



UNIVERSIDADE FEDERAL DA BAHIA
INSTITUTO DE BIOLOGIA

CURSO DE CIÊNCIAS BIOLÓGICAS

**A evolução do domínio RHOGAP e o neurodesenvolvimento em
primatas**

por
LUANA SANTOS SOBRAL

SALVADOR
2019

LUANA SANTOS SOBRAL

A evolução do domínio RHOGAP e o neurodesenvolvimento em primatas

Trabalho de Conclusão de Curso apresentado ao Instituto de Biologia da Universidade Federal Bahia como exigência para obtenção do grau de Bacharel em Ciências Biológicas.

Orientadora: Prof.^a Dr.^a Vanessa Rodrigues Paixão-Cortes

SALVADOR

2019

A evolução do domínio RHOGAP e o neurodesenvolvimento em primatas

Trabalho de Conclusão de Curso apresentado ao Instituto de Biologia da Universidade Federal Bahia como exigência para obtenção do grau de Bacharela em Ciências Biológicas.

Data da Defesa: 03 de dezembro de 2019

Banca Examinadora:

Prof.^a Dr.^a Vanessa Rodrigues Paixão-Cortes – Orientadora
Universidade Federal da Bahia

Prof.^a Dr.^a Renata Lúcia Leite Ferreira de Lima
Universidade Federal da Bahia

Prof. Dr. Rodrigo Barban Zucoloto
Universidade Federal da Bahia

RESUMO

A família RhoGAP, presente em todos os eucariotos, está envolvida na regulação de diversos processos biológicos, incluindo a organização citoesquelética e desenvolvimento neuronal; envolvido na migração neuronal, na formação dos axônios e das sinapses. Duplicações em RHOGAPs que são expressos no cérebro em desenvolvimento (*SRGAP2* e *ARHGAP11*) são associadas a emergência de novidades adaptativas na função cerebral da linhagem humana. Este trabalho tem como objetivo entender a relação entre a perda/ganho dos RHOGAPs e o neurodesenvolvimento em primatas. Utilizamos o Ensembl em busca de todos os genes em primatas que possuíam o domínio proteico RhoGAP. A confirmação da presença/ausência dos genes foi feita a partir da busca em outros três bancos de dados, o UniProt (proteínas), o Genomicus e o GenBank/NCBI. Verificamos a expressão de todos os genes encontrados, utilizando o banco de dados GenBank, escolhendo os genes onde o tecido cerebral fosse pelo menos o terceiro tecido onde ocorre a maior expressão do gene. Recuperamos as sequências nucleotídicas de todos os genes, utilizando o programa MEGA X alinhamos estas pelo algoritmo Muscle, determinamos o melhor modelo evolutivo, o qual o programa indicou o modelo Kimura-2 + G e construímos a filogenia por máxima verossimilhança de todos os domínios RHOGAPs com bootstrap=1000. Encontramos 23 genes expressos significativamente no cérebro em 24 primatas. Vinte e três primatas possuem todos os genes, o Tarsier, é o único primata em que faltam 4 genes (*ARHGAP23*, *ARHGAP35*, *SRGAP3* e *PIK3R2*), devido a cobertura e o estágio de anotação que se encontra esse primata. Recuperamos múltiplas cópias de 3 genes em alguns dos primatas. Dentre eles o *ARHGAP5* que apresenta uma duplicação em 14 dos 16 os primatas do velho mundo estudados (Catarrhini). E o *ARHGAP21* que aparece nos Homininae. Estudando a árvore gerada pelo alinhamento de aproximadamente 529 domínios RHOGAP, observou-se que as regiões que codificam os domínios formavam grupos, em sua maioria, de parálogos da mesma espécie; inicialmente suspeitamos que seria resultado da alta conservação do domínio e em alguns casos eventos de conversão gênica, pois obtivemos alguns ramos sem suporte e outros com altos valores de bootstrap, entretanto existem muitas regiões específicas de cada gene que foram retiradas da análise e isto deve ter sido o motivo do agrupamento dos parálogos de uma mesma espécie.

Palavras-chave: *ARHGAP5*, *ARHGAP21*, duplicações gênicas, pseudogenização

ABSTRACT

The RhoGAP family, present in all eukaryotes, is involved in the regulation of various biological processes, including cytoskeletal organization and neuronal development; participating in neuronal migration, in the formation of axons and synapses. Duplications in RHO GAPs that are expressed in the developing brain (*SRGAP2* and *ARHGAP11*) are associated with emergencies of adaptive novelties in brain function of the human lineage. This paper aims to understand the relationship between loss/gain of RhoGapS and neurodevelopment in primates. We used Ensembl to search for all genes in primates that had the RhoGAP protein domain. Confirmation of the presence/absence of genes was made by searching three other databases, UniProt (proteins), Genomicus and GenBank / NCBI. We checked the expression of all genes found using the GenBank database by selecting only genes in that brain tissue was at least third tissue of highest expression of gene. We retrieve the nucleotide sequences of all genes, we alignments, and we determine the best evolutionary model and build maximum likelihood phylogeny of all *RHO GAPs* domains using the Muscle algorithm and the tools available in the MEGA X program, Kamura-2 + G model, with bootstrap = 1000. We found 23 genes expressed significantly in the brain in 24 primates. Twenty-three primates have all genes, the only one exception is the Tarsier, which is missing four genes (*ARHGAP23*, *ARHGAP35*, *SRGAP3*, and *PIK3R2*), and due to the coverage and annotation study being this primate. We recovered duplicates in three genes in some of the primates. *ARHGAP5*, which has a duplication in several old world primates (Catarrhini), we identified that there has been a pseudogenization throughout the group. In addition, the *ARHGAP21* that appears in Hominae. Studying a tree generated by the alignment of approximately 529 RhoGap domains we observed that domains are grouped, most of the time, of paralogs of the same species; initially we suspect it would be a result of high domain conservation and possibly in some cases gene conversion events. There was some unsupported branches and others with high bootstrap values, but there was many specific regions of each gene that are removed from the analysis, this artifact should be the reason for grouping the paralogous of the same species.

Keywords: *ARHGAP5*, *ARHGAP21*, gene duplications, pseudogenization

AGRADECIMENTOS

A minha família, pelo apoio incondicional;

Aos meus amigos, que tornaram a caminhada mais leve;

Aos professores que proporcionaram uma visão crítica ao estudo da vida e que por meio de seus esforços me tornaram uma estudante melhor;

A minha orientadora, que aceitou me orientar mesmo depois de tantas idas e vindas;

A Universidade Federal da Bahia por me fornecer a possibilidade de aprendizagem e experiências maravilhosas;

A todos e todas que permitiram que eu chegasse até aqui.

SUMÁRIO

AGRADECIMENTOS	I
SUMÁRIO	II
LISTA DE FIGURAS E TABELAS	III
1. INTRODUÇÃO	1
1.1. FAMÍLIA RHOGAP	1
1.2. RHOGAPS E O NEURODESENVOLVIMENTO	2
1.3. EVOLUÇÃO GÊNICA	3
1.4. GENES DO NEURODESENVOLVIMENTO EM PRIMATAS	4
2. OBJETIVOS	7
OBJETIVO GERAL	7
OBJETIVOS ESPECÍFICOS	7
3. METODOLOGIA	8
3.1. OBTENÇÃO DAS SEQUÊNCIAS	8
3.2. ESCOLHA DOS GENES	8
3.3. CONFIRMAÇÃO DE AUSÊNCIA/PRESENÇA E DUPLICAÇÃO DOS GENES	8
3.4. ALINHAMENTO E ANÁLISE EVOLUTIVA	9
4. RESULTADOS	10
4.1. AMOSTRAS	10
4.2. OBTENÇÃO DAS SEQUÊNCIAS	10
4.3. GENES ESCOLHIDOS	11
4.4. AUSÊNCIA E PRESENÇA DOS GENES	12
4.5. ANÁLISE EVOLUTIVA	12
5. DISCUSSÃO	15
6. CONCLUSÃO	20
7. REFERÊNCIAS	21

LISTA DE FIGURAS E TABELAS

Tabela 1- Primatas utilizados no trabalhop.11

Tabela 2 - Genes de trabalho e suas respectivas descrições das proteínas baseadas em informações do NCBI.....p. 12

Tabela 3- Presença e ausência de genes.....p. 13

Figura 1 - Mecanismos de regulação das Rho Gtpeases - Retirado de Moon e Zheng, (2003) 1

Figura 2 - Processos pelo qual passa um gene duplicado - Retirado de Hurles (2004)..... 4

Figura 3 - Evolutionary analysis by Maximum Likelihood method. - TRECHOS DA ÁRVORE GERADA14

1. INTRODUÇÃO

1.1. FAMÍLIA RHOGAP

A superfamília Ras corresponde a um grande grupo de proteínas que estão envolvidas na organização e sinalização celular, mudanças na estrutura e regulação dessas proteínas podem modificar como esta sinalização é realizada. A superfamília Ras é composta pelas famílias Ran, Ras, Rab, Rho e Arf (ROJAS et al., 2012).

As proteínas Rho Gtpases atuam em diversas atividades celulares englobando a polimerização da actina, a adesão celular, o ciclo celular, a polaridade da membrana entre outras coisas. As proteínas Rho Gtpases são reguladas pelos grupos de proteínas efetoras GEFs, GDIs e GAPs. As GEFs que promovem a liberação do GDP permitindo a ligação do GTP, as GDIs que se ligam ao GDP ligado impedindo que ele se desligue da Rho Gtpase, e as proteínas de nosso interesse são as efetoras GAPs, conhecidas como RhoGAPs. Elas inativam as proteínas Rho Gtpases aumentando a taxa intrínseca de hidrólise do GTP (LAMACHE e HALL,1994; PECK et al.,2002).

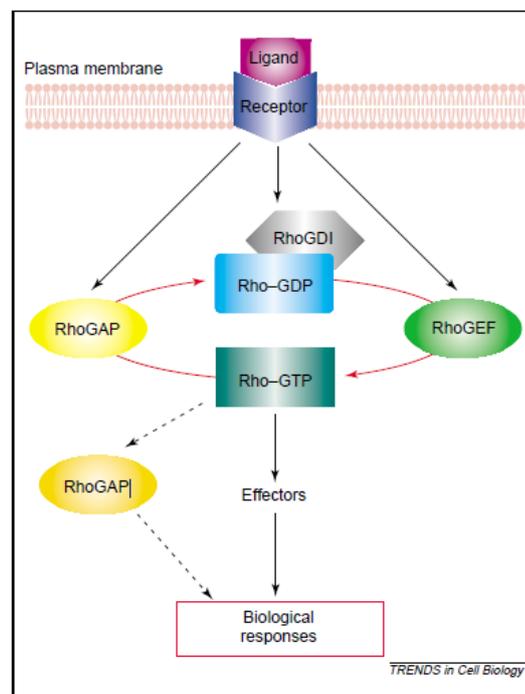


Figura 1 - Mecanismos de regulação das Rho Gtpases - Retirado de Moon e Zheng, (2003)

RhoGAPs presentes em todos os eucariotos, estão envolvidas na regulação de diversos processos biológicos, incluindo a organização citoesquelética e desenvolvimento neuronal, participando na migração neuronal, na formação dos axônios e das sinapses. Devido a estas funções as RhoGAPs são boas candidatas para se compreender o neurodesenvolvimento (MOON e ZHENG, 2003). A família dos genes RHOGAPS possui pelo menos 51 membros parálogos em humanos, com algumas duplicações específicas na linhagem. Os RHOGAPS possuem suas sequências conservadas ao longo da evolução em primatas não humanos e humanos.

1.2. RHOGAPS E O NEURODESENVOLVIMENTO

Vários genes da família estão envolvidos com mudanças no citoesqueleto, na migração dos neurônios, durante o desenvolvimento do cérebro; e muitos genes da família RHOGAP são expressos no cérebro durante a vida, nesse estudo focamos nos genes que possuem uma expressão significativa no tecido cerebral.

Exemplificando, um dos genes ativos durante o desenvolvimento o *SRGAP1* (SLIT-ROBO Rho GTPase activating protein 1) regula a atividade das proteínas RAC em neurônios, conseqüentemente, a migração dos neurônios (YAMAZAKI, et al, 2013). O *ARHGAP5* (Rho GTPase activating protein 5) regula negativamente a Rho GTPases e assim conduz a mudanças no citoesqueleto de células neurais (DEMALI et al, 2003). Outro que pode ser citado é o *SRGAP3* (SLIT-ROBO Rho GTPase activating protein 3), que influencia o crescimento das espinhas dendríticas(filópodes) dos neurônios. Já foi hipotetizado que *SRGAP3* pode estar envolvido no processo que tornou os humanos diferenciados dos outros primatas, como a capacidade cognitiva e o aprendizado cumulativo, que podem ser uma consequência de uma maior quantidade de conexões entre os neurônios (BACON, 2013).

Outros genes importantes para a maior complexidade das estruturas cerebrais são as cópias do *SRGAP2* (SLIT-ROBO Rho GTPase activating protein 2), que interagem com *RAC1*, que ligam-se aos neurônios e deforma as

membranas, controlando a dinâmica das fibras de actina para regular a migração e a diferenciação dos neurônios. Esses genes duplicados desempenham um papel importante em diferentes aspectos neurais, entre eles a morfogênese e, principalmente, a migração durante o desenvolvimento do córtex cerebral (crescimento de axônios e dendritos e maturação de espinhas dendrítica) (DENNIS et al, 2012). Uma das cópias, o *SRGAP2C*, está presente apenas em humanos modernos e arcaicos, agindo de forma antagonista ao gene original no desenvolvimento neuronal cortical. Ele está envolvido na maturação das espinhas dendríticas (alterando a morfologia e densidade), o que resulta em implicações para a cognição, aprendizagem e memória (DENNIS et al, 2012). Outro exemplo de gene duplicado exclusivo da linhagem humana implicado ao neurodesenvolvimento é o *ARHGAP11B*, este expressa uma proteína específica de Hominídeos que promove o desenvolvimento e expansão evolutiva do neocórtex cerebral (FLORIO, et al 2015).

1.3. EVOLUÇÃO GÊNICA

Eventos de duplicações no genoma ou do genoma completo geram material genético que servem como substratos para a evolução molecular e impactam o processo evolutivo. A organização do genoma dos vertebrados sofreu influência das sucessivas duplicações do genoma inteiro, embora sejamos diploides, alguns genes foram perdidos, outros foram mantidos duplicados, formando e/ou expandindo famílias de genes (PAIXÃO-CÔRTEZ, 2015).

A duplicação de genes oferece novas oportunidades evolutivas, fornecendo uma região nova onde há um relaxamento da pressão evolutiva. Depois da duplicação podem ocorrer três processos: a neofuncionalização, a subfuncionalização e a degradação do gene duplicado (HURLES, 2004; LEVINE e TIJAN, 2003). A posterior fixação do gene duplicado leva a formação de famílias genicas, possibilitando que surjam diferentes genes produzindo proteínas diversas que exercem atividades distintas (HURLES, 2004). Uma duplicação do gene *RAC* ancestral deu origem: ao *CDC42* (controla a polaridade

das células) e a família *RHO* (envolvida na citocinese), sendo um bom exemplo de como a duplicação favorece o surgimento de genes com novas funções. (JAFFE e HALL, 2005). Dada a importância das duplicações para a evolução de novas funções, fica o questionamento se é possível relacionar as duplicações ao surgimento de novidades adaptativas documentando a trajetória destas duplicações.

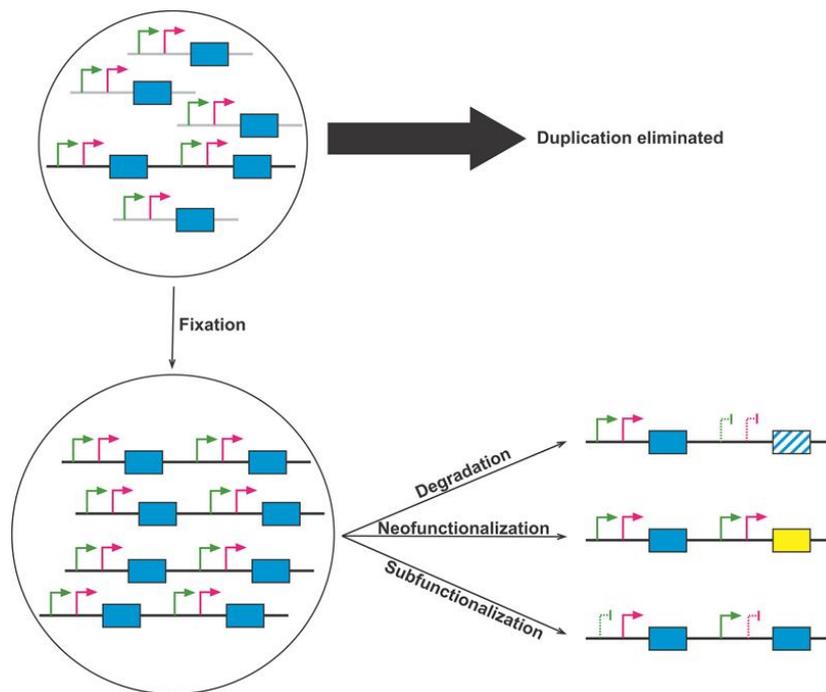


Figura 2 - Processos pelo qual passa um gene duplicado - Retirado de Hurles (2004)

1.4. GENES DO NEURODESENVOLVIMENTO EM PRIMATAS

Fortna (2004) encontrou especificamente 134 genes que possuem mais de uma cópia em humanos (*Homo sapiens*) e não estão presentes nos outros membros da família *Hominidae* à qual pertence o Bonobo (*Pan paniscus*), Chimpanzés (*Pan troglodytes*), Gorila (*Gorilla gorilla*) e o Orangotango (*Pongo pygmaeus abelii*). Essas duplicações foram associadas à diferenciação da

linhagem humana e associadas a evolução de novas funções adaptativas. (DENNIS, 2016; DOUGHERTY, 2018).

Um desses genes duplicados é *SRGAP2* (Slit-Robo Rho GTPase activating protein), citado acima, expresso no cérebro sua função está relacionada ao desenvolvimento do neocórtex, possui múltiplas cópias em humanos (*SRGAP2A*, *SRGAP2B*, *SRGAP2C*, *SRGAP2D*), foi associado a maior capacidade do cérebro humano de gerar mais conexões neurais e, conseqüentemente, ter uma maior complexidade anatômica e fisiológica, tendo assim maior capacidade cognitiva (TYLER-SMITH e XUE, 2012). Recentemente foi descoberto que o *ARHGAP11* também sofreu duplicação e deu origem a uma proteína específica de homínídeos que promove o desenvolvimento e expansão evolutiva do neocórtex cerebral (FLORIO, et al 2015).

O *SRGAP2* e o *ARHGAP11* são apontados como uma importante fonte de mudança fenotípica e evolução adaptativa (CHARRIER et al, 2012). Dennis et al, (2012), mostram que os seres humanos possuem quatro cópias não-identicas (nomeadas A-D) do *SRGAP2* em diferentes loci no cromossomo 1. As cópias *SRGAP2B* e *SRGAP2D* possuem expressividade baixa e são mais propensas a variações na sequência entre os seres humanos, do que as cópias A e C (TYLER-SMITH e XUE, 2012), sugerindo assim que o *SRGAP2C* pode ter desempenhado um papel importante no surgimento da linhagem Homo de 2 a 3 milhões de anos atrás, quando especializações cerebrais humanas, entre elas, as que conduzem ao desenvolvimento da linguagem, cognição social e resolução de problemas, provavelmente, evoluíram (GUERRIER et al, 2009). Florio, et al (2015), identificaram 56 genes expressos preferencialmente em células da glia apical e basal em humanos e que não possuem ortólogos (genes com um ancestral comum, compartilhados por duas ou mais espécies, basicamente o mesmo gene em diferentes espécies) em camundongo. Entre eles está o *ARHGAP11B*. Esse gene é expresso durante o desenvolvimento do neocórtex humano. Promove o aumento da ampliação de progenitores basais na zona subventricular, e assim, produz mais neurônios durante a corticogênese fetal. O *ARHGAP11B* surgiu da duplicação parcial do *ARHGAP11A*, não possui todos os domínios proteicos, após a separação da linhagem de chimpanzé.

Uma das questões mais interessantes do ponto de vista evolutivo é como a linhagem humana desenvolveu sua capacidade cognitiva. Um dos maiores desafios da biologia na atualidade, é poder relacionar alterações nos genes à mudanças fenotípicas em organismos complexos. Os genes da família RHOGAP, seriam bons candidatos para o estudo do surgimento de novidades adaptativas nas linhagens dos primatas.

2. OBJETIVOS

OBJETIVO GERAL

Associar eventos de perdas e ganhos dos genes da família RHOGAP e o surgimento de novidades adaptativas relacionadas ao neurodesenvolvimento em primatas.

OBJETIVOS ESPECÍFICOS

1. Quantificar o número de genes contendo o domínio RHOGAP em primatas;
2. Determinar quais genes da família RHOGAP estão associados ao neurodesenvolvimento e/ou são expressos significativamente no cérebro;
3. Estabelecer as perdas e duplicações dos genes expressos no cérebro dos primatas;
4. Associar perdas/duplicações a possíveis traços específicos em uma linhagem;
5. Construir uma filogenia dos domínios RHOGAP para tentar elucidar padrões evolutivos da família.

3. METODOLOGIA

3.1 OBTENÇÃO DAS SEQUÊNCIAS

Para encontrar todos os genes alocados na plataforma que tivesse o domínio de interesse, realizou-se uma busca por IPR000198 em todas as espécies disponíveis no Ensembl Release 98¹. O código IPR000198 refere-se ao InterPro² do domínio, uma plataforma que faz análise funcional das proteínas e a classifica em domínios. Obtidos todos os animais e seus respectivos genes com o domínio utilizamos a ferramenta BIOMART do Ensembl para verificar a presença/ausência e a recuperação das sequências de todos os genes que continham em sua estrutura o domínio de interesse, o RhoGap. Optamos trabalhar com o *Ensembl*, por ser uma plataforma onde se encontra o genoma dos animais, não apenas informações isoladas de regiões específicas do DNA. Depois de obtidos todos os genes que continham o domínio RhoGap, as sequências que estavam sem notação foram identificadas por homologia e receberam notação manual, animais que continham genes não identificados foram descartados. Além disso escolheu-se um grupo focal para o trabalho, os primatas.

3.2 ESCOLHA DOS GENES

Incluimos no estudo genes que continham o domínio RhoGap e que preenchiam o critério estabelecido: genes onde o tecido cerebral era pelo menos o terceiro tecido de maior expressão deste gene baseadas em informações de expressão obtidos pelo GenBank.

3.3 CONFIRMAÇÃO DE AUSÊNCIA/PRESENÇA E DUPLICAÇÃO DOS GENES

Para confirmar se determinado gene estava ausente no genoma do animal procurou-se a sequência em outros três bancos de dados o *UniProt*³

(proteínas), o *Genomicus*⁴ e o *GenBank/NCBI*⁵. Dessa forma a chance de a ausência ser por inexistência tem maior probabilidade de ser condizente com a realidade do que por ausência de dados.

Para a confirmação das duplicações se analisou em outros bancos como *Genome Browser UCSC*⁶ (*BLAT*) e o *Genomicus*.

3.4 ALINHAMENTO E ANÁLISE EVOLUTIVA

Para entender a dinâmica de ganhos e perdas de genes e domínios proteicos, determinamos os domínios presentes em cada proteína utilizando o banco *UniProt*. Alinhamos as sequências do domínio RhoGap, único domínio presente em todos os genes da família estudados, pelo algoritmo Muscle e destacamos a região do domínio para construir a filogenia a partir delas. Determinamos o melhor modelo evolutivo e construímos o qual o programa indicou o modelo Kimura-2 + G. Construímos a filogenia por máxima verossimilhança de todos os domínios com bootstrap=1000. Todos esses passos foram utilizados a partir de ferramentas no programa Mega X.

4. RESULTADOS

4.1 AMOSTRAS

Das 113 espécies de vertebrados disponíveis no *Ensembl databank* que possuem genes com domínio RhoGAP. Para a realização do trabalho foi determinado como grupo focal os primatas, os quais foram selecionados 24, estes 24 foram aqueles que estavam no Ensembl e tiveram todas as sequências identificadas (Tabela 1).

4.2 OBTENÇÃO DAS SEQUÊNCIAS

Recuperamos¹ a partir da busca no Ensembl em 113 animais 6.660 sequências que possuem o domínio RhoGap (que serão utilizadas no projeto da orientadora). Retiramos os animais sem anotação e escolhemos apenas os 24 primatas para esse trabalho, com 1.600 sequências. Os animais e a quantidade de genes/sequências podem ser conferidos na tabela abaixo:

Nome Comum	Espécie	Nº de genes com domínio RhoGap
<u>Bushbaby</u>	<i>Otolemur garnettii</i>	63
<u>Mouse Lemur</u>	<i>Microcebus murinus</i>	64
<u>Coquerel's sifaka</u>	<i>Propithecus coquereli</i>	64
<u>Marmoset</u>	<i>Callithrix jacchus</i>	65
<u>Bolivian squirrel monkey</u>	<i>Saimiri boliviensis boliviensis</i>	63
<u>Capuchin</u>	<i>Cebus capucinus imitator</i>	66
<u>Ma's night monkey</u>	<i>Aotus nancymae</i>	66
<u>Sooty mangabey</u>	<i>Cercocebus atys</i>	66
<u>Drill</u>	<i>Mandrillus leucophaeus</i>	68
<u>Olive baboon</u>	<i>Papio anubis</i>	68
<u>Crab-eating macaque</u>	<i>Macaca fascicularis</i>	68

¹ Embora esse trabalho de conclusão siga as regras ABNT como exigido pelo colegiado, ele está escrito na voz ativa da terceira pessoa, pois há um movimento da comunidade científica para que alunos e pesquisadores assumam os seus trabalhos e descobertas.

<u>Macaque</u>	<i>Macaca mulatta</i>	66
<u>Pig-tailed macaque</u>	<i>Macaca nemestrina</i>	70
<u>Vervet-AGM</u>	<i>Chlorocebus sabaues</i>	64
<u>Angola colobus</u>	<i>Colobus angolensis palliatus</i>	67
<u>Black snub-nosed monkey</u>	<i>Rhinopithecus bieti</i>	63
<u>Golden snub-nosed monkey</u>	<i>Rhinopithecus roxellana</i>	66
<u>Orangutan</u>	<i>Pongo abelii</i>	66
<u>Gorilla</u>	<i>Gorilla gorilla gorilla</i>	71
<u>Bonobo</u>	<i>Pan paniscus</i>	68
<u>Chimpanzee</u>	<i>Pan troglodytes</i>	73
<u>Human</u>	<i>Homo sapiens</i>	79
<u>Gibbon</u>	<i>Nomascus leucogenys</i>	69
<u>Tarsier</u>	<i>Carlito syrichta</i>	57

Tabela 1- Primatas utilizados no trabalho

*Somente foram curados para verificar se eram realmente genes e não transcritos alternativos os genes que preenchiam o critério de inclusão no trabalho

4.3 GENES ESCOLHIDOS

Considerando que o trabalho procura relacionar o neurodesenvolvimento em primatas com o domínio RhoGap, escolheu-se aqueles genes onde havia expressão significativa e que pudesse direcionar fortemente a evolução do gene, por esse motivo escolheu-se aqueles genes onde o tecido nervoso/cerebral era até o terceiro tecido onde ocorria a maior expressão. Estabelecido esse critério, de todos os genes em primatas que continha o gene RhoGap que ao todo eram 66 genes, apenas 23 respeitava o critério estabelecido e portanto estes foram escolhidos, resultando em um total de 529 sequências. Os genes escolhidos podem ser conferidos na tabela seguinte (tabela 2).

GENE	DESCRIÇÃO DAS PROTEÍNAS	DOMÍNIOS PRESENTES ALÉM DO RHOGAP
ABR	Ativadora da RhoGEF e GTPase	DH, PH, C2
ARHGAP5	Ativadora da RhoGTPase 5	FF1, FF2, FF3, FF4, pG1, pG2
ARHGAP12	Ativadora da RhoGTPase 12	SH3, WW1, WW2, PH
ARHGAP20	Ativadora da RhoGTPase 20	PH, Ras-associating
ARHGAP21	Ativadora da RhoGTPase 21	PDZ, PH
ARHGAP22	Ativadora da RhoGTPase 22	PDZ, PH
ARHGAP23	Ativadora da RhoGTPase 23	
ARHGAP26	Ativadora da RhoGTPase 26	PH, SH3
ARHGAP33	Ativadora da RhoGTPase 33	PX, SH3

ARHGAP35	Ativadora da RhoGTPase 35	FF1, FF2, FF3, FF4, pG1, pG2
ARHGAP36	Ativadora da RhoGTPase 36	-
ARHGAP39	Ativadora da RhoGTPase 39	WW1, WW2, MyTH4
ARHGAP44	Ativadora da RhoGTPase 44	BAR
BCR	Ativadora da RhoGEF e GTPase	DH, PH, C2
CHN1	Ativadora da GTPase para p21-rac	SH2
CHN2	Ativadora de GTPase para rac	SH2
FAM13B	Família com similaridade de sequência 13 membro B	-
OPHN1	Ativadora da RhoGTPase	PH
PIK3R1	Subunidade reguladora 1 da fosfoinositídeo-3-cinase	SH3, 2 CÓPIAS DE SH2
PIK3R2	Subunidade reguladora 2 da fosfoinositídeo-3-cinase	SH3, 2 CÓPIAS DE SH2
SRGAP1	Ativadora 1 da SLIT-ROBO Rho GTPase	F-BAR, SH3
SRGAP2	Ativadora 2 da SLIT-ROBO Rho GTPase	F-BAR, SH3
SRGAP3	Ativadora 3 da SLIT-ROBO Rho GTPase	F-BAR, SH3

Tabela 2 - Genes de trabalho e suas respectivas descrições das proteínas baseadas em informações do NCBI. Informações sobre os domínios retiradas do UniProt.

4.4 AUSÊNCIA E PRESENÇA DOS GENES

Para garantir que o gene realmente estava ausente foi feita a busca em outros três bancos de dados procurando por sequências homologas aos supostos genes ausentes. Com exceção do *Tarsier* todos os outros primatas possuem os 23 genes. Abaixo, na tabela 3, segue a lista de supostas duplicações e ausência do gene.

4.5 ANÁLISE EVOLUTIVA

Analizamos a árvore gerada pelo alinhamento de 529 domínios e observamos que os domínios formaram grupos, na sua maioria, da mesma espécie (Figura 1), a árvore possui alguns ramos sem suporte nenhum e outros com altos valores de bootstrap.

ESPÉCIE	ABR	ARHGAP5	ARHGAP12	ARHGAP20	ARHGAP21	ARHGAP23	ARHGAP35	BCR	PIK3R2	SRGAP1	SRGAP3
<i>Cercocebus atys</i>		*									
<i>Mandrillus leucophaeus</i>		*									
<i>Papio anubis</i>		*									
<i>Macaca fascicularis</i>		*									
<i>Macaca mulatta</i>		*	*								
<i>Macaca nemestrina</i>		*								*	
<i>Chlorocebus sabaeus</i>								*			
<i>Colobus angolensis palliatus</i>											
<i>Rhinopithecus bieti</i>		*									
<i>Rhinopithecus roxellana</i>		*									
<i>Pongo abelii</i>		*									
<i>Gorilla gorilla gorilla</i>		*			*						
<i>Pan paniscus</i>		*		*	*						
<i>Pan troglodytes</i>		*			*						
<i>Homo sapiens</i>		*			*	*					
<i>Nomascus leucogenys</i>		*									
<i>Carlito syrichta</i>						-	-		-		-

Tabela 3- Presença e ausência de genes. Os asteriscos indicam os genes com supostas duplicações. Traço indica ausência. Células sem nenhuma indicação possui apenas uma cópia do gene.

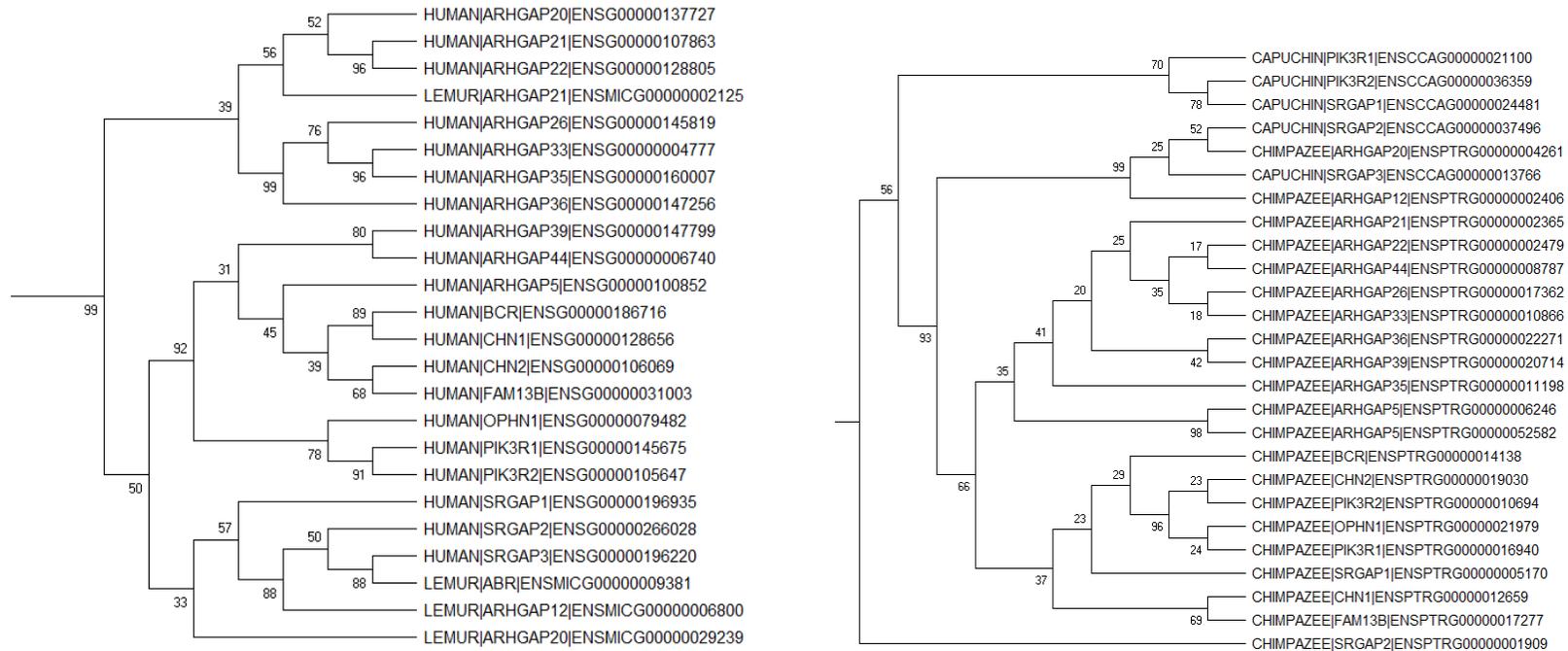


Figura 3 - Evolutionary analysis by Maximum Likelihood method. - TRECHOS DA ÁRVORE GERADA

5. DISCUSSÃO

Encontramos em vinte e três primatas todos os genes, o único que não possui os 23 genes é o *Tarsier*, em que faltam 4 genes (*ARHGAP23*, *ARHGAP35*, *SRGAP3* E *PIK3R2*), supomos que essa ausência seja devido a cobertura e o estágio de anotação que se encontra o genoma desse primata. Em algumas espécies, eventos adicionais de duplicação de genes produziu um acúmulo de sequências paralogas (duplicação de gene ancestral produz dois parálogos, são as cópias dentro da mesma espécie), que resulta em variação entre espécies em termos do número de cópias do gene.

Encontramos supostas cópias em 8 genes em 15 primatas. Verificamos que as supostas duplicações nos genes *ARHGAP12* no Rhesus, *ABR* e *ARHGAP23* nos seres humanos e o *ARHGAP20* no Bonobo encontradas no Ensembl se mostraram como artifício de anotação dos genes, isso ocorre quando o mesmo gene com notações distintas estão presentes em um mesmo banco de dado, podendo ser um transcrito alternativo ou mesmo sequências parciais dos genes anotadas independentemente pelos sistemas automatizados. Verificamos que isso ocorre nos conjuntos de genomas mais recentes, como o hg38 em humanos, onde existe uma sequência "alternativa" ou de "correção".

Encontramos também uma suposta duplicação no *Macaca nemestrina* no *SRGAP1* aparecendo tanto *Ensembl* quanto no *Genomicus*, no entanto o genoma da *M. nemestrina* foi sequenciado pelo método *shotgun* que quebra todo o genoma em pedaços, depois une eles em sequências conhecidas como *shotfall*. *Shotfall* em alguns casos podem se sobrepor, aparecendo então dois *shotfalls* que representa na verdade apenas uma região do genoma. Por tanto, acreditamos que não há uma duplicação e sim um erro de anotação.

Identificamos que a duplicação do *BCR* do *Vervet Monkey* está no mesmo cromossomo 19 em outro segmento do cromossomo (Chro19: 5,324,117-5,327,478), a cópia de referência se encontra na região Chro19: 2,566,190-2,702,920. Sugerimos que a duplicação possivelmente ocorreu num evento de recombinação, já que as cópias estão presentes no mesmo cromossomo. Descobrimos que a duplicação desse gene no Vervet só possui o domínio RhoGAP e apresenta um ORF. A proteína

codificada pela duplicação é bem menor possuindo 271 amino ácidos (aa), enquanto que a proteína produzida pela cópia original possui 1064 aa.

O *ARHGAP21* está envolvido nas atividades que envolvem o citoesqueleto de actina, incluindo migração, adesão, transporte intracelular e secreção de insulina (ROSA et al., 2018). Conferimos que as cópias do *ARHGAP21* do gorila e do bonobo estão no cromossomo 10 e no cromossomo 6. Não recuperamos cópias nos outros *Homininae*, inicialmente, na busca realizada pelo *UniProtKB*. Entretanto quando utilizamos o algoritmo *BLAT* do browser *UCSC* encontramos cópias em outros primatas da subfamília. Foram encontrados o *ARHGAP21* no chimpanzé e em humanos nos mesmos cromossomos do bonobo e do gorila indicando uma duplicação que ocorreu em um ramo ancestral desta linhagem. O *ARHGAP21* é fortemente expresso no cérebro sendo o tecido onde ocorre a maior expressão do gene segundo os dados no NCBI e do Protein Atlas⁷.

Salientamos, que embora encontramos a sequência correspondente em humanos (chr6:80064288-80070202) com o mesmo tamanho da sequência funcional (1958 aa) esta possui 3 códons de parada, o primeiro uma troca de uma serina na posição 23 por um códon de terminação. Na duplicação do chimpanzé e do bonobo o *ARHGAP21* há 3 códons de parada nas mesmas posições, e um específico em cada uma das espécies. Não encontramos códons de parada na duplicação do gorila. Poderíamos estar observando um evento do surgimento de pseudogenes na subfamília *Homininae* (*Pan* e *Homo*). O gorila apresenta quatro ORFs, e chimpanzé possui um, o bonobo possui quatro. É muito provável que o ser humano também apresente transcritos. Ressaltamos que mesmo cópias “defeituosas” têm a possibilidade de serem funcionais, citamos como exemplo os *SRGAP2s* onde as cópias interferem na funcionalidade da proteína original e essa interação parece estar envolvida com o surgimento de novidades adaptativas em humanos (DENNIS et al., 2012).

Enfatizamos a duplicação que ocorre com o *ARHGAP5* que apresenta duplicação nos seguintes primatas: *Macaca nemestrina*, *Rhinopithecus bieti*, *Rhinopithecus roxellana*, *Pongo abelii*, *Pan troglodytes*, *Nomascus leucogenys* (6 primatas). Duplicações essas encontradas na busca realizada no Ensembl. Todos estes primatas são do velho mundo (*Catarrhini*) analisados. Após uma busca por

BLAT da UCSC observou-se que a duplicação está presente nos 14 primatas do velho mundo dos 16 analisados, estando ausentes apenas nos *Chlorocebus sabaeus* e *Colobus angolensis palliatus*. Por estar amplamente difundido nos *Catarrhini* a duplicação deve ter ocorrido anterior a divergência das espécies aqui presentes, sendo posteriormente perdida nas linhagens das *Chlorocebus sabaeus* e *Colobus angolensis palliatus*. As múltiplas cópias encontradas no *ARHGAP5* pode ter ocorrido por duplicação insercional ou translocação, essas possibilidades também são válidas para o *ARHGAP21*.

Dois *scaffolds* contendo o *ARHGAP5* aparece no *R. roxellana* e *R. bieti*, no *Cercocebus atys*, e no *Mandrillus leucophaeus*. O *Nomascus leucogenys* apresenta três cópias do gene, estando situados nos cromossomos 1, 17 e 25 da espécie. No *Papio anubis*, e nas 3 espécies de Macaca analisadas as cópias aparecem nos cromossomo 7 e 3. Nos 5 grandes primatas (*Hominidae; Apes*) as cópias aparecem no cromossomo 14 e no cromossomo 7, sendo este último o gene duplicado.

Rastreamos as cópias do *ARHGAP5*, observamos que vem ocorrendo a degradação das mesmas. *Macaca mulata (Rhesus)* apresenta 12 códons de paradas, *Papio anubis (Baboon)* a duplicação possui 16 códons de parada. A duplicação da *R. roxellana* possui 10 códons de parada. Nos gorilas há 3 códons. Nos chimpanzés há 6 códons. Nos Bonobos há 3. Supreendentemente nos *H. sapiens* há apenas 2 códons de paradas, mas elas se encontram adjacentes. Esse padrão indica que a degeneração aconteceu rapidamente na maior parte dos grupos dentro dos *Catarrhini* com exceção dos *Hominoidea* onde a degeneração se deve a eventos mais recentes. O chimpanzé e o gorila possuem um transcrito, é muito provável que ocorram transcritos também nos bonobos e humanos.

O *ARHGAP5* decodifica uma fosfoproteína, na porção N-terminal ela possui vários motivos GTPase e na porção C-terminal se encontra o domínio RhoGap, o RhoGap possui atividade GAP para as proteínas Rac1, Cdc42 e RhoA. O *ARHGAP5* é um regulador de Rho que medeia a sinalização de adesão dependente de integrina em células cultivadas (BURBELO et al.,1995). E apresenta papel importante na morfogênese neuronal dependentes da actina em camundongos. Os camundongos sem o *ARHGAP5* funcional apresentaaram problemas no fechamento do tubo neural

craniano, na fusão de hemisférios na linha média do prosencéfalo e no fechamento da fissura óptica (BROUNS et al., 2000).

A maioria dos estudos sobre esse gene é focado no seu papel no câncer. Foi observado que o microRNA miR-494 está envolvido na desregulação do *ARHGAP5*, diminuindo a sua expressão, e subseqüentemente o aumento de invasão de gliomas, A expressão reduzida aumentava a expressão do *EGFR*. Esse estudo aponta o quão crítico é o *ARHGAP5* na regulação de tumores (KWAK et al, 2014). Já em cânceres pulmonares é observado o contrário, a expressão aumentada do *ARHGAP5* induz a invasão do câncer (WANG, et al.,2014). Já a superexpressão do gene pode levar a tumorigênese e a haplossuficiência inibe o tumor MMTV-Neu (HECKMAN-STODDARD, et al.,2009).

O crescimento dos ductos mamários são influenciados pelo *ARHGAP5*, camundongos heterozigotos para o alelo apresentaram formação menos acentuada de ductos mamários e aqueles sem o alelo não apresentaram a formação dos ductos, mas o fenótipo dos heterozigotos poderia ser recuperado quando atingida a maturidade sexual (CHAKRAVARTY, et al., 2003). A proteína *ARHGAP5* também é fortemente expressa nos testículos e o RNA na tireoide segundo o *The human protein atlas*.

Analisamos a árvore gerada pelo alinhamento de aproximadamente 529 domínios RhoGap, observamos que os domínios formaram grupos, na sua maioria, de parálogos da mesma espécie. Pensamos inicialmente que seria resultado da alta conservação do domínio e possivelmente em alguns casos eventos de conversão gênica, pois temos alguns ramos sem suporte nenhum e outros com altos valores de bootstrap. Mas existem muitas regiões específicas de cada gene que foram retiradas da análise, pois no alinhamento só foram utilizadas as sequências contendo somente a região do domínio de interesse, este artefato deve ser o motivo do agrupamento dos parálogos de uma mesma espécie. Estamos procurando uma metodologia alternativa para solucionar o problema.

Mudanças nos genes que ocorreram nos 15 milhões de anos na evolução dos grandes primatas devem ser responsáveis por algumas das características marcantes que distinguem as espécies. Nesse trabalho se destaca o *ARHGAP5* e o *ARGGAP21*. Duplicações de genoma deveriam ser facilmente identificadas através do surgimento

de seguimentos coincidentes dentro de uma filogenia de muitas famílias de genes. Infelizmente, esse sinal é complicado pela subsequente perda gradual e ganho de membros da família de genes (HURLES, 2004).

6. CONCLUSÃO

Quantificamos o número de genes contendo o domínio RhoGap em primatas, os quais possuem poucas diferenças. Vinte e três genes da família RHOGAP são expressos significativamente no cérebro, muitos deles ativos durante o neurodesenvolvimento, recuperamos todos esses genes na maioria dos primatas, verificamos duplicações e perdas específicas nos grandes primatas dos genes *ARHGAP5* e *ARHGAP21*, não foi possível associar perdas/duplicações a possíveis traços específicos nessa linhagem. A filogenia dos domínios RhoGap não foi informativa para determinar os padrões evolutivos da família.

Nosso trabalho fica limitado a espécies que possuíam o genoma completo e depende completamente do estágio de anotação do genoma. Uma anotação mais atualizada e homogênea facilitaria identificar melhor as sequências e as suas eventuais cópias. Nossas buscas preliminares encontraram duplicações por puro artifício da anotação dos genes e, também foram negligenciados alguns genes parálogos. Embora diversas alternativas foram utilizadas para refinar as buscas. Um estudo mais abrangente, ou seja, utilizando mais espécies poderia ajudar elucidar melhor a trajetória da evolução desses genes nos primatas.

7. REFERÊNCIAS

- BACON, Claire; ENDRIS, Volker; RAPPOLD, Gudrun A. The cellular function of srGAP3 and its role in neuronal morphogenesis. *Mechanisms of development*, v. 130, n. 6-8, p. 391-395, 2013.
- BROUNS, M. R. et al. The adhesion signaling molecule p190 RhoGAP is required for morphogenetic processes in neural development. *Development*, v. 127, n. 22, p. 4891-4903, 2000.
- BURBELO, Peter D. et al. p190-B, a new member of the Rho GAP family, and Rho are induced to cluster after integrin cross-linking. *Journal of Biological Chemistry*, v. 270, n. 52, p. 30919-30926, 1995.
- CÁCERES, Mario et al. Elevated gene expression levels distinguish human from non-human primate brains. *Proceedings of the National Academy of Sciences*, v. 100, n. 22, p. 13030-13035, 2003.
- CHARRIER, Cécile et al. Inhibition of SRGAP2 function by its human-specific paralogs induces neoteny during spine maturation. *Cell*, v. 149, n. 4, p. 923-935, 2012.
- CHAKRAVARTY, Geetika et al. p190-B RhoGAP regulates mammary ductal morphogenesis. *Molecular Endocrinology*, v. 17, n. 6, p. 1054-1065, 2003.
- DEMALI, Kris A.; WENNERBERG, Krister; BURRIDGE, Keith. Integrin signaling to the actin cytoskeleton. *Current opinion in cell biology*, v. 15, n. 5, p. 572-582, 2003.
- DENNIS, Megan Y.; EICHLER, Evan E. Human adaptation and evolution by segmental duplication. *Current opinion in genetics & development*, v. 41, p. 44-52, 2016.
- DENNIS, Megan Y. et al. Evolution of human-specific neural SRGAP2 genes by incomplete segmental duplication. *Cell*, v. 149, n. 4, p. 912-922, 2012.
- DOUGHERTY, Max L. et al. Transcriptional fates of human-specific segmental duplications in brain. *Genome research*, v. 28, n. 10, p. 1566-1576, 2018.
- FLORIO, Marta et al. Human-specific gene ARHGAP11B promotes basal progenitor amplification and neocortex expansion. *Science*, v. 347, n. 6229, p. 1465-1470, 2015.
- FORTNA, Andrew et al. Lineage-specific gene duplication and loss in human and great ape evolution. *PLoS biology*, v. 2, n. 7, p. e207, 2004.

GUERRIER, Sabrice et al. The F-BAR domain of srGAP2 induces membrane protrusions required for neuronal migration and morphogenesis. *Cell*, v. 138, n. 5, p. 990-1004, 2009.

HECKMAN-STODDARD, Brandy M. et al. Haploinsufficiency for p190B RhoGAP inhibits MMTV-Neu tumor progression. *Breast Cancer Research*, v. 11, n. 4, p. R61, 2009.

HURLES, Matthew. Gene duplication: the genomic trade in spare parts. *PLoS biology*, v. 2, n. 7, p. e206, 2004.

JAFFE, Aron B.; HALL, Alan. Rho GTPases: biochemistry and biology. *Annu. Rev. Cell Dev. Biol.*, v. 21, p. 247-269, 2005.

KWAK, Seo-Young et al. Ionizing radiation-inducible miR-494 promotes glioma cell invasion through EGFR stabilization by targeting p190B rhoGAP. *Biochimica et Biophysica Acta (BBA)-Molecular Cell Research*, v. 1843, n. 3, p. 508-516, 2014.

LAMARCHE, Nathalie; HALL, Alan. GAPs for rho-related GTPases. *Trends in Genetics*, v. 10, n. 12, p. 436-440, 1994.

LEVINE, Michael; TJIAN, Robert. Transcription regulation and animal diversity. *Nature*, v. 424, n. 6945, p. 147, 2003.

MOON, Sun Young; ZHENG, Yi. Rho GTPase-activating proteins in cell regulation. *Trends in cell biology*, v. 13, n. 1, p. 13-22, 2003.

PAIXAO-CORTES, Vanessa R.; SALZANO, Francisco M.; BORTOLINI, Maria Catira. Origins and evolvability of the PAX family. In: *Seminars in cell & developmental biology*. Academic Press, 2015. p. 64-74.

PECK, Jeremy et al. Human RhoGAP domain-containing proteins: structure, function and evolutionary relationships. *FEBS letters*, v. 528, n. 1-3, p. 27-34, 2002.

ROJAS, Ana Maria et al. The Ras protein superfamily: evolutionary tree and role of conserved amino acids. *J cell Biol*, v. 196, n. 2, p. 189-201, 2012.

ROSA, Lucas RO et al. ARHGAP21 as a master regulator of multiple cellular processes. *Journal of cellular physiology*, v. 233, n. 11, p. 8477-8481, 2018.

TYLER-SMITH, Chris; XUE, Yali. Sibling rivalry among paralogs promotes evolution of the human brain. *Cell*, v. 149, n. 4, p. 737-739, 2012.

WANG, J. et al. Downregulation of miR-486-5p contributes to tumor progression and metastasis by targeting protumorigenic ARHGAP5 in lung cancer. *Oncogene*, v. 33, n. 9, p. 1181, 2014.

YAMAZAKI, Daisuke et al. srGAP1 regulates lamellipodial dynamics and cell migratory behavior by modulating Rac1 activity. *Molecular biology of the cell*, v. 24, n. 21, p. 3393-3405, 2013

REFERÊNCIA ONLINE

1 – Ensembl – <https://www.ensembl.org/index.html>

2 – InterPro – <http://www.ebi.ac.uk/interpro/>

3 – UniProt – <https://www.uniprot.org/>

4 – Genomicus – <http://www.genomicus.biologie.ens.fr/genomicus-92.01/cgi-bin/search.pl>

5 – GenBank – <https://www.ncbi.nlm.nih.gov/>

6 – Genome Browser UCSC – <https://genome.ucsc.edu/>

7 – The Protein Atlas – <https://www.proteinatlas.org/>