da membrana.

Assim, como esperado, a origem de ligantes de receptores cognatos é separada no tempo. Todavia, contrariando o esperado, a origem de IFN γ antecede a origem de seu receptor. A partir de dados da literatura, sabe-se que essa citocina se liga exclusivamente ao seu receptor. É plausível que genes, de um modo geral, tenham origem em outros pré-existentes, embora atualmente se saiba que também é possível a origem gênica a partir de regiões não-codificadoras [118, 119].

Então, foi feito um estudo com buscas com PSI-BLAST com o intuito de encontrar possíveis ancestrais mais remotos para as cadeias IFNGR1 e IFNGR2. O fato de ambas as cadeias do receptor de IFN γ apontarem para duas cadeias de receptor de IL-10 reforça que teria havido esta origem e sugere estudos em peixes para verificar se nesses

organismos há ligação de um homólogo de IFN γ a receptores de IL-10 dessas espécies, já que não encontramos homólogos para as cadeias do receptor cognato.

A evolução dos IFNs e seus receptores é precedida pela existência de uma gama de IRFs. Em outras palavras, os IFNs supostamente evoluíram já com os reguladores do regulon presentes. Contudo, uma vez que o regulon antecede os efetadores da resposta, trabalhos futuros serão necessários para inferir possíveis funções dessas moléculas anteriormente à evolução do sistema.

É importante notar que os filtros utilizados sempre propagam a anotação a partir da *Seed* humana e, portanto, somente vão considerar a origem a partir do ponto onde a similaridade com a proteína atual é significativa. Portanto, é possível que proteínas sob uma intensa plasticidade façam parte do sistema mais antigo e sejam fixadas no momento em que sua função atual é obtida.

O complexo receptor de células T (TCR) aparece tardiamente em Eutheria, uma vez que este é o LCA encontrado para o precursor desse receptor em LT imaturos. O co-receptor CD3 tem as suas subunidades que aparecem em Euteleostomi e Amniota.

Os receptores CD4 e CD8 têm ambos seu LCA em Amniota e são imprescindíveis para que a resposta seja efetuada, mesmo não sendo componentes do TCR. O complexo liga-se a Ags virais e tumorais, que estão associados com MHC classe I, enquanto agentes patogênicos bacterianos e outros Ags estão associados a MHC de classe II.

Os componentes para ambos MHC surgiram em Euteleostomi, entretanto dados da literatura mostram a presença de MHC classe I em invertebrados [paper miguel 2], contudo é interessante notar que o estabelecimento e estabilidade da resposta, como observada em humanos, se faz em Euteleostomi. Além disso, notou-se que o alinhamento de MHC classe I da espécie D. rerio (peixe-zebra) mostra similaridade com MHC de

tubarão. Isso sugere que o estabelecimento do MHC classe I, como é observado em humanos, possa ter surgido em Euteleostomi, todavia a primeira origem teria sido anterior, em Gnathostomata, mas com uma sequência não suficientemente similar à humana para ser agrupado com ela. Com o constante aumento do número de organismos sequenciados, estudos futuros serão capazes de elucidar essa questão. Por fim, é interessante notar o possível surgimento do MHC classe I em congruência com o principal efetor da resposta imune anti-viral, IFN α , tendo ambos sua origem em Euteleostomi.

Quanto ao BCR, o heterodímero Ig-alfa/Ig-beta apareceu juntamente com a origem dos ossos, em Euteleostomi. Muitos outros receptores transmembrana que são conhecidos por modular elementos específicos da sinalização BCR surgiram em diferentes clados: CD45 em Bilateria e FcGRIIB em Eutheria; há também um complexo de correceptores constituído por CD21, CD19 originados em Amniota e Eutheria, respectivamente. Esse quadro mostra aquilo que foi observado nos demais sistemas estudados: a constituição da sinalização na membrana é sempre a última a se estabelecer.

Tratando-se dos marcadores de superfície de NK, notou-se contribuição bastante recente, em Boreoeutheria e Euarchontoglires. Esse quadro mostra aquilo que foi observado nos demais sistemas estudados: a constituição da sinalização na membrana é sempre a última a se estabelecer. E, ainda, na resposta imune inata, foi observada a ampla distribuição dos TLRs ao longo dos clados Bilateria e Euarchontoglires (Figura 20, em Resultados). É importante notar que a origem do TLR5 diverge de sua posição na filogenia. Isso sugere que, se uma cópia de expansão é mais recente mas leva poucos passos para subfuncionalizar, ela pode aparecer na filogenia como uma proteína mais ancestral. Uma possibilidade alternativa para justificar esse quadro seria organismos

Bilateria que não evoluíram até Euteleostomi terem perdido a expansão (TLR5) de TLR3, que só se fixou em algumas espécies descendentes de Bilateria. Entretanto, essa seria uma alternativa menos favorável.

A abordagem utilizada para o agrupamento dos homólogos usa ferramentas de propagação a partir de uma sequência bem conhecida, a *Seed*. Em vista disso, o presente trabalho não se resume apenas na busca pelo último ancestral comum a uma proteína, mas desde qual momento na evolução essa proteína se torna suficientemente similar à *Seed* para que seja agrupada como homóloga. Portanto, faz-se necessário entender que proteínas precursoras podem ter existido, uma vez que um possível mecanismo de origem de um gene é que o mesmo deriva de outros pré-existentes. Contudo, o presente trabalho sustenta a idéia de que no momento em que uma proteína guarda similaridade alta com a *Seed*, ela participa de forma mais estável no sistema.

Anteriormente, a mesma abordagem foi utilizada num trabalho preliminar para mostrar que a regulação do desenvolvimento embrionário tem a participação de genes mais ancestrais, como NANOG, que possui homólogos em *Drosófila*, e outros mais recentes como Oct4, que ocorre só em Euteleostomi [74]. Além disso, embora fragmentos de composição de um cenário evolutivo sejam encontrados em diversos trabalhos na literatura, essa é a primeira vez que um sistema é escolhido, o sistema imune, e uma análise ampla é dedicada à inferência de sua formação, contribuindo para a pesquisa sobre o origem de genes a partir do fornecimento de um cenário completo, um campo de prova para trabalhos futuros.

XI. CONCLUSÕES

A clusterização de homólogos mostrou a evolução do sistema imunológico, demonstrando que o estabelecimento da resposta engloba, de um modo geral, o período que se inicia juntamente com a origem dos ossos, em Euteleostomi até o advento dos Eutheria, sendo que a fixação do sistema sempre se conclui com os elementos de membrana. Essa distribuição concorda com a origem total dos genes de H. sapiens.

- A resposta imune anti-viral é uma nova adição na evolução. As principais adições foram obtidas em Euteleostomi e Tetrapoda. A principal transdução de sinal teve origem concomitante ao primeiro interferon e seu receptor cognato, em Euteleostomi. Ficou claro que esse sistema continuou a ser aperfeiçoado com as adições de outras moléculas importantes, como os IFNs tipo III.
- A interação celular na apresentação de antígenos é uma adição nova na evolução. As adições principais foram obtidas com o aparecimento dos amniota, mas a origem do precursor de TCR em Eutheria sugere que esse resposta só foi possível, nos moldes como conhecemos, nesse clado.
- O estabelecimento da resposta associada ao BCR começou em Euteleostomi, mas se tornou completa e efetiva, nos moldes como se conhece hoje em H. sapiens, apenas com o advento dos Eutheria.
- Em relação aos seus marcadores de superfície, as NK são uma afição recente no cenário evolutivo, com receptores cuja ancestralidade varia de Euteleostomi à Euarchontoglires.

- Os TLRs tem uma ampla distribuição ao longo dos clados, sendo que o mais ancestral deles, TLR3, tem origem em Bilateria, sugerindo que a resposta imune natural precede a celular evolutivamente.

XII. PERSPECTIVAS

Algumas perguntas ainda puderam ser levantadas ao final das análises apresentadas neste trabalho, por exemplo, qual a origem das proteínas do sistema complemento? Qual o padrão de surgimento das interleucinas? Seria a origem dos marcadores de células epiteliais e células dendríticas mais ancestral que as demais células apresentadas neste trabalho? A mesma abordagem será aplicada para que essas perguntas sejam respondidas em trabalhos futuros. Este estudo, fornecendo um cenário de aparecimento de proteínas mais recentes na evolução, permite que estudos sejam focados sobre mecanismos de origem de genes específicos. Alguns podem ser associados com genes mais ancestrais, mas é possível que outros tenham tido origem a partir de outros possíveis mecanismos. E, por fim, o estudo abre perspectiva para que sejam analisados os elementos regulatórios dos genes recentes, e se levante a questão: genes que surgem são regulados essencialmente por fatores de transcrição mais novos, ou usam combinações de fatores de transcrição bem remotos? Nosso estudo sugere que proteínas ancestrais passam a ser reguladas por elementos responsivos a interferon. Em conclusão, abrimos perspectivas de retratar mais sub-sistemas da resposta imune, um sistema suficientemente recente na evolução para que tenha sido revelada uma certa ordem de aparecimento que pode subsidiar estudos posteriores.

XIII. REFERÊNCIAS

1. Abbas, Abul K., Andrew H. Lichtman, and Shiv Pillai. **Imunologia celular e molecular**. Elsevier Brasil, 2008.
2. Parslow, Tristram G., et al. **Imunologia médica**. Guanabara Koogan, 2004.
3. Flajnik MF, L du Pasquier. **Evolution of innate and adaptive immunity: can we draw a line?** Trends in Immunology 25:640-644, 2004.
4. Peconick, Ana Paula, and Tutor Karlos Henrique Martins Kalks. **Sistema imune inato**. Disponível em <projetotics.cead.ufla.br> Acesso em: 07/10/2014.
5. Leandro, Carol Góis, et al. **Mecanismos adaptativos do sistema imunológico em resposta ao treinamento físico**. Rev Bras Med Esporte 13.5 (2007): 343-348.
6. Universidade Federal de Ouro Preto. **IMUNIDADE INATA**. Disponível em: <www.nupeb.ufop.br/lip/pdf/imunidadeinata.pdf> Acesso em: 07/10/2014
7. Witko-Sarsat, Véronique, et al. **Neutrophils: molecules, functions and pathophysiological aspects**. Laboratory investigation 80.5 (2000): 617-653.
8. Van Furth, R., et al. **The mononuclear phagocyte system: a new classification of macrophages, monocytes, and their precursor cells**. Bulletin of the World Health Organization 46.6 (1972): 845.
9. Banchereau, Jacques, and Ralph M. Steinman. **Dendritic cells and the control of immunity**. Nature 392.6673 (1998): 245-252.
10. Lanier, Lewis L., and Joseph H. Phillips. **Natural killer cells**. Current opinion in immunology 4.1 (1992): 38-42.
11. Iturry-Yamamoto, G. R., and C. P. Portinho. **Sistema complemento: ativação, regulação e deficiências congênitas e adquiridas**. Revista da Associação Médica Brasileira 47.1 (2001): 41-51.
12. Alberts, Bruce; Alexander Johnson, Julian Lewis, Martin Raff, Keith Roberts, and Peter Walters. **Molecular Biology of the Cell**. Fourth Edition. New York and London: Garland Science, 2002.
13. Varella, Pedro PV, and Wilma C. Neves Forte. **Citocinas: revisão**. Rev bras alergologia imunopatol 24 (2001): 146-154.

14. Kraychete, Durval Campos, Maria Thais de Andrade Calasans, and Camila Motta Leal Valente. **Citocinas pró-inflamatórias e dor**. Rev. bras. reumatol 46.3 (2006): 199-206.
15. Júnior, Danilo Mesquita, et al. **Sistema Imunitário–Parte II Fundamentos da resposta imunológica mediada por linfócitos T e B**. Rev Bras Reumatol 50.5 (2010): 552-80.
16. Parham, Peter. **O Sistema Imune-3**. McGraw Hill Brasil, 2011.
17. Magalhães, P. S. C., Maristela Böhlke, and Fernando Neubarth. **Complexo Principal de Histocompatibilidade (MHC): codificação genética, bases estruturais e implicações clínicas**. Rev. Med. UCPEL (Pelotas) 2.5 (2004): 59.
18. USC School of Medicine, 2011. **Microbiology and Immunology On-Line Textbook**. Disponível em <pathmicro.med.sc.edu/ghaffar/complement.htm> Acesso em: 07/10/2014.
19. ZANON, Renata Graciele et al. **Expressão do complexo de histocompatibilidade principal de classe I (MHC I) no sistema nervoso central: plasticidade sináptica e regeneração**. Coluna/Columna, São Paulo , v. 9, n. 2, June 2010 .
20. Lanzavecchia, Antonio. **Antigen-specific interaction between T and B cells**. (1985): 537-539.
21. **Mecanismos efetores da imunidade humoral**. Disponível em <pt.slideshare.net/GuilhermeSchmittdeAndrade/mecanismos-efetores-da-imunidade-humoral> Acesso em: 20/10/2014.
22. Prof. Dr. Durval Damiani. Major **Histocompatibility Complex and T Cell Receptor**. Disponível em <www.sistemanervoso.com/pagina.php?secao=3&materia_id=550&materiaver=1> Acesso em: 20/10/2014.
23. Julius, Michael, Christiane R. Maroun, and Lorelee Haughn. **Distinct roles for CD4 and CD8 as co-receptors in antigen receptor signalling**. Immunology today 14.4 (1993): 171-176.
24. Borgulya, P., et al. **Development of the CD4 and CD8 lineage of T cells: instruction versus selection**. The EMBO journal 10.4 (1991): 913.
25. SILVA, Dayse Danielle de Oliveira. **Detecção in situ de citocinas de perfil TH1 e TH2 por imunoistoquímica e relação com a atividade de macrófagos nas formas polares da doença de Hansen**. (2009).
26. Dantas, Valéria Cristina Ribeiro. **IgA sérica, secretora, IgE total e estado nutricional em crianças com infecções por enteroparasitas**. (2003).

27. Tan, Joyce T., et al. **Interleukin (IL)-15 and IL-7 jointly regulate homeostatic proliferation of memory phenotype CD8+ cells but are not required for memory phenotype CD4+ cells.** *The Journal of experimental medicine* 195.12 (2002): 1523-1532.
28. Stark, George R., et al. **How cells respond to interferons.** *Annual review of biochemistry* 67.1 (1998): 227-264.
29. Muller, Ulrike, et al. **Functional role of type I and type II interferons in antiviral defense.** *Science* 264.5167 (1994): 1918-1921.
30. Donnelly, Raymond P., and Sergei V. Kotenko. **Interferon-lambda: a new addition to an old family.** *Journal of Interferon & Cytokine Research* 30.8 (2010): 555-564.
31. Ghislain, Julien J., et al. **The interferon-inducible Stat2: Stat1 heterodimer preferentially binds in vitro to a consensus element found in the promoters of a subset of interferon-stimulated genes.** *Journal of Interferon & Cytokine Research* 21.6 (2001): 379-388.
32. Reth, Michael. **B cell antigen receptors.** *Current opinion in immunology* 6.1 (1994): 3-8.
33. Hombach, Joachim, et al. **Molecular components of the B-cell antigen receptor complex of the IgM class.** (1990): 760-762.
34. DeFranco AL. **B-cell activation 2000.** *Immunological Reviews* 176:5-9, 2000.
35. Gene Mayer. **IMUNOGLOBULINAS – ESTRUTURA E FUNÇÃO.** Disponível em <pathmicro.med.sc.edu/portuguese/immuno-port-chapter4.htm> Acesso em: 20/10/2014.
36. **Mecanismos efetores da imunidade humoral.** Disponível em <pt.slideshare.net/GuilhermeSchmittdeAndrade/mecanismos-efetores-da-imunidade-humoral> Acesso em: 20/10/2014.
37. Lochmiller, Robert L., and Charlotte Deerenberg. **Trade-offs in evolutionary immunology: just what is the cost of immunity?.** *Oikos* 88.1 (2000): 87-98.
38. Du Pasquier, L., and J. Schwager. **Evolutions of the immune system.** *Progress in Immunology.* Springer Berlin Heidelberg, 1989. 1246-1255.
39. Du Pasquier, Louis. **The immune system of invertebrates and vertebrates.** *Comparative Biochemistry and Physiology Part B: Biochemistry and Molecular Biology* 129.1 (2001): 1-15.

40. Bang, Frederik B. **Immune reactions among marine and other invertebrates.** *Bioscience* 23.10 (1973): 584-589.
41. Cannon, John P., et al. **The phylogenetic origins of the antigen-binding receptors and somatic diversification mechanisms.** *Immunological reviews* 200.1 (2004): 12-22.
42. Pancer, Zeev, and Max D. Cooper. **The evolution of adaptive immunity.** *Annu. Rev. Immunol.* 24 (2006): 497-518.
43. SMITH, A.. **Oxford Dictionary of Biochemistry and Molecular Biology.** New York: Oxford University Press, 2000. 672 p.
44. de Oliveira, Karina Zupo. **Construção de filogenias baseadas em genomas completos.** Diss. Master's thesis, Institute of Computing, University of Campinas, 2010..
45. Altenhoff AM, Dessimoz C. Inferring orthology and paralogy. *Methods Mol Biol.* 2012;14:259–279.
46. Guedes RLM, Ortega JM: **Desenvolvimento das ferramentas SeedServer, para agrupamento de sequências protéicas homólogas e U-MAGE, para propagação de ontologia funcional.** 2012.
47. Altschul SF, Madden TL, Schäffer AA, Zhang J, Zhang Z, Miller W, Lipman DJ: **Gapped BLAST and PSI-BLAST: a new generation of protein database search programs.** *Nucleic acids research* 1997, **25**:3389–402.
48. Ostlund G, Schmitt T, Forslund K, Köstler T, Messina DN, Roopra S, Frings O, Sonnhammer ELL: **InParanoid 7: new algorithms and tools for eukaryotic orthology analysis.** *Nucleic acids research* 2010, **38**:D196–203.
49. Waterhouse RM, Zdobnov EM, Tegenfeldt F, Li J, Kriventseva EV: **OrthoDB: the hierarchical catalog of eukaryotic orthologs in 2011.** *Nucleic acids research* 2011, **39**:D283–8.
50. Li L, Stoeckert C, Roos D: **OrthoMCL: identification of ortholog groups for eukaryotic genomes.** *Genome research* 2003:2178–2189.
51. Vilella AJ, Severin J, Ureta-Vidal A, Heng L, Durbin R, Birney E: **EnsemblCompara GeneTrees: Complete, duplication-aware phylogenetic trees in vertebrates.** *Genome research* 2009, **19**:327–35.
52. Datta RS, Meacham C, Samad B, Neyer C, Sjölander K: **Berkeley PHOG: PhyloFacts orthology group prediction web server.** *Nucleic acids research* 2009, **37**:W84–9.
53. Altenhoff AM, Dessimoz C: **Phylogenetic and functional assessment of orthologs inference projects and methods.** *PLoS computational biology* 2009, **5**:e1000262.

54. Barbosa-Silva A, Satagopam VP, Schneider R, Ortega JM: **Clustering of cognate proteins among distinct proteomes derived from multiple links to a single seed sequence.** *BMC bioinformatics* 2008, **9**:141.
55. **The Universal Protein Resource (UniProt).** *Nucleic acids research* 2007, **35**:D193–7.
56. Leinonen,R., Diez,F.G., Binns,D., Fleischmann,W., Lopez,R. and Apweiler,R. (2004) **UniProt archive.** *Bioinformatics*, **20**, 3236–3237.
57. Disponível em <<http://www.ebi.ac.uk/uniprot>>. Acesso em: 05/09/2014
58. Zaha, Arnaldo, Henrique Bunselmeyer Ferreira, and Luciane MP Passaglia.**Biologia Molecular Básica-5.** Artmed Editora, 2014.
59. Kanehisa M, Goto S: **KEGG: kyoto encyclopedia of genes and genomes.** *Nucleic Acids Res* 2000, **28**:27-30.
60. Kanehisa M, Araki M, Goto S, Hattori M, Hirakawa M, Itoh M, Katayama T, Kawashima S, Okuda S, Tokimatsu T, et al. **KEGG for linking genomes to life and the environment.** *Nucleic Acids Res.* 2008;**36**:D480-D484.
61. Federhen, S. (2012) **The NCBI taxonomy database.** *Nucleic Acids Res* **40**: D136–D143.
62. BRUSCA, R.C. & BRUSCA, G.J. 2007. **Invertebrados.** 2a.ed., Ed. Guanabara Koogan, Rio de Janeiro. 968p.
63. Melo HV, Ortega JM: **Ferramentas e serviços online para a análise da origem cladística de genes e vias metabólicas.** 2014.
64. Pico AR, Kelder T, van Iersel MP, Hanspers K, Conklin BR, Evelo C. **WikiPathways: pathway editing for the people.** *PLoS Biol.* 2008;**6**:e184.
65. **MediaWiki History 2014.** Disponível em: <www.mediawiki.org/w/index.php?title=MediaWiki_history&oldid=618636>. Acessado em: 23/09/2014.
66. Van Iersel, M. P.; Pico, A. R.; Kelder, T.; Gao, J.; Ho, I.; Hanspers, K.; Conklin, B. R.; Evelo, C. T. (2010). **The BridgeDb framework: standardized access to gene, protein and metabolite identifier mapping services.** *BMC Bioinformatics*
67. Fernandes GR, Ortega JM: **Integração de bases de dados de genes homólogos e aplicações em análises de sequências.** 2011.
68. Fernandes GR, Barbosa DVC, Prosdociami F, Pena IA, Santana-Santos L, Coelho Junior O, Barbosa-Silva A, Velloso HM, Mudado MA, Natale DA, Faria-Campos AC, Aguiar SCV, Ortega JM: **A procedure to recruit members to enlarge protein family databases--the**

building of UECOG (UniRef-Enriched COG Database) as a model. *Genetics and molecular research : GMR* 2008, 7:910–24.

69. Disponível em: <<http://www.lab-bioinfo.com/tools/ueko/>> Acesso em: 15/10/2014

70. 2002. Disponível em: <www.biology.wustl.edu/gcg/psiblast.html> Acesso em: 15/10/2014

71. Ribeiro HAL, Ortega JM: **Desenvolvimento de um serviço de análise de sequências utilizando um modelo baseado em atributos de resultados de PSI-BLAST.** 2014.

72. 2015. Disponível em <<http://www.megasoftware.net/>> Acesso em: 15/10/2014

73. Guedes, R. L. M., et al. **Amino acids biosynthesis and nitrogen assimilation pathways: a great genomic deletion during eukaryotes evolution.** *BMC genomics* 12.Suppl 4 (2011): S2.

74. Donnard, Elisa, et al. **Preimplantation development regulatory pathway construction through a text-mining approach.** *BMC genomics* 12.Suppl 4 (2011): S3.

75. Tamura, Koichiro, et al. **MEGA6: molecular evolutionary genetics analysis version 6.0.** *Molecular biology and evolution* 30.12 (2013): 2725-2729.

76. Edgar, Robert C. **MUSCLE: multiple sequence alignment with high accuracy and high throughput.** *Nucleic acids research* 32.5 (2004): 1792-1797.

77. Barber, G. N. "Host defense, viruses and apoptosis." *Cell death and differentiation* 8.2 (2001): 113-126.

78. Pestka, Sidney, Christopher D. Krause, and Mark R. Walter. **Interferons, interferon-like cytokines, and their receptors.** *Immunological reviews* 202.1 (2004): 8-32.

79. Nguyen, Khuong B., et al. **Interferon α/β -mediated inhibition and promotion of interferon γ : STAT1 resolves a paradox.** *Nature immunology* 1.1 (2000): 70-76.

80. Aaronson, David S., and Curt M. Horvath. **A road map for those who don't know JAK-STAT.** *Science* 296.5573 (2002): 1653-1655.

81. Torpey, Nicholas, et al. **Interferon α but not interleukin 12 activates STAT4 signaling in human vascular endothelial cells.** *Journal of Biological Chemistry* 279.25 (2004): 26789-26796.

82. Chen, Jiabing, Ehtesham Baig, and Eleanor N. Fish. **Diversity and relatedness among the type I interferons.** *Journal of interferon & cytokine research* 24.12 (2004): 687-698.

83. Parmar, Simrit, and Leonidas C. Platanias. **Interferons: mechanisms of action and clinical applications.** *Current opinion in oncology* 15.6 (2003): 431-439.

84. David, Michael. **Signal transduction by type I interferons.** *Biotechniques* 33.Suppl (2002): 58-65.
85. Cao, Weiping, et al. **Toll-like receptor-mediated induction of type I interferon in plasmacytoid dendritic cells requires the rapamycin-sensitive PI (3) K-mTOR-p70S6K pathway.** *Nature immunology* 9.10 (2008): 1157-1164.
86. Dimberg, Anna, et al. **Ser727/Tyr701-phosphorylated Stat1 is required for the regulation of c-Myc, cyclins, and p27Kip1 associated with ATRA-induced G0/G1 arrest of U-937 cells.** *Blood* 102.1 (2003): 254-261.
87. Decker, Thomas, Pavel Kovarik, and Andreas Meinke. **GAS elements: a few nucleotides with a major impact on cytokine-induced gene expression.** *Journal of interferon & cytokine research* 17.3 (1997): 121-134.
88. Kotenko, Sergei V., et al. **IFN-λs mediate antiviral protection through a distinct class II cytokine receptor complex.** *Nature immunology* 4.1 (2002): 69-77.
89. Gong, Q-M., et al. **Association study of IFNAR2 and IL10RB genes with the susceptibility and interferon response in HBV infection.** *Journal of viral hepatitis* 16.9 (2009): 674-680.
90. Witte, Katrin, et al. **IL-28A, IL-28B, and IL-29: promising cytokines with type I interferon-like properties.** *Cytokine & growth factor reviews* 21.4 (2010): 237-251.
91. Lin, Joseph, and Arthur Weiss. **T cell receptor signalling.** *Journal of cell science* 114.2 (2001): 243-244.
92. Vilcek, Jan. **Novel interferons.** *Nature immunology* 4.1 (2003): 8-9. Lin, Joseph, and Arthur Weiss. **T cell receptor signalling.** *Journal of cell science* 114.2 (2001): 243-244.
93. Davis, Mark M., and Pamela J. Bjorkman. **T-cell antigen receptor genes and T-cell recognition.** *Nature* 334.6181 (1988): 395-402.
94. Hancock, John F. **Ras proteins: different signals from different locations.** *Nature Reviews Molecular Cell Biology* 4.5 (2003): 373-385.
95. Davis, Mark M. **A new trigger for T cells.** *Cell* 110.3 (2002): 285-287.
96. Okkenhaug, K., et al. **Phosphoinositide 3-kinase in T cell activation and survival.** *Biochemical Society Transactions* 32.2 (2004): 332-335.
97. Kuchroo, Vijay K., et al. **B7-1 and B7-2 costimulatory molecules activate differentially the Th1/Th2 developmental pathways: application to autoimmune disease therapy.** *Cell* 80.5 (1995): 707-718.

98. Linsley, Peter S., et al. **Human B7-1 (CD80) and B7-2 (CD86) bind with similar avidities but distinct kinetics to CD28 and CTLA-4 receptors.** *Immunity* 1.9 (1994): 793-801.
99. Okkenhaug, Klaus, and Bart Vanhaesebroeck. **PI3K in lymphocyte development, differentiation and activation.** *Nature Reviews Immunology* 3.4 (2003): 317-330.
100. FESKE, Stefan; RAO, Anjana; HOGAN, Patrick G. **The Ca²⁺-calcineurin–NFAT signalling pathway.**
101. Gong, Qian, et al. **Disruption of T cell signaling networks and development by Grb2 haploid insufficiency.** *Nature immunology* 2.1 (2001): 29-36.
102. Alonso, Andres, et al. **Tyrosine phosphorylation of VHR phosphatase by ZAP-70.** *nature immunology* 4.1 (2002): 44-48.
103. Weil, Robert, and Alain Israël. **T-cell-receptor-and B-cell-receptor-mediated activation of NF-κB in lymphocytes.** *Current opinion in immunology* 16.3 (2004): 374-381.
104. SCHMITZ, M. LIENHARD, SUSANNE BACHER, and OLIVER DIENZ. **NF-κB activation pathways induced by T cell costimulation.** *The FASEB journal* 17.15 (2003): 2187-2193.
105. Pfrepper, Klaus-Ingmar, et al. **Structural and functional dissection of the cytoplasmic domain of the transmembrane adaptor protein SIT (SHP2-interacting transmembrane adaptor protein).** *European journal of immunology* 31.6 (2001): 1825-1836.
106. Lin, Hua, et al. **Cytotoxic T lymphocyte antigen 4 (CTLA4) blockade accelerates the acute rejection of cardiac allografts in CD28-deficient mice: CTLA4 can function independently of CD28.** *The Journal of experimental medicine* 188.1 (1998): 199-204.
107. Rudd, Christopher E. **Adaptors and molecular scaffolds in immune cell signaling.** *Cell* 96.1 (1999): 5-8.
108. Mongini, Patricia KA, and John K. Inman. **Cytokine dependency of human B cell cycle progression elicited by ligands which coengage BCR and the CD21/CD19/CD81 costimulatory complex.** *Cellular immunology* 207.2 (2001): 127-140.
109. Richards, James D., et al. **Reconstitution of B cell antigen receptor-induced signaling events in a nonlymphoid cell line by expressing the Syk protein-tyrosine kinase.** *Journal of Biological Chemistry* 271.11 (1996): 6458-6466.

- 110.** Marshall, A. J., et al. **Regulation of B-cell activation and differentiation by the phosphatidylinositol 3-kinase and phospholipase Cgamma pathway.** *Immunological reviews* 176 (2000): 30-46.
- 111.** Neshat MS, Raitano AB, Wang H-G, Reed JC, Sawyers CL. **The Survival Function of the Bcr-Abl Oncogene Is Mediated by Bad-Dependent and -Independent Pathways: Roles for Phosphatidylinositol 3-Kinase and Raf.** *Molecular and Cellular Biology* 2000;20(4):1179-1186.
- 112.** **Genetic analysis of B cell antigen receptor signaling.** *Annual review of immunology* 17.1 (1999): 555-592.
- 113.** Inabe, Kazunori, and Tomohiro Kurosaki. **Tyrosine phosphorylation of B-cell adaptor for phosphoinositide 3-kinase is required for Akt activation in response to CD19 engagement.** *Blood* 99.2 (2002): 584-589.
- 115.** Sutherland, Claire L., et al. **Differential activation of the ERK, JNK, and p38 mitogen-activated protein kinases by CD40 and the B cell antigen receptor.** *The Journal of Immunology* 157.8 (1996): 3381-3390.
- 116.** Medzhitov, Ruslan. **Toll-like receptors and innate immunity.** *Nature Reviews Immunology* 1.2 (2001): 135-145.
- 117.** Beutler, Bruce. **Inferences, questions and possibilities in Toll-like receptor signalling.** *Nature* 430.6996 (2004): 257-263.
- 118.** Kaessmann, Henrik. **Origins, evolution, and phenotypic impact of new genes.** *Genome research* 20.10 (2010): 1313-1326.
- 119.** Knowles, David G., and Aoife McLysaght. **Recent de novo origin of human protein-coding genes.** *Genome research* 19.10 (2009): 1752-1759.

XIV. PRODUÇÃO CIENTÍFICA

1. Trabalhos apresentados em congressos

- Costa, V. R. M.; Vialle, R. A. and Ortega J. M. **The origin of B-cell antigen receptor pathway**, ISCB-Latin America x-Meeting on Bioinformatics with BSB and SoiBio, Belo Horizonte/MG.
- Katia De Paiva Lopes, Ricardo Assunção Vialle, Verônica R M Costa, J Miguel Ortega. **Computational approach for amino acids visualization according to codon usage**, ISCB-Latin America x-Meeting on Bioinformatics with BSB and SoiBio, Belo Horizonte/MG.
- COSTA, VRM.; Tamin, S; Vo, D; Qiao, M; Lopes, KP; RIBEIRO, HAL; Velloso, H; GUEDES, R.LM.; Fernandes, GR; Penalva, LOF; ORTEGA, JM. **Most RNAi target genes are more ancient than RNAi origin: miR 137 as a model**, 59th Brazilian Conference for Genetics. Águas de Lindóia/SP.
- LOPES, K. P. ; COSTA, V. R. M. ; RIBEIRO, H. A. L. ; SANTOS, M. A. ; ORTEGA, J. M. **Singular value decomposition (SVD) from tissue-specific gene expression data**, 59th Brazilian Conference for Genetics. Águas de Lindóia/SP.
- COSTA, VRM.; RIBEIRO, HA.L.; Tamin, S; Vo, D; Qiao, M; Lopes, KP; Penalva, LOF; ORTEGA, JM. **On the pursuit of mir137 target sequence acquisition in recently originated genes**. International conference of the AB3C and Brasiliam Symposiun on Bioinformatics, Recife/PE.
- COSTA, V. R. M.; GUEDES, R. L. M.; RIBEIRO, H. A. L.; ORTEGA, J. M. **Depicting the origin of interferons, its receptors and regulatory transcriptional factors**. 58ª Brazilian Conference for Genetics, Foz do Iguaçu/PR.
- COSTA, VRM.; GUEDES, RLM.; ORTEGA, J M. **The rise of cell-to-cell contacts in the presentation of antigens to t-cells in human evolution**. Internacional Conference of the Brazilian Association for Bioinformatics and Computational Biology, Campinas/SP.

- Tetsu Sakamoto; Verônica RM Costa; Rafael LM Guedes; ORTEGA, JM. **Occurrence and phylogenetic studies of the microcystin biosynthesis genes**. Internacional Conference of the Brazilian Association for Bioinformatics and Computational Biology, Campinas/SP.
- COSTA, VRM; GUEDES, RLM; ORTEGA, JM. **The rise of interferons, its receptors and regulatory factors throughout human evolution**. National Meeting for Biochemistry and Immunology, Belo Horizonte/MG.

2. Prêmios

- Honorable mention for the best oral presentation in the area of Genomics and Bioinformatics/Brazilian society of genetics
- Best Oral Presentation in the area of Genetics and Evolution Human Genetics/Brazilian society of genetics.
- Honorable Mention award for the best poster in the area of Genomics, Sequence Analysis, Evolution and Phylogeny/Brazilian Association Bioinformatics and Computational Biology (AB3C).

XV. ANEXOS

Anexo 1: Via de sinalização TCR

Extra celular					
Intercellular adhesion molecule 1	ICAM-1	P05362	Eutheria		
T-lymphocyte activation antigen CD80	B7-1	P33681	Amniota		
T-lymphocyte activation antigen CD86	B7-2	P42081	Amniota		
Membrana					
Integrin alpha-L	LFA-1	P20701	Euteleostomi	13	
Cytotoxic T-lymphocyte protein 4	CTLA4	P16410	Amniota	17	
Signaling threshold-regulating transmembrane adapter 1	SIT	Q9Y3P8	Mammalia	18	
T-cell-specific surface glycoprotein CD28	CD28	P10747	Tetrapoda	16	
T-cell surface glycoprotein CD4	CD4	P01730	Amniota	17	
T-cell surface glycoprotein CD8 alpha chain	CD8A	P01732	Amniota	17	CD8+
T-cell surface glycoprotein CD8 beta chain	CD8B	P10966	Amniota	17	
T-cell surface glycoprotein CD3 delta chain	CD3D	P04234	Amniota	17	
T-cell surface glycoprotein CD3 epsilon chain	CD3E	P07766	Amniota	17	
T-cell surface glycoprotein CD3 zeta chain	CD247	P20963	Euteleostomi	13	
Phosphoprotein associated with glycosphingolipid-enriched microdomains 1	PAG	Q9NWQ8	Tetrapoda	16	
T-cell receptor-associated transmembrane adapter 1	TRIM	Q6PIZ9	Theria	19	
Linker for activation of T-cells family member 1	LAT	O43561	Theria	19	
Voltage-dependent P/Q-type calcium channel subunit alpha-1A	CACNA1A	O00555	Euteleostomi	13	CaCn
Voltage-dependent N-type calcium channel subunit alpha-1B	CACNA1B	Q00975	Gnathostomata	11	
Voltage-dependent L-type calcium channel subunit alpha-1C	CACNA1C	Q13936	Chordata	8	

Voltage-dependent L-type calcium channel subunit alpha-1D	CACNA1D	Q01668	Euteleostomi	13
Voltage-dependent R-type calcium channel subunit alpha-1E	CACNA1E	Q15878	Amniota	17
Voltage-dependent L-type calcium channel subunit alpha-1F	CACNA1F	O60840	Mammalia	18
Voltage-dependent T-type calcium channel subunit alpha-1G	CACNA1G	O43497	Theria	19
Voltage-dependent T-type calcium channel subunit alpha-1H	CACNA1H	O95180	Bilateria	6
Voltage-dependent T-type calcium channel subunit alpha-1I	CACNA1I	Q9P0X4	Euteleostomi	13
Voltage-dependent L-type calcium channel subunit alpha-1S	CACNA1S	Q13698	Eukaryota	2
Voltage-dependent calcium channel subunit alpha-2/delta-1	CACNA2D1	P54289	Bilateria	6
Voltage-dependent calcium channel subunit alpha-2/delta-2	CACNA2D2	Q9NY47	Euteleostomi	13
Voltage-dependent calcium channel subunit alpha-2/delta-3	CACNA2D3	Q8IZS8	Eumetazoa	5
Voltage-dependent calcium channel subunit alpha-2/delta-4	CACNA2D4	Q7Z3S7	Metazoa	4
Voltage-dependent L-type calcium channel subunit beta-1	CACNB1	Q02641	Chordata	8
Voltage-dependent L-type calcium channel subunit beta-2	CACNB2	Q08289	Eumetazoa	5
Voltage-dependent L-type calcium channel subunit beta-3	CACNB3	P54284	Euteleostomi	13
Voltage-dependent L-type calcium channel subunit beta-4	CACNB4	O00305	Euteleostomi	13
Voltage-dependent calcium channel gamma-1 subunit	CACNG1	Q06432	Euteleostomi	13
Voltage-dependent calcium channel gamma-2 subunit	CACNG2	Q9Y698	Euteleostomi	13
Voltage-dependent calcium channel gamma-3 subunit	CACNG3	O60359	Euteleostomi	13
Voltage-dependent calcium channel gamma-4 subunit	CACNG4	Q9UBN1	Euteleostomi	13
Voltage-dependent calcium channel gamma-5 subunit	CACNG5	Q9UF02	Bilateria	6
Voltage-dependent calcium channel gamma-6 subunit	CACNG6	Q9BXT2	Euteleostomi	13
Voltage-dependent calcium channel gamma-7 subunit	CACNG7	P62955	Bilateria	6

Citosol

Tyrosine-protein phosphatase non-receptor type 11	SHP2	Q06124	Opisthokonta	3
Tyrosine-protein kinase Lck	Lck	P06239	Deuterostomia	7
Tyrosine-protein kinase Fyn	Fyn	P06241	Opisthokonta	3
Tyrosine-protein kinase CSK	CSK	P41240	Opisthokonta	3
Tyrosine-protein kinase ZAP-70	ZAP70	P43403	Opisthokonta	3

Phosphatidylinositol 3-kinase regulatory subunit alpha	PIK3R1	P27986	Bilateria	6	PI3K
Phosphatidylinositol 3-kinase regulatory subunit beta	PIK3R2	O00459	Opisthokonta	3	
Phosphatidylinositol 3-kinase regulatory subunit gamma	PIK3R3	Q92569	Bilateria	6	
Phosphoinositide 3-kinase regulatory subunit 4	PIK3R4	Q99570	Eukayota	2	
Phosphoinositide 3-kinase regulatory subunit 5	PIK3R5	Q8WYR1	Bilateria	6	
Phosphatidylinositol 4-phosphate 3-kinase C2 domain-containing subunit gamma	PIK3C2G	O75747	Eukaryota	2	
Phosphatidylinositol 4,5-bisphosphate 3-kinase catalytic subunit alpha isoform	PIK3CA	P42336	Eukaryota	2	
Phosphatidylinositol 4-phosphate 3-kinase C2 domain-containing subunit alpha	PIK3C2A	O00443	Eukaryota	2	
Phosphatidylinositol 3-kinase catalytic subunit type 3	PIK3C3	Q8NEB9	Eukaryota	2	
Phosphatidylinositol 4-phosphate 3-kinase C2 domain-containing subunit beta	PIK3C2B	O00750	Eukaryota	2	
Phosphatidylinositol 4,5-bisphosphate 3-kinase catalytic subunit delta isoform	PIK3CD	O00329	Eukaryota	2	
Phosphatidylinositol 4,5-bisphosphate 3-kinase catalytic subunit beta isoform	PIK3CB	P42338	Eukaryota	2	
Calmodulin	Calm	P62158	Eukaryota	2	
Serine/threonine-protein phosphatase 2B catalytic subunit alpha isoform	Calcineurin	Q08209	Eukaryota	2	Calcineurin
Serine/threonine-protein phosphatase 2B catalytic subunit beta isoform		P16298	Eukaryota	2	
Serine/threonine-protein phosphatase 2B catalytic subunit gamma isoform		P48454	Eukaryota	2	
Calcineurin subunit B type 1		P63098	Eukaryota	2	
Calcineurin subunit B type 2		Q96LZ3	Eukaryota	2	
Growth factor receptor-bound protein 2	GRB2	P62993	Opisthokonta	3	
Son of sevenless homolog 1	SOS1	Q07889	Opisthokonta	3	SOS
Son of sevenless homolog 2	SOS2	Q07890	Opisthokonta	3	
1-phosphatidylinositol 4,5-bisphosphate phosphodiesterase gamma-1	PCLG1	P19174	Metazoa	4	
GRB2-related adapter protein 2	GADS	O75791	Euteleostomi	13	
Tyrosine-protein kinase ITK/TSK	ITK	Q08881	Metazoa	4	
FYN-binding protein	Fyb	O15117	Euteleostomi	13	
Lymphocyte cytosolic protein 2	SLP76	Q13094	Euteleostomi	13	
RAC-alpha serine/threonine-protein kinase	Akt1	P31749	Eukaryota	2	Akt

RAC-beta serine/threonine-protein kinase	Akt2	P31751	Eukaryota	2	
RAC-gamma serine/threonine-protein kinase	Akt3	Q9Y243	Eukaryota	2	
RAF proto-oncogene serine/threonine-protein kinase	Raf1	P04049	Euteleostomi	13	
GTPase HRas	HRAS	P01112	Eukaryota	2	Ras
GTPase KRas	KRAS	P01116	Eukaryota	2	
GTPase NRas	NRAS	P01111	Eukaryota	2	
Ras-related protein R-Ras	RRAS	P10301	Bilateria	6	
Ras-related protein M-Ras	MRAS	O14807	Eukaryota	2	
Ras-related protein R-Ras2	RRAS2	P62070	Eukaryota	2	
Dual specificity mitogen-activated protein kinase kinase 1	MEK1	Q02750	Eukaryota	2	
Dual specificity mitogen-activated protein kinase kinase 2	MEK2	P36507	Eukaryota	2	
Mitogen-activated protein kinase 3	ERK1	P27361	Eukaryota	2	
Mitogen-activated protein kinase 1	ERK2	P28482	Eukaryota	2	
Protein kinase C theta type	PCK0	Q04759	Opisthokonta	3	
Dual specificity protein phosphatase 3	VHR	P51452	Eukaryota	2	
Caspase recruitment domain-containing protein 11	CARD11	Q9BXL7	Euteleostomi	13	
B-cell lymphoma/leukemia 10	BCL10	O95999	Eumetazoa	5	
Mucosa-associated lymphoid tissue lymphoma translocation protein 1	MALT1	Q9UDY8	Metazoa	4	
Serine/threonine-protein kinase PAK 1	PAK1	Q13153	Eukaryota	2	PAK
Serine/threonine-protein kinase PAK 2	PAK2	Q13177	Opisthokonta	3	
Serine/threonine-protein kinase PAK 3	PAK3	O75914	Eukaryota	2	
Serine/threonine-protein kinase PAK 4	PAK4	O96013	Bilateria	6	
Serine/threonine-protein kinase PAK 7	PAK5	Q9P286	Euteleostomi	13	
Serine/threonine-protein kinase PAK 6	PAK6	Q9NQU5	Euteleostomi	13	
Cell division control protein 42 homolog	CDC42	P60953	Eukaryota	2	
Mitogen-activated protein kinase kinase kinase 1	MAP3K1	Q13233	Chordata	8	MAP3Ks
Mitogen-activated protein kinase kinase kinase 2	MAP3K2	Q9Y2U5	Bilateria	6	
Mitogen-activated protein kinase kinase kinase 3	MAP3K3	Q99759	Eukaryota	2	

Mitogen-activated protein kinase kinase kinase 4	MAP3K4	Q9Y6R4	Bilateria	6	
Mitogen-activated protein kinase kinase kinase 5	MAP3K5	Q99683	Opisthokonta	3	
Mitogen-activated protein kinase kinase kinase 6	MAP3K6	O95382	Eutheria	20	
Nuclear factor of activated T-cells, cytoplasmic 1	NFATC1	O95644	Eumetazoa	5	NFAT
Nuclear factor of activated T-cells, cytoplasmic 2	NFATC2	Q13469	Eumetazoa	5	
Nuclear factor of activated T-cells, cytoplasmic 3	NFATC3	Q12968	Eumetazoa	5	
Nuclear factor of activated T-cells, cytoplasmic 4	NFATC4	Q14934	Eumetazoa	5	
Nuclear factor of activated T-cells 5	NFAT5	O94916	Eumetazoa	5	
Inositol 1,4,5-trisphosphate receptor type 1	ITPR1	Q14643	Eukaryota	2	IP3R
Inositol 1,4,5-trisphosphate receptor type 2	ITPR2	Q14571	Metazoa	4	
Inositol 1,4,5-trisphosphate receptor type 3	ITPR3	Q14573	Eukaryota	2	
Inhibitor of nuclear factor kappa-B kinase subunit alpha	CHUK	O15111	Metazoa	4	INKs
Inhibitor of nuclear factor kappa-B kinase subunit beta	IKBKB	O14920	Bilateria	6	
Inhibitor of nuclear factor kappa-B kinase subunit epsilon	IKBKE	Q14164	Bilateria	6	
NF-kappa-B essential modulator	IKBKG	Q9Y6K9	Chordata	8	
NF-kappa-B inhibitor alpha	NFKBIA	P25963	Bilateria	6	IKB
NF-kappa-B inhibitor beta	NFKBIB	Q15653	Bilateria	6	
NF-kappa-B inhibitor epsilon	NFKBIE	O00221	Chordata	8	
Transcription factor p65	RELA	Q04206	Bilateria	6	NF-KB
Transcription factor RelB	RELB	Q01201	Bilateria	6	
Nuclear factor NF-kappa-B p105 subunit	NFKB1	P19838	Eumetazoa	5	
Nuclear factor NF-kappa-B p100 subunit	NFKB2	Q00653	Metazoa	4	
Proto-oncogene c-Rel	REL	Q04864	Bilateria	6	
Ras-related C3 botulinum toxin substrate 1	RAC1	P63000	Eukaryota	2	Rac
Ras-related C3 botulinum toxin substrate 2	RAC2	P15153	Eukaryota	2	
Ras-related C3 botulinum toxin substrate 3	RAC3	P60763	Eukaryota	2	
Mitogen-activated protein kinase kinase kinase 1	MEKK1	Q13233	Chordata	8	
Dual specificity mitogen-activated protein kinase kinase 4	MK4	P45985	Opisthokonta	3	

Dual specificity mitogen-activated protein kinase kinase 3	MEKK3	P46734	Opisthokonta	3	
Dual specificity mitogen-activated protein kinase kinase 6	MEKK6	P52564	Opisthokonta	3	
Mitogen-activated protein kinase 14	MAPK14	Q16539	Eukaryota	2	p38
Mitogen-activated protein kinase 11	MAPK11	Q15759	Eukaryota	2	
Mitogen-activated protein kinase 12	MAPK12	P53778	Eukaryota	2	
Mitogen-activated protein kinase 13	MAPK13	O15264	Eukaryota	2	
Mitogen-activated protein kinase 8	MAPK8	P45983	Metazoa	4	JNK
Mitogen-activated protein kinase 9	MAPK9	P45984	Metazoa	4	
Mitogen-activated protein kinase 10	MAPK10	P53779	Metazoa	4	
Núcleo					
Transcription factor AP-1	JUN	P05412	Metazoa	4	c-Jun
Cyclic AMP-dependent transcription factor ATF-2	ATF2	P15336	Metazoa	4	
Transcription factor p65	RELA	Q04206	Bilateria	6	NF-KB
Transcription factor RelB	RELB	Q01201	Bilateria	6	
Nuclear factor NF-kappa-B p105 subunit	NFKB1	P19838	Eumetazoa	5	
Nuclear factor NF-kappa-B p100 subunit	NFKB2	Q00653	Metazoa	4	
Proto-oncogene c-Rel	REL	Q04864	Bilateria	6	
Serum response factor	SRF	P11831	Metazoa	4	
ETS domain-containing protein Elk-1	Elk1	P19419	Metazoa	4	
Nuclear factor of activated T-cells, cytoplasmic 2	NFATC2	Q13469	Eumetazoa	5	NFAT
Nuclear factor of activated T-cells, cytoplasmic 1	NFATC1	O95644	Eumetazoa	5	
Nuclear factor of activated T-cells, cytoplasmic 4	NFATC4	Q14934	Eumetazoa	5	
Nuclear factor of activated T-cells, cytoplasmic 3	NFATC3	Q12968	Eumetazoa	5	
Nuclear factor of activated T-cells 5	NFAT5	O94916	Eumetazoa	5	

Anexo 2: Via de sinalização BCR

Membrana				
Complement receptor type 2	CD21	P20023	Amniota	17
B-lymphocyte antigen CD19	CD19	P15391	Eutheria	20
Low affinity immunoglobulin gamma Fc region receptor II-b	FCGR2B/CD32	P31994	Eutheria	20
B-cell antigen receptor complex-associated protein alpha chain	Ig-alpha/CD79a	P11912	Euteleostomi	13
B-cell antigen receptor complex-associated protein beta chain	Ig-beta/CD79b	P40259	Euteleostomi	13
Receptor-type tyrosine-protein phosphatase C	CD45	P08575	Bilateria	6
B-cell receptor CD22	CD22	P20273	Euteleostomi	13
Citosol				
CD81 antigen	CD81	P60033	Eukaryota	2
GTPase HRas Ras	HRas	P01112	Eukaryota	2
GTPase KRas	KRas	P01116	Eukaryota	2
GTPase NRas	NRas	P01111	Eukaryota	2
Ras-related protein R-Ras	RRas	P10301	Bilateria	6
Ras-related protein M-Ras	MRas	O14807	Eukaryota	2
Ras-related protein R-Ras2	RRas2	P62070	Eukaryota	2
Growth factor receptor-bound protein 2	GRB2	P62993	Opisthokonta	3
Son of sevenless homolog 1	SOS1	Q07889	Opisthokonta	3
Son of sevenless homolog 2	SOS2	Q07890	Opisthokonta	3
B-cell receptor-associated protein 31	Bap31	P51572	Euteleostomi	13
CD81 antigen	CD81	P60033	Eukaryota	2
B-cell linker protein	SLP-65	Q8WV28	Bilateria	6
Tyrosine-protein kinase Lyn	Lyn	P07948	Opisthokonta	3
T-lymphocyte activation antigen CD80	B7-1	P33681	Amniota	17
T-lymphocyte activation antigen CD86	B7-2	P42081	Amniota	17
B-cell linker protein	SLP-65	Q8WV28	Bilateria	6
RAF proto-oncogene serine/threonine-protein kinase	Raf1	P04049	Euteleostomi	13
Dual specificity mitogen-activated protein kinase kinase 1	MEK1	Q02750	Eukaryota	2

Dual specificity mitogen-activated protein kinase kinase 2	MEK2	P36507	Eukaryota	2	
Mitogen-activated protein kinase 3	ERK1	P27361	Eukaryota	2	
Mitogen-activated protein kinase 1	ERK2	P28482	Eukaryota	2	
Tyrosine-protein phosphatase non-receptor type 6	PTPN6	P29350	Opisthokonta	3	
Phosphatidylinositol 3,4,5-trisphosphate 5-phosphatase 1	SHIP-1	Q92835	Opisthokonta	3	
Tyrosine-protein kinase Lyn	Lyn	P07948	Opisthokonta	3	
Tyrosine-protein kinase SYK	SYK	P43405	Metazoa	4	
Tyrosine-protein kinase Fyn	Fyn	P06241	Opisthokonta	3	
Tyrosine-protein kinase Blk	BLK	P51451	Opisthokonta	3	
GRB2-associated-binding protein 1	GAB1	Q13480	Eumetazoa	5	GRB2
GRB2-associated-binding protein 2	GAB2	Q9UQC2	Euteleostomi	13	
GRB2-associated-binding protein 3	GAB3	Q8WWW8	Euteleostomi	13	
Tyrosine-protein kinase BTK	BTK	Q06187	Opisthokonta	3	
Proto-oncogene vav	VAV 1	P15498	Opisthokonta	3	Vav
Guanine nucleotide exchange factor	VAV2	P52735	Opisthokonta	3	
Guanine nucleotide exchange factor	VAV3	Q9UKW4	Opisthokonta	3	
Cytoplasmic protein NCK2	NCK2	O43639	Eumetazoa	5	NCK
Cytoplasmic protein NCK1	NCK1	P16333	Opisthokonta	3	
Dual adapter for phosphotyrosine and 3-phosphotyrosine and 3-phosphoinositide	hDAPP1	Q9UN19	Metazoa	4	
Phosphatidylinositol 3,4,5-trisphosphate 3-phosphatase and dual-specificity protein phosphatase PTEN	PTEN	P60484	Eukaryota	2	
1-phosphatidylinositol 4,5-bisphosphate phosphodiesterase gamma-1	PCLG1	P19174	Metazoa	4	PCLg
1-phosphatidylinositol 4,5-bisphosphate phosphodiesterase gamma-2	PCLG2	P16885	Gnathostomata	11	
3-phosphoinositide-dependent protein kinase 1	hPDK1	O15530	Eukaryota	2	
Serine/threonine-protein kinase	mTOR	P42345	Eukaryota	2	
Ribosomal protein S6 kinase beta-1	p70S6K	P23443	Eukaryota	2	
RAC-alpha serine/threonine-protein kinase	Akt1	P31749	Eukaryota	2	Akt
RAC-beta serine/threonine-protein kinase	Akt2	P31751	Eukaryota	2	
RAC-gamma serine/threonine-protein kinase	Akt3	Q9Y243	Eukaryota	2	

PKC	PKC-A	P17252	Opisthokonta	3	PCK
Protein kinase C gamma type	PKC-gamma	P05129	Opisthokonta	3	
Protein kinase C iota type	PRKC-lambda/iota	P41743	Metazoa	4	
Serine/threonine-protein kinase D1	PRKD1	Q15139	Metazoa	4	
Protein kinase C zeta type	PRKCZ	Q05513	Metazoa	4	
Protein kinase C epsilon type	PRKCE	Q02156	Opisthokonta	3	
Serine/threonine-protein kinase D3	PRKD3	O94806	Metazoa	4	
Protein kinase C delta type	PRKCD	Q05655	Opisthokonta	3	
Protein kinase C eta type	PRKCH	P24723	Bilateria	6	
Protein kinase C theta type	PRKCQ	Q04759	Opisthokonta	3	
Protein kinase C beta type	PRKCB	P05771	Opisthokonta	3	
Calmodulin	CaM	P62158	Eukaryota	2	
Serine/threonine-protein phosphatase 2B catalytic subunit alpha isoform	PPP3CA	Q08209	Eukaryota	2	Calcineurin
Serine/threonine-protein phosphatase 2B catalytic subunit beta isoform	PPP3CB	P16298	Eukaryota	2	
Serine/threonine-protein phosphatase 2B catalytic subunit gamma isoform	PPP3CC	P48454	Eukaryota	2	
Calcineurin subunit B type 1	PPP3R1	P63098	Eukaryota	2	
Calcineurin subunit B type 2	PPP3R2	Q96LZ3	Eukaryota	2	
Caspase recruitment domain-containing protein 10	CARD10	Q9BWT7	Amniota	17	
B-cell lymphoma/leukemia 10	BCL10	O95999	Eumetazoa	5	
Mucosa-associated lymphoid tissue lymphoma translocation protein 1	MALT1	Q9UDY8	Metazoa	4	
Glycogen synthase kinase-3 beta	GSK-3 beta	P49841	Eukaryota	2	GSK3
Glycogen synthase kinase-3 alpha	GSK-3 alpha	P49840	Eukaryota	2	
Inhibitor of nuclear factor kappa-B kinase subunit alpha	CHUK	O15111	Metazoa	4	INKs
Inhibitor of nuclear factor kappa-B kinase subunit beta	IKBKB	O14920	Bilateria	6	
Inhibitor of nuclear factor kappa-B kinase subunit epsilon	IKBKE	Q14164	Bilateria	6	
NF-kappa-B essential modulator	IKBKG	Q9Y6K9	Chordata	8	
Phosphatidylinositol 3-kinase regulatory subunit alpha	PIK3R1	P27986	Bilateria	6	PI3K
Phosphatidylinositol 3-kinase regulatory subunit beta	PIK3R2	O00459	Opisthokonta	3	

Phosphatidylinositol 3-kinase regulatory subunit gamma	PIK3R3	Q92569	Bilateria	6	
Phosphoinositide 3-kinase regulatory subunit 4	PIK3R4	Q99570	Eukayota	2	
Phosphoinositide 3-kinase regulatory subunit 5	PIK3R5	Q8WYR1	Bilateria	6	
Phosphatidylinositol 4-phosphate 3-kinase C2 domain-containing subunit gamma	PIK3C2G	O75747	Eukaryota	2	
Phosphatidylinositol 4,5-bisphosphate 3-kinase catalytic subunit alpha isoform	PIK3CA	P42336	Eukaryota	2	
Phosphatidylinositol 4-phosphate 3-kinase C2 domain-containing subunit alpha	PIK3C2A	O00443	Eukaryota	2	
Phosphatidylinositol 3-kinase catalytic subunit type 3	PIK3C3	Q8NEB9	Eukaryota	2	
Phosphatidylinositol 4-phosphate 3-kinase C2 domain-containing subunit beta	PIK3C2B	O00750	Eukaryota	2	
Phosphatidylinositol 4,5-bisphosphate 3-kinase catalytic subunit delta isoform	PIK3CD	O00329	Eukaryota	2	
Phosphatidylinositol 4,5-bisphosphate 3-kinase catalytic subunit beta isoform	PIK3CB	P42338	Eukaryota	2	
Tyrosine-protein kinase ABL1	ABL1	P00519	Euteleostomi	13	Abl
Abelson tyrosine-protein kinase 2	ABL2	P42684	Euteleostomi	13	
Cell division control protein 42 homolog	CDC42	P60953	Eukaryota	2	
Ras-related C3 botulinum toxin substrate 1	Rac1	P63000	Eukaryota	2	Rac
Ras-related C3 botulinum toxin substrate 2	Rac2	P15153	Eukaryota	2	
Ras-related C3 botulinum toxin substrate 3	Rac3	P60763	Eukaryota	2	
Serine/threonine-protein kinase PAK 1	PAK	Q13153	Eukaryota	2	
Dual specificity mitogen-activated protein kinase kinase 1	MEK1	Q02750	Eukaryota	2	MKKs
Dual specificity mitogen-activated protein kinase kinase 3	MAPKK3	P46734	Opisthokonta	3	
Dual specificity mitogen-activated protein kinase kinase 6	MEKK6	P52564	Opisthokonta	3	
Dual specificity mitogen-activated protein kinase kinase 4	MK4	P45985	Opisthokonta	3	
Dual specificity mitogen-activated protein kinase kinase 2	MEK2	P36507	Eukaryota	2	
Dual specificity mitogen-activated protein kinase kinase 5	MAPKK5	Q13163	Opisthokonta	3	
Dual specificity mitogen-activated protein kinase kinase 7	MKK7	O14733	Eumetazoa	5	
Mitogen-activated protein kinase 8	MAPK8	P45983	Metazoa	4	
Mitogen-activated protein kinase 9	MAPK9	P45984	Metazoa	4	
Mitogen-activated protein kinase 14	MAPK14	Q16539	Eukaryota	2	p38
Mitogen-activated protein kinase 11	MAPK11	Q15759	Eukaryota	2	

Mitogen-activated protein kinase 12	MAPK12	P53778	Eukaryota	2	
Mitogen-activated protein kinase 13	MAPK13	O15264	Eukaryota	2	
Núcleo					
Transcription factor p65	RELA	Q04206	Bilateria	6	NF-KB
Transcription factor RelB	RELB	Q01201	Bilateria	6	
Nuclear factor NF-kappa-B p105 subunit	NFKB1	P19838	Eumetazoa	5	
Nuclear factor NF-kappa-B p100 subunit	NFKB2	Q00653	Metazoa	4	
Proto-oncogene c-Rel	REL	Q04864	Bilateria	6	
ETS domain-containing protein Elk-1	Elk1	P19419	Metazoa	4	
Nuclear factor of activated T-cells, cytoplasmic 2	NFATC2	Q13469	Eumetazoa	5	NFAT
Nuclear factor of activated T-cells, cytoplasmic 1	NFATC1	O95644	Eumetazoa	5	
Nuclear factor of activated T-cells, cytoplasmic 4	NFATC4	Q14934	Eumetazoa	5	
Nuclear factor of activated T-cells, cytoplasmic 3	NFATC3	Q12968	Eumetazoa	5	
Nuclear factor of activated T-cells 5	NFAT5	O94916	Eumetazoa	5	
Transcription factor AP-1	JUN	P05412	Metazoa	4	c-Jun
Cyclic AMP-dependent transcription factor ATF-2	ATF2	P15336	Metazoa	4	
Cyclic AMP-responsive element-binding protein 1	CREB1	P16220	Eumetazoa	5	CREB
Cyclic AMP-dependent transcription factor	ATF4	P18848	Metazoa	4	
Cyclic AMP-responsive element-binding protein 3	CREB3	O43889	Metazoa	4	
Cyclic AMP-responsive element-binding protein 3-like protein 4	CREB3L4	Q8TEY5	Metazoa	4	

Anexo 3: Via de sinalização IFNs

Extra celular

Proteína	Short name	Uniprot_ID	LCA	Level
Interferon alpha-1/13	IFNA1\IFNA13	P01562	Euteleostomi	13
Interferon alpha-2	IFNA2	P01563	Euteleostomi	13
Interferon alpha-4	IFNA4	P05014	Euteleostomi	13
Interferon alpha-5	IFNA5	P01569	Euteleostomi	13
Interferon alpha-6	IFNA6	P05013	Euteleostomi	13
Interferon alpha-7	IFNA7	P01567	Euteleostomi	13
Interferon alpha-8	IFNA8	P32881	Euteleostomi	13
Interferon alpha-10	IFNA10	P01566	Euteleostomi	13
Interferon alpha-14	IFNA14	P01570	Euteleostomi	13
Interferon alpha-16	IFNA16	P05015	Euteleostomi	13
Interferon alpha-17	IFNA17	P01571	Euteleostomi	13
Interferon alpha-21	IFNA21	P01568	Euteleostomi	13
Interferon beta	IFNB	P01574	Amniota	17
Interferon omega 1	IFNW1	P05000	Eutheria	20
Interferon kappa	IFNK	Q9P0W0	Amniota	17
Interferon epsilon	IFNE	Q86WN2	Amniota	17
Interferon gamma	IFNG	P01579	Euteleostomi	13
Interferon lambda-1	IL29	Q8IU54	Amniota	17
Interferon lambda-2	IL28A	Q8IZJ0	Amniota	17
Interferon lambda-3	IL28B	Q8IZI9	Amniota	17
Tyrosine-protein kinase Lck	Lck	P06239	Deuterostomia	7

Membrana

Interferon receptor 1	IFNAR1	P17181	Euteleostomi	13
-----------------------	--------	--------	--------------	----

Interferon receptor 2	IFNAR2	P48551	Tetrapoda	16	
Interferon gamma receptor 1	IFNGR1	P15260	Tetrapoda	16	
Interferon gamma receptor 2	IFNGR2	P38484	Tetrapoda	16	
interleukin 10 receptor beta	IL10RB	Q08334	Tetrapoda	16	
interleukin 28 receptor alpha 2	IL28RA	Q8IU57	Tetrapoda	16	
Tyrosine-protein kinase ZAP-70	ZAP70	P43403	Opisthokonta	3	
Citosol					
Janus kinase 1	JAK1	P23458	Opisthokonta	3	
Janus kinase 2	JAK2	O60674	Metazoa	4	
Signal transducer and activator of transcription 1	STAT1	P42224	Euteleostomi	13	
Signal transducer and activator of transcription 2	STAT2	P52630	Tetrapoda	16	
Signal transducer and activator of transcription 3	STAT3	P40763	Euteleostomi	13	
Signal transducer and activator of transcription 4	STAT4	Q14765	Euteleostomi	13	
Signal transducer and activator of transcription 5B	STAT5B	P51692	Opisthokonta	3	
Non-receptor tyrosine-protein kinase	TYK2	P29597	Euteleostomi	13	
Interferon regulatory factor 9	IRF9	Q00978	Euteleostomi	13	
Suppressor of cytokine signaling 1	SOCS1	O15524	Euteleostomi	13	
Suppressor of cytokine signaling 3	SOCS3	O14543	Euteleostomi	13	
PIK3R1	Subunidade 1	P27986	Bilateria	6	Complexo PI3K
PIK3R2	Subunidade 2	O00459	Opisthokonta	3	
PIK3R3	Subunidade 3	Q92569	Bilateria	6	
PIK3R4	Subunidade 4	Q99570	Eukayota	2	
PIK3R5	Subunidade 5	Q8WYR1	Bilateria	6	
Phosphatidylinositol 4-phosphate 3-kinase C2 domain-containing subunit gamma	PIK3C2G	O75747	Eukaryota	2	
Phosphatidylinositol 4,5-bisphosphate 3-kinase catalytic subunit alpha isoform	PIK3CA	P42336	Eukaryota	2	
Phosphatidylinositol 4-phosphate 3-kinase C2 domain-containing subunit alpha	PIK3C2A	O00443	Eukaryota	2	
Phosphatidylinositol 3-kinase catalytic subunit type 3	PIK3C3	Q8NEB9	Eukaryota	2	
Phosphatidylinositol 4-phosphate 3-kinase C2 domain-containing subunit beta	PIK3C2B	O00750	Eukaryota	2	

Phosphatidylinositol 4,5-bisphosphate 3-kinase catalytic subunit delta isoform	PIK3CD	O00329	Eukaryota	2	
Phosphatidylinositol 4,5-bisphosphate 3-kinase catalytic subunit beta isoform	PIK3CB	P42338	Eukaryota	2	
Protein kinase C delta type	PKCδ	Q05655	Opisthokonta	3	
Proto-oncogene vav	VAV 1	P15498	Opisthokonta	3	Complexo VAV
Guanine nucleotide exchange factor	VAV2	P52735	Opisthokonta	3	
Guanine nucleotide exchange factor	VAV3	Q9UKW4	Opisthokonta	3	
Ras-related C3 botulinum toxin substrate 1	Rac1	P63000	Eukaryota	2	
Rho guanine nucleotide exchange factor 1	ARHGEF1	Q92888	Euteleostomi	13	Complexo GEFs
Rho guanine nucleotide exchange factor 2	ARHGEF2	Q92974	Euteleostomi	13	
Rho guanine nucleotide exchange factor 3	ARHGEF3	Q9NR81	Opisthokonta	3	
Rho guanine nucleotide exchange factor 4	ARHGEF4	Q9NR80	Eukaryota	2	
Rho guanine nucleotide exchange factor 5	ARHGEF5	Q12774	Theria	19	
Rho guanine nucleotide exchange factor 6	ARHGEF6	Q15052	Bilateria	6	
Rho guanine nucleotide exchange factor 7	ARHGEF7	Q14155	Bilateria	6	
Neuroepithelial cell-transforming gene 1 protein	NET1	Q7Z628	Opisthokonta	3	
Rho guanine nucleotide exchange factor 9	ARHGEF9	O43307	Eumetazoa	5	
Rho guanine nucleotide exchange factor 10	ARHGEF10	O15013	Bilateria	6	
Rho guanine nucleotide exchange factor 11	ARHGEF11	O15085	Metazoa	4	
Rho guanine nucleotide exchange factor 12	ARHGEF12	Q9NZN5	Euteleostomi	13	
Guanine nucleotide exchange factor DBS	MCF2L	O15068	Metazoa	4	
Rho guanine nucleotide exchange factor 15	ARHGEF15	O94989	Euteleostomi	13	
Rho guanine nucleotide exchange factor 16	ARHGEF16	Q5VV41	Opisthokonta	3	
Mitogen-activated protein kinase kinase kinase 1	MAP3K1	Q13233	Chordata	8	
Dual specificity mitogen-activated protein kinase kinase 3	MAPKK3	P46734	Opisthokonta	3	
Dual specificity mitogen-activated protein kinase kinase 6	MAPKK6	P52564	Opisthokonta	3	
Ribosomal protein S6 kinase alpha-5	MSK1	O75582	Opisthokonta	3	
Ribosomal protein S6 kinase alpha-4	MSK2	O75676	Opisthokonta	3	
Insulin receptor substrate 1	IRS 1	P35568	Metazoa	4	Complexo IRS

Insulin receptor substrate 2	IRS2	Q9Y4H2	Metazoa	4	
Eukaryotic translation initiation factor 4E-binding protein 1	4EBP1	Q13541	Eumetazoa	5	
Eukaryotic initiation factor 4A-I	eIF4A1	P60842	Eukaryota	2	
Eukaryotic translation initiation factor 4E	eIF4E	P06730	Eukaryota	2	
Eukaryotic translation initiation factor 4B	eIF4B	P23588	Eukaryota	2	
Ribosomal protein S6 kinase beta-1	p70S6K	P23443	Eukaryota	2	
40S ribosomal protein S6	RPS6	P62753	Cellular organisms	1	
Serine/threonine-protein kinase	mTOR	P42345	Eukaryota	2	Complexo mTORc1
Target of rapamycin complex subunit LST8	LST8	Q9BVC4	Eukaryota	2	
Regulatory-associated protein of mTOR	rpTOR	Q8N122	Eukaryota	2	
Programmed cell death protein 4	PDCD4	Q53EL6	Eukaryota	2	
Guanine nucleotide-binding protein subunit beta-2-like 1	GNB2L1	P63244	Eukaryota	2	
GRB2-associated-binding protein 2	GAB2	Q9UQC2	Euteleostomi	13	
Protein arginine N-methyltransferase 1	PRMT1	Q99873	Cellular organisms	1	
Tyrosine-protein phosphatase non-receptor type 6	PTPN6	P29350	Opisthokonta	3	
Tyrosine-protein phosphatase non-receptor type 11	SHP2	Q06124	Opisthokonta	3	
E3 ubiquitin-protein ligase CBL	CBL	P22681	Opisthokonta	3	
Tyrosine-protein kinase Fyn	Fyn	P06241	Opisthokonta	3	
Adapter molecule crk	CRK	P46108	Opisthokonta	3	
Crk-like protein	CrkL	P46109	Opisthokonta	3	
Rap guanine nucleotide exchange factor 1	C3G	Q13905	Eukaryota	2	
Ras-related protein Rap-1A	RAP1A	P62834	Eukaryota	2	
Núcleo					
Cyclic AMP-responsive element-binding protein 1	CREB1	P16220	Eumetazoa	5	Complexo CREB
Cyclic AMP-dependent transcription factor	ATF4	P18848	Metazoa	4	
Cyclic AMP-responsive element-binding protein 3	CREB3	O43889	Metazoa	4	
Cyclic AMP-responsive element-binding protein 3-like protein 4	CREB3L4	Q8TEY5	Metazoa	4	
E3 SUMO-protein ligase PIAS1	PIAS1	O75925	Eukaryota	2	

E3 SUMO-protein ligase PIAS3	PIAS3	Q9Y6X2	Eukaryota	2	
Transcription factor p65	RELA	Q04206	Bilateria	6	Complexo NF-KB
Transcription factor RelB	RELB	Q01201	Bilateria	6	
Nuclear factor NF-kappa-B p105 subunit	NFKB1	P19838	Eumetazoa	5	
Nuclear factor NF-kappa-B p100 subunit	NFKB2	Q00653	Metazoa	4	
Proto-oncogene c-Rel	REL	Q04864	Bilateria	6	
Produtos ab					
Proteasome subunit beta type-8	PSB8	P28062	Cellular Organisms	1	
Inositol hexakisphosphate kinase 2	IP6K2	Q9UHH9	Eukaryota	2	
Interferon-stimulated gene 20 kDa protein	ISG20	Q96AZ6	Mammalia	18	
Double-stranded RNA-specific adenosine deaminase	ADAR	P55265	Eumetazoa	5	
Ubiquitin-like protein ISG15	ISG15	P05161	Theria	19	
2'-5'-oligoadenylate synthase 1	OAS1	P00973	Eumetazoa	5	
2'-5'-oligoadenylate synthase 2	OAS2	P29728	Eumetazoa	5	
Interferon-induced 35 kDa protein	IFI35	P80217	Euteleostomi	13	
Interferon-induced GTP-binding protein Mx1	MX1	P20591	Eukaryota	2	
Interferon-induced GTP-binding protein Mx2	MX2	P20592	Eukaryota	2	
Interferon-induced protein with tetratricopeptide repeats 2	IFIT2	P09913	Euarchontoglires	22	
Interferon-induced transmembrane protein 1	IFITM1	P13164	Euteleostomi	13	
Interferon-induced transmembrane protein 2	IFITM2	Q01629	Euteleostomi	13	
Interferon-induced transmembrane protein 3	IFITM3	Q01628	Euteleostomi	13	
Interferon-induced protein with tetratricopeptide repeats 3	IFIT3	O14879	Eutheria	20	
Interferon-induced guanylate-binding protein 2	GBP2	P32456	Eukaryota	2	
XIAP-associated factor 1	XAF1	Q6GPH4	Eukaryota	2	
2-5A-dependent ribonuclease	RNASEL	Q05823	Amniota	17	
Early growth response protein 1	EGR1	P18146	Bilateria	6	
Interferon alpha-inducible protein 6	IFI6	P09912	Mammalia	18	
Interferon regulatory factor 1	IRF1	P10914	Gnathostomata	11	

Interferon regulatory factor 2	IRF2	P14316	Euteleostomi	13	
Interferon regulatory factor 3	IRF3	Q14653	Euteleostomi	13	
Produtos g					
MHC class II transactivator	C2TA	P33076	Euteleostomi	13	
Protein PML	PML	P29590	Eutheria	20	
Nuclear autoantigen Sp-100	SP100	P23497	Simiiformes	25	
Metallothionein-2	MT-2	P02795	Euteleostomi	13	
Interferon-induced guanylate-binding protein 1	GBP1	P32455	Euteleostomi	13	
Interferon-induced guanylate-binding protein 2	GBP2	P32456	Eukaryota	2	
HLA class II histocompatibility antigen, DM alpha chain	HLA-DMA	P28067	Euteleostomi	13	MHC Classe II
HLA class II histocompatibility antigen, DM beta chain	HLA-DMB	P28068	Euteleostomi	13	
HLA class II histocompatibility antigen, DO alpha chain	HLA-DOA	P06340	Euteleostomi	13	
HLA class II histocompatibility antigen, DO beta chain	HLA-DOB	P13765	Euteleostomi	13	
HLA class II histocompatibility antigen, DP alpha 1 chain	HLA-DPA1	P20036	Euteleostomi	13	
HLA class II histocompatibility antigen, DQ alpha 1 chain	HLA-DQA1	P01909	Euteleostomi	13	
HLA class II histocompatibility antigen, DQ alpha 2 chain	HLA-DQA2	P01906	Euteleostomi	13	
HLA class II histocompatibility antigen, DQ beta 1 chain	HLA-DQB1	P01920	Euteleostomi	13	
HLA class II histocompatibility antigen, DR alpha chain	HLA-DRA	P01903	Euteleostomi	13	
HLA class II histocompatibility antigen, DRB1-15 beta chain	HLA-DRB1	P01911	Euteleostomi	13	
HLA class II histocompatibility antigen, DRB1-3 chain	HLA-DRB1	P01912	Euteleostomi	13	
HLA class II histocompatibility antigen, DRB1-1 beta chain	HLA-DRB1	P04229	Euteleostomi	13	
HLA class II histocompatibility antigen, DRB1-8 beta chain	HLA-DRB1	Q30134	Euteleostomi	13	
HLA class II histocompatibility antigen, DRB1-16 beta chain	HLA-DRB1	Q29974	Euteleostomi	13	
HLA class II histocompatibility antigen, DRB1-11 beta chain	HLA-DRB1	P20039	Euteleostomi	13	
HLA class II histocompatibility antigen, DRB1-10 beta chain	HLA-DRB1	Q30167	Euteleostomi	13	
HLA class II histocompatibility antigen, DRB1-13 beta chain	HLA-DRB1	Q5Y7A7	Euteleostomi	13	
HLA class II histocompatibility antigen, DRB1-12 beta chain	HLA-DRB1	Q95IE3	Euteleostomi	13	
HLA class II histocompatibility antigen, DRB1-14 beta chain	HLA-DRB1	Q9GIY3	Euteleostomi	13	

HLA class II histocompatibility antigen, DRB1-7 beta chain	HLA-DRB3	P13761	Euteleostomi	13	
HLA class II histocompatibility antigen, DR beta 3 chain	HLA-DRB3	P79483	Euteleostomi	13	
HLA class II histocompatibility antigen, DRB1-4 beta chain	HLA-DRB4	P13760	Euteleostomi	13	
HLA class II histocompatibility antigen, DR beta 4 chain	HLA-DRB4	P13762	Euteleostomi	13	
HLA class II histocompatibility antigen, DR beta 5 chain	HLA-DRB5	Q30154	Euteleostomi	13	
HLA class II histocompatibility antigen, DRB1-9 beta chain	HLA-DRB5	Q9TQE0	Euteleostomi	13	
2'-5'-oligoadenylate synthase 1	OAS1	P00973	Eumetazoa	5	
2'-5'-oligoadenylate synthase 2	OAS2	P29728	Eumetazoa	5	
2'-5'-oligoadenylate synthase 3	OAS3	Q9Y6K5	Eumetazoa	5	
Low affinity immunoglobulin gamma Fc region receptor III-A	FCGR3A	P08637	Theria	19	Complexo FCGR
Low affinity immunoglobulin gamma Fc region receptor III-B	FCGR3B	O75015	Theria	19	
Low affinity immunoglobulin gamma Fc region receptor II-a	FCGR2A	P12318	Eutheria	20	
Low affinity immunoglobulin gamma Fc region receptor II-c	FCGR2C	P31995	Theria	19	
High affinity immunoglobulin gamma Fc receptor I	FCGR1A	P12314	Eutheria	20	
Low affinity immunoglobulin gamma Fc region receptor II-b	FCGR2B	P31994	Eutheria	20	

Anexo 4: Marcadores de NK e TLRs

Marcadores NK cell

Proteína	Short name	Uniprot	LCA	Level
Low affinity immunoglobulin gamma Fc region receptor III-A	CD16a	P08637	Theria	19
Low affinity immunoglobulin gamma Fc region receptor III-B	CD16b	O75015	Theria	19
Neural cell adhesion molecule 1	CD56	P13591	Metazoa	4
Killer cell immunoglobulin-like receptor 2DL1	CD158a	P43626	Eutheria	20
Killer cell immunoglobulin-like receptor 2DL5A	CD158f	Q8N109	Eutheria	20
Natural cytotoxicity triggering receptor 1	NCR1	O76036	Eutheria	20
Natural cytotoxicity triggering receptor 2	NCR2	O95944	Eutheria	20
Natural cytotoxicity triggering receptor 3	NCR3	O14931	Tetrapoda	16
Natural cytotoxicity triggering receptor 3 ligand 1	NCR3LG1	Q68D85	Gnathostomata	11
Natural killer cells antigen CD94	KLRD1	Q13241	Boreoeutheria	21
NKG2-A/NKG2-B type II integral membrane protein	CD159a	P26715	Boreoeutheria	21
CD160 antigen	CD160	O95971	Euarchothoglires	22
Natural killer cell receptor 2B4	CD244	Q9BZW8	Eutheria	20
C-X-C chemokine receptor type 1	CXCR1	P25024	Euteleostomi	13
Granulysin	GPLY	P22749	Eutheria	20
Sialic acid-binding Ig-like lectin 7	p75	Q9Y286	Euarchothoglires	22
Interleukin-15 receptor subunit alpha	IL15RA	Q13261	Euteleostomi	13
Interleukin-21	IL21	Q9HBE4	Euteleostomi	13
Interleukin-27 subunit alpha	IL27A	Q8NEV9	Eutheria	20
Interleukin-27 subunit beta	IL27B	Q14213	Euteleostomi	13
Interleukin-27 receptor subunit alpha	IL27RA	Q6UWB1	Theria	19
Interleukin-27 receptor subunit beta	IL27RB	Q14213	Euteleostomi	13

Toll-like receptor 1	TLR1	Q15399	Vertebrata	10
Toll-like receptor 2	TLR2	O60603	Vertebrata	10
Toll-like receptor 3	TLR3	O15455	Bilateria	6
Toll-like receptor 4	TLR4	O00206	Euteleostomi	13
Toll-like receptor 5	TLR5	O60602	Euteleostomi	13
Toll-like receptor 6	TLR6	Q9Y2C9	Tetrapoda	16
Toll-like receptor 8	TLR8	Q9NR97	Euteleostomi	13
Toll-like receptor 9	TLR9	Q9NR96	Euteleostomi	13
Toll-like receptor 10	TLR10	Q9BXR5	Euteleostomi	13