

GUILHERME BASTOS GOMES

Classificação e Evolução de GTPases da classe TRAFAC: a história evolutiva das septinas/parasseptinas

Dissertação apresentada ao Programa de Pós-Graduação em Microbiologia do Instituto de Ciências Biomédicas da Universidade de São Paulo para obtenção do Título de Mestre em Ciências

SÃO PAULO
2020

GUILHERME BASTOS GOMES

Classificação e Evolução de GTPases da classe TRAFAC: a história evolutiva das septinas/parasseptinas

Dissertação apresentada ao Programa de Pós-Graduação em Microbiologia do Instituto de Ciências Biomédicas da Universidade de São Paulo para obtenção do Título de Mestre em Ciências.

Área de Concentração: Microbiologia

Orientador: Prof. Dr. Robson Francisco de Souza

Versão Original

SÃO PAULO
2020

CATALOGAÇÃO NA PUBLICAÇÃO (CIP)
Serviço de Biblioteca e informação Biomédica
do Instituto de Ciências Biomédicas da Universidade de São Paulo

Ficha Catalográfica elaborada pelo(a) autor(a)

Bastos Gomes, Guilherme
Classificação e Evolução de GTPases da classe
TRAFAC: a história evolutiva das
septinas/parasseptinas / Guilherme Bastos Gomes;
orientador Robson Francisco de Souza. -- São Paulo,
2020.
100 p.

Dissertação (Mestrado)) -- Universidade de São
Paulo, Instituto de Ciências Biomédicas.

1. classificação. 2. evolução. 3. GTPases. 4.
septinas. 5. parasseptinas. I. Francisco de Souza,
Robson, orientador. II. Título.

UNIVERSIDADE DE SÃO PAULO
INSTITUTO DE CIÊNCIAS BIOMÉDICAS

Candidato(a): Guilherme Bastos Gomes

Título da Dissertação/Tese: CLASSIFICAÇÃO E EVOLUÇÃO DE GTPASES DA CLASSE TRAFAC: A HISTÓRIA EVOLUTIVA DAS SEPTINAS/PARASSEPTINAS

Orientador: Prof. Dr. Robson Francisco de Souza

A Comissão Julgadora dos trabalhos de Defesa da Dissertação de Mestrado/Tese de Doutorado, em sessão pública realizada em/...../....., considerou o(a) candidato(a):

() **Aprovado(a)** () **Reprovado(a)**

Examinador(a): Assinatura:
Nome:
Instituição:

Examinador(a): Assinatura:
Nome:
Instituição:

Examinador(a): Assinatura:
Nome:
Instituição:

Presidente: Assinatura:
Nome:
Instituição:



M

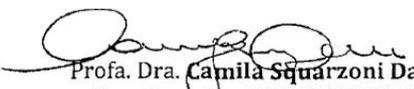
Cidade Universitária "Armando de Salles Oliveira", Butantã, São Paulo, SP - Av. Professor Lineu Prestes, 2415 - ICB III - 05508 000
Comissão de Ética em Pesquisa - Telefone (11) 3091-7733 - e-mail: cep@icb.usp.br

CERTIFICADO DE ISENÇÃO

Certificamos que o Protocolo CEP-ICB nº **944/2018** referente ao projeto intitulado: ***"Genômica comparativa de famílias de peptidases que atuam em sistemas toxina-antitoxina e sistemas de secreção bacterianos"*** sob a responsabilidade de ***Guilherme Bastos Gomes*** e orientação do(a) Prof.(a) Dr.(a) ***Robson Francisco de Souza***, do Departamento de Microbiologia, foi analisado pela **CEUA** - Comissão de Ética no Uso de Animais e pelo **CEPSH** - Comitê de Ética em Pesquisa com Seres Humanos, tendo sido deliberado que o referido projeto não utilizará animais que estejam sob a égide da Lei nº 11.794, de 8 de outubro de 2008, nem envolverá procedimentos regulados pela Resolução CONEP nº 466/2012.

São Paulo, 21 de março de 2018.


Profa. Dra. **Luciane Valéria Sita**
Coordenadora CEUA ICB/USP


Profa. Dra. **Camila Squarizoni Dale**
Coordenadora CEPSH ICB/USP



Dedico este trabalho a minha família que sempre me proveu o bastante para manter minha caminhada rumo ao conhecimento. Também aos gigantes dos quais subi nos ombros para enxergar mais longe.

AGRADECIMENTOS

O presente trabalho foi realizado com apoio da Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - Brasil (CAPES) - Projeto Biocomputacional - Edital nº 51/2013 - Biologia Computacional entre 01/04/2018 e 31/12/2018, e Regulamento do PROEX – IES Públicas, entre 01/03/2019 e 31/03/2020, portanto agradeço todo o apoio e suporte da CAPES. Também agradeço muito ao Professor Dr. Robson Francisco de Souza que aceitou me acompanhar e guiar ao longo dessa jornada, garantindo boas conversas e discussões a respeito de ciência e evolução, sempre buscando excelência e perfeição, dedicando seu tempo, disposição, e garantindo uma infraestrutura robusta para que esta pesquisa pudesse ser realizada.

Aos meus companheiros de laboratório que me acompanharam e guiaram nesta jornada científica. Um especial agradecimento aos Drs. Gianluca Nicastro e Gilberto Kaihama, grandes cientistas que dedicaram seu tempo para me ensinarem sobre a academia e os processos relacionados. Obrigado Aureliano Guedes, Rodolfo Alvarenga e Gabriel Hueck, por tantos momentos de descontração e desenvolvimento, verdadeiros amigos, sem vocês eu não teria sido capaz de terminar este trabalho. Agradeço a Prof. Dra. Cristiane Guzzo, por nos prover uma ótima atmosfera de trabalho e desenvolvimento científico, também à Dra. Ethel Bayer, e aos seus estudantes que logo se tornaram grandes amigos, obrigado: Tânia, Gabriel, Angel, Iris, Nathália, Aline, Camila, Julia, Stephanie, Ana Paula, Anacleto, Daniel e Alexandre, pelos melhores momentos e conversas durante os bandejões, bolos e cafés.

Aos meus familiares, principalmente à minha mãe, meu pai que sei que estaria orgulhoso de mim, meus irmãos, meus padrinhos e minha madrinha, e a todos que me deram suporte emocional para a realização deste trabalho.

Aos meus amigos que ficam e aos que se foram, mas que sempre me apoiaram custe o que custar: Alexandre, Cody, Marcel, Maurício, Matheus e André, que nos deixou muitas saudades.

Aos meus professores.

Ao meu querido amigo, que considero um irmão, Mateus que me ajudou a minimizar o uso de orações subordinadas ao longo do texto e por checar minha sanidade mental vez ou outra.

E meu maior e mais importante agradecimento à minha sensacional esposa/namorada/amiga/de facto, Nathália, que me deu todo suporte necessário para começar, desenvolver e concluir minha dissertação. Seu apoio e todo o tempo que dedicou para me ajudar significa tudo para mim, você é luz nas noites mais escuras e sombra preciosa nos dias mais quentes, eu a amo muito, muito obrigado.

“Nada na Biologia faz
sentido exceto à luz da
Evolução”
Theodosius Dobzhansky

RESUMO

GOMES, G. B. **Classificação e Evolução de GTPases da classe TRAFAC: a história evolutiva das septinas/parasseptinas.** 2020. 99 f. Dissertação (Mestrado em Ciências) – Instituto de Ciências Biomédicas, Universidade de São Paulo, São Paulo, 2020.

Existem diversas proteínas capazes de ligar e hidrolisar moléculas de nucleotídeo fosfato e grande parte delas são necessárias para a manutenção da vida. As GTPases são proteínas do *fold* P-loop NTPases com capacidade de catalisar moléculas de GTP e ATP para efetuar e regular processos celulares como a iniciação e alongamento da tradução, remodelamento ou organização de membranas, sinalização, síntese de nutrientes e imunidade. As GTPases são divididas em duas classes com base em suas sinapomorfias: TRAFAC e SIMIBI. Neste trabalho nos concentramos na história evolutiva de GTPases da classe TRAFAC (de *translation factors*), que agrega diversas famílias de proteínas altamente divergentes que partilharam ancestralidade comum como: os fatores de tradução, a superfamília Ras de transdutores de sinal, a família FeoB de transportadores de ferro, as dinaminas e as septinas que participam de processos de remodelamento e tráfego de vesículas, entre outras. A última classificação massiva desses grupos de GTPases baseada nas estruturas e caracteres compartilhados a nível de sequência foi proposta no início deste século, precisamente em 2002, e por esse motivo esse trabalho objetivou comparar as sequências e estruturas disponíveis dessas GTPases desenvolvendo uma metodologia reprodutível baseada na similaridade entre as sequências e em perfis de sequências baseados em HMMs, seguidas de análises de redes, de modo a construir uma classificação evolutiva natural e atualizada e inferir hipóteses evolutivas a respeito da origem e diversificação desses grupos de genes amplamente divergentes. Além disso, aprofundamos o debate sobre a história evolutiva das septinas nos organismos eucariotos e da relação evolutiva desse grupo com as famílias semelhantes as septinas como GIMAP e Toc, e de um grupo relacionado ainda não caracterizado experimentalmente das parasseptinas bacterianas.

Palavras-chave: Classificação. Evolução. GTPases. Septinas. Parasseptinas.

ABSTRACT

GOMES, G. B. **Classification and Evolution of the TRAFAC class GTPases: the evolutionary history of septins/paraseptins.** 2020. 99 f. Dissertação (Mestrado em Ciências) – Instituto de Ciências Biomédicas, Universidade de São Paulo, São Paulo, 2020.

There are a diverse range of proteins which are able to bind and hydrolyse nucleotide triphosphate molecules and a majority of them are required to the maintenance of life. GTPases are proteins from the P-loop NTPase fold that are able to catalyze GTP and ATP molecules in order to effectuate or regulate several cellular processes such as translation elongation and initiation, membrane remodelling and reorganisation, signaling, nutrient synthesis and immunity. There are two classes of GTPases built on the basis of their synapomorphies: TRAFAC and SIMIBI. In this work we focus on the evolutionary history of the TRAFAC (after translation factor) class of GTPases, that comprise several highly diversified protein families including the translation factors, the Ras extended superfamily of signal transducers, the FeoB family of iron transporters, dynamins and septins that participate in processes of membrane remodelling and vesicle trafficking, among others. The last massive classification of these groups of GTPases based on structure and shared characters was released in the beginning of this century, precisely in 2002, and for this reason this work aims to compare available sequences and structures of these GTPases by developing a reproducible methodology based on the sequence similarity HMM sequence profiles, followed by sequence similarity networks, in order to build an updated and natural evolutionary classification and generate evolutionary hypothesis of the main events regarding the origins and diversification of these groups of highly divergent genes. Beyond that, we dig deeply into the evolutionary story of septins in eukaryotes and into the evolutionary relation between septins and the other septin-like families known as GIMAP and Toc, and with a related group not yet experimentally characterized named bacterial paraseptins.

Key-words: Classification. Evolution. GTPases. Septin. Paraseptin.

Lista de Figuras

Figura 1 - Representação gráfica das P-loop GTPases da classe TRAFAC com ênfase nas estruturas secundárias e motivos conservados.	16
Figura 2 - Representação gráfica da estrutura secundária das duas classes de GTPases	19
Figura 3 - Arquitetura geral das Septinas com ênfase no domínio GTPase e Diagrama da classificação das Septinas	22
Figura 4 - Filogenia das famílias de Septinas e a topologia	26
Figura 5 - Arquitetura de domínio das GIMAPs presentes em humanos	27
Figura 6 - Árvore filogenética dos eucariotos (Eukarya) e descrição filogenética das Asgard	29
Figura 7 - Workflow simplificado da análise desenvolvida	31
Figura 8 - Estrutura da GTPase da família ERA 5IER_A	37
Figura 9 - Estrutura do domínio GTPase da proteína 5IRR_A	31
Figura 10 - Representação da estrutura em 3D da proteína 3V70_A	40
Figura 11 - Representação da estrutura em 3D da proteína 3BB3_A	41
Figura 12 - Alinhamento de duas sequências da família GIMAP	45
Figura 13 - Gráfico em barras mostrando a frequência relativa de arquiteturas de domínios da comunidade 1	48
Figura 14 - Rede de similaridade de sequências do supergrupo das P-loop GTPases.	51
Figura 15 - Representação do espaço de modelos do nosso banco de dados de homólogos de GTPases	55
Figura 16 - Árvore das distâncias entre os modelos de pHMMs das famílias de GTPases	59
Figura 17 - Árvore filogenética dos eucariotos	69
Figura 18 - Árvore filogenética de uma das famílias de GIMAP	72

Sumário

1 INTRODUÇÃO	12
1.1 P-loop NTPases	12
1.2 A superclasse das P-loop GTPases	14
1.2.1 Classes de P-loop GTPases	14
1.2.1.1 Núcleo do domínio GTPase: cinco motivos conservados a nível de sequência	15
1.2.1.2 Classe SIMIBI	17
1.2.1.3 Classe TRAFAC	18
1.3 Evolução das GTPases	20
1.4.1 Septinas	21
1.4.2 Paraseptinas	24
1.5 Mudanças no paradigma das relações evolutivas entre Eucariotos e Procariotos	28
2 MATERIAL E MÉTODOS	30
3 JUSTIFICATIVA	33
3.1 Porquê compreender a evolução dessas famílias de proteínas?	33
4 OBJETIVOS	34
4.1 Objetivos Gerais	34
4.2 Objetivos específicos	34
5 RESULTADOS E DISCUSSÃO	35
5.1 Coleta de homólogos e características das famílias de interesse	35
5.2 Classificação das GTPases	42
5.2.1 Identificação e separação das famílias de P-loop GTPases	42
5.2.2 Agrupamento hierárquico e a relação entre as famílias de GTPases da classe TRAFAC	57
5.3 Superfamília das Septinas	63
5.3.1 Ampla distribuição da superfamília das Septinas no grupo coroa dos Eucariotos	66
5.4 GIMAP	71
6 CONCLUSÃO	73
7 REFERÊNCIAS	75
Anexo - Classificação das GTPases estruturalmente caracterizadas (PDBs classificados)	85

1 INTRODUÇÃO

1.1 P-loop NTPases

Diversas proteínas são capazes de ligar e hidrolisar moléculas de nucleotídeo trifosfato e muitas são extremamente necessárias para a manutenção da vida. Essas proteínas podem ser classificadas como parte de diversos dobramentos, sendo eles: o *fold* de ligação a dinucleotídeos *Rossmann-like* e os *folds* relacionados a Tubulina/Ftsz; o *fold* P-loop NTPase; o *fold* das proteínas Kinases; o *fold* histidina kinase/HSP90/Topoll; e o *fold* das HSP70/RNase H (Gorbalenya et al. 1990; Milner-White et al. 1991; Saraste et al. 1990; Schulz 1992; Schweins et al. 1994; Vetter et al. 1999).

O *fold* P-loop NTPase é conhecido por possuir uma estrutura de *loop* localizada na região amino-terminal. P-loop é uma abreviação para *Phosphate-binding-loop* ou alça de ligação a fosfato, atividade chave no processo de interação com o grupo fosfato-gama de nucleotídeos. Esta região também é conhecida como Walker A e possui o motivo $GxxxxGK[T/S]$ entre a primeira fita β conservada do *fold* e a primeira α -hélice, que cercam um *loop* flexível cuja função é posicionar corretamente o trifosfato do nucleotídeo para atividade catalítica dessas proteínas (Saraste, M. 1990; Milner-White, E. J., 1991; Walker, J. E. 1982). Distal ao motivo anterior, o motivo Walker B é caracterizado pela presença de um ácido aspártico ou ácido glutâmico (menos comum) conservado, situado no fim de uma fita-beta e precedendo um *loop* que faz pontes de hidrogênio com um íon de magnésio (Mg^{+2}) (Walker, J. E. 1982).

As P-loop NTPases, juntamente com as proteínas do *fold* *Rossmann-like*, são provavelmente as duas famílias de proteínas com a distribuição taxonômica mais ampla conhecidas até o momento. Ambas famílias possuem semelhanças na composição e organização dos elementos de estrutura secundária do núcleo conservado de seus domínios catalíticos, e também em seus motivos de ligação ao grupo fosfato, o que sugere que sua descendência a partir de um ancestral comum, à exclusão de outros grupos de proteínas, não pode ser descartada. Embora a

literatura sempre tenha reconhecido algumas semelhanças estruturais, esses dois grandes grupos de proteínas sempre foram consideradas como derivadas de origens independentes no mundo pré-LUCA (Last Universal Common Ancestor, ou Último Ancestral Comum Universal), devido à aparente ausência de identidade entre elas. Uma análise recente, porém, propõe uma possível relação dessas duas famílias, que podem ter sido derivadas a partir de um ancestral comum pré-LUCA, caracterizado por um enovelamento (*fold*) ancestral Beta-Alfa-Beta, que inclui o motivo Walker A e é compartilhado pelas duas famílias. Esses autores, porém, admitem que, mesmo com o grande volume de dados de sequência acumulados e a análise cuidadosa das similaridades entre as duas famílias, ainda é difícil descartar a possibilidade de convergência evolutiva desses *fold*s que evoluíram há tanto tempo e que perderam a maior parte do sinal que evidenciaria um ancestral comum para ambos os grupos (Longo et al. 2020).

Até o momento, 26 grupos de P-Loop NTPase já foram identificados e classificados, e pelo menos 6 dentre estes podem ter se originado antes de LUCA (Iyer et al. 2004). Análises de sequência e estrutura sugerem que um dos primeiros eventos evolutivos de diversificação na evolução das P-loop NTPases resultou na cladogênese de dois grupos. O primeiro, denominado ASCE (A*dditional* s*trand* c*onserved* E, ou fita adicional, com um resíduo “E” conservado), foi estruturalmente caracterizado por uma fita-beta adicional na folha central, localizada entre a fita-beta que precede o loop P (Walker A) e a fita-beta que precede o motivo Walker B, e a nível de sequência por um resíduo adicional de ácido glutâmico (E) catalítico no mesmo motivo Walker B (*hhhhE/DE*) (Iyer et al. 2004).

O outro grupo é conhecido pela denominação KG (de Kinase – GTPase, ou Quinase - GTPase) inclui as quinases, GTPases e outros grupos de proteínas relacionados, que compartilham características estruturais, como o posicionamento dos *loops* Walker (A e B), responsáveis pelo acesso e estabilização da ligação ao trifosfato nucleotídeo, analisadas neste trabalho e detalhadas à frente (Iyer et al. 2004).

1.2 A superclasse das P-loop GTPases

A superclasse das GTPases e proteínas relacionadas constitui um grupo monofilético dentro do *fold* das P-loop NTPases. A relação dos membros desse grupo, que inclui famílias extremamente divergentes e com valores de similaridade par-a-par no limiar dos métodos de detecção de similaridade, é suportado por vários estudos que empregaram a comparação de estruturas e a análise de motivos conservados na sequência do núcleo catalítico destas proteínas. Membros desse grupo conservam a capacidade de ligação a nucleotídeos, geralmente GTP, e muitos são capazes de clivar a ligação entre o pirofosfato e o fosfato γ , atividade que é essencial para conferir a estas proteínas a característica de “interruptor molecular” (molecular switch), ou seja, a capacidade de transitar entre dois estados (“ativo” e “inativo”) que permitem aos membros desse grupo atuar como reguladores de diversos processos, como pontes em redes de transdução de sinal ou ainda na dinâmica de estruturas do citoesqueleto.

1.2.1 Classes de P-loop GTPases

A base atual para a classificação das GTPases e proteínas relacionadas foi introduzida por Leipe, em 2002, que subdividiu o grupo em duas classes, que possuem características estruturais e de sequência exclusivos de cada subgrupo.

O primeiro subgrupo, denominado TRAFAC (de *translation-factor related*, ou proteínas relacionadas a fatores de tradução), compreende a maioria dos membros da superfamília com atividade de GTPase conhecidas, incluindo diversas famílias de fatores de tradução, como fatores de iniciação e alongamento; a superfamília Ras, com membros que atuam na transdução de sinais e motilidade celular em procariotos; a superfamília das Septinas, que desempenham importante papel na divisão celular de eucariotos; as dinaminas, envolvidas em remodelamento e tráfego de vesículas; e algumas ATPases, como as miosinas e quinesinas, que funcionam como motores celulares e no transporte intracelular em eucariotos.

A segunda classe, é conhecida como SIMIBI, um acrônimo derivado do nome de seus três grandes grupos: proteínas de reconhecimento de Sinal (SRP), as famílias de proteínas similares a MInD e a BioD (Leipe et al., 2002). O grupo das SIMIBI inclui GTPases de reconhecimento de sinais associados à secreção (SRPs), a superfamília MRP/MinD, que também inclui os homólogos das proteínas de particionamento de cromossomos conhecidas como ParA, e diversas outras enzimas metabólicas, como BioD e outras enzimas da via de síntese de biotina, adenilosuccinato sintetase, e formil tetrahidrofolato ligase.

Cada classe de GTPases é caracterizada por uma estrutura ancestral que as define, ou seja, por elementos de estrutura secundária que são compartilhados por todos os clados de cada classe, por exemplo, as TRAFAC são definidas estruturalmente por possuírem duas fitas betas, a segunda (β 2-Figura 1) e a terceira (β 3-Figura 1), posicionadas no sentido anti-paralelo uma à outra. Enquanto na classe das SIMIBI, a topologia dos elementos equivalentes se posicionam de tal modo que as mesmas fitas se encontrem paralelas uma à outra (conforme detalhado mais à frente) (Leipe, 2002).

1.2.1.1 Núcleo do domínio GTPase: cinco motivos conservados a nível de sequência

Dentro do núcleo catalítico de todas as proteínas da superfamília das GTPases, é possível encontrar motivos altamente conservados, denominados G1 a G5, que compreendem os resíduos diretamente envolvidos com a ligação ao substrato e a atividade de hidrolase. O primeiro destes motivos, denominado G1, se localiza na primeira fita-beta, e inclui o motivo Walker A no primeiro *loop* após essa fita. O motivo G2 é o *loop* posicionado na região aminoterminal da segunda fita-beta. G3 está localizado na região C-terminal da terceira fita-beta e inclui o variante do motivo Walker B que é característico das proteínas da família das GTPases (*hhhhDxxG*). G4 é a quinta fita-beta contendo os resíduos conservados *xKxD*, e G5 é a fita 6 geralmente associada a uma serina conservada no décimo

loop, embora a posição desse resíduo varie bastante entre os subgrupos de GTPases, conforme descrito na Figura 1.

Além disso, as TRAFAC usualmente possuem um resíduo conservado que pode ser uma Treonina ou uma Serina (T/S), localizada no *loop* entre a segunda e a terceira fitas beta, responsável por fazer uma ponte de hidrogênio com um cátion, o qual é necessário para a atividade de hidrólise da molécula de GTP. Nas SIMIBI, outro resíduo é responsável pela mesma função, conforme descrito a seguir.

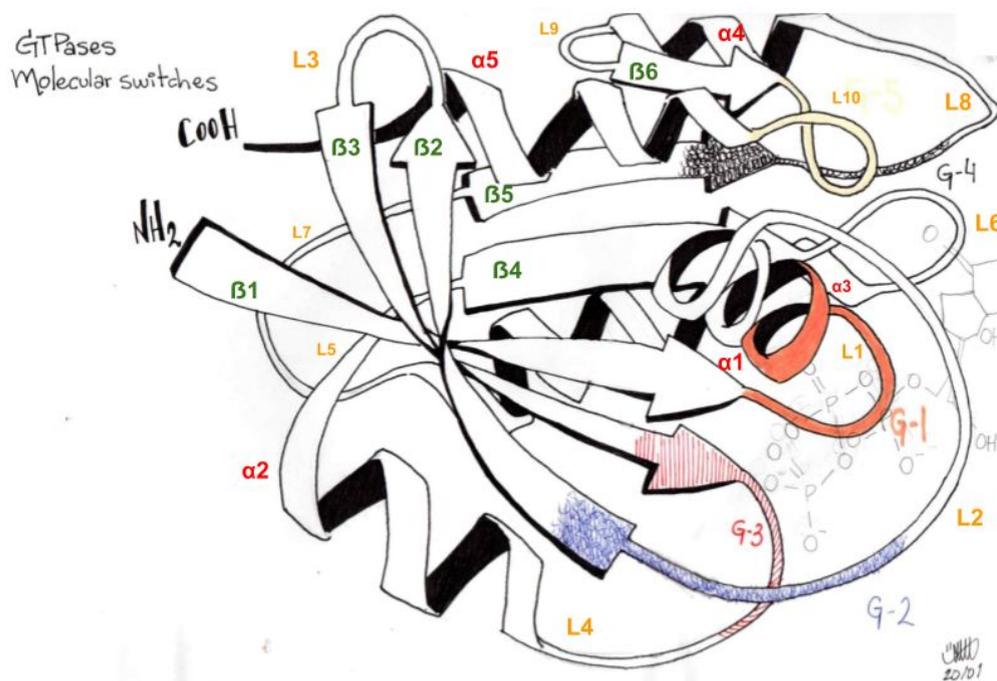


Figura 1. Representação gráfica desenhada com nanquim e caneta da menor estrutura conhecida das P-loop GTPases da classe TRAFAC com ênfase nas estruturas secundárias e motivos conservados dessa classe. Em verde estão descritas as fitas-beta, da primeira até a sexta, em vermelho estão descritas as alfas-hélices, da primeira até a quinta, e em amarelo os loops que se posicionam entre essas estruturas secundárias, do primeiro até o décimo, conforme orientação da região N-terminal para C-terminal. Nesta representação é possível observar que a segunda fita-beta ($\beta 2$) se posiciona em orientação antiparalela à terceira fita-beta ($\beta 3$), que precede o motivo de sequência G3 ou Walker B, característica evidente em todas as GTPases da classe TRAFAC. Hachurados estão os motivos de sequência G1-G5, em laranja o motivo G1, ou Walker A, em azul G2, em vermelho G3 ou Walker B, em preto G4 ou xKxD, em amarelo G5, um motivo mais divergente entre as famílias de GTPases. O título “GTPases, Molecular Switches”, em português “GTPases, interruptores moleculares” se refere a atividade de regulação de proteínas da família Ras, cuja ligação e hidrólise de GTP controlam diversos mecanismos celulares. Em grafite está representada a molécula de Guanosina trifosfato com os fosfatos posicionados em relação aos motivos com capacidade de interação e catálise dessa molécula. A assinatura da representação é do autor deste trabalho e essa versão do desenho foi concluída em 20/01/2019.

1.2.1.2 Classe SIMIBI

As famílias de proteínas da classe SIMIBI são estruturalmente caracterizadas por possuírem as duas fitas-beta paralelas à fita-beta do motivo Walker B orientadas no mesmo sentido, ao contrário das TRAFAC, cuja a fita que precede a fita-beta do motivo Walker B se posiciona em sentido antiparalelo a fita que constitui o motivo Walker B (Figura 2).

Em nível de sequência também é possível observar diferenças entre SIMIBI e TRAFAC, a primeira delas é um resíduo de ácido aspártico na região C-terminal da segunda fita-beta, presente nas proteínas da classe SIMIBI que contribui na formação de uma ponte de hidrogênio com um íon de Mg, conforme evidenciado pelas estruturas de SRP, MinD, ArsA e na proteína ferro-nitrogenase NifH (Montoya et al. 2000; Zhou et al. 2000; Hayashi et al. 2001; Schindelin et al. 1997). A função desse resíduo é similar a do resíduo de treonina presente no loop anterior a segunda fita-beta nas TRAFAC (Figura 1 - G2).

As proteínas da classe SIMIBI também conservam uma variação do motivo Walker A, no qual é possível encontrar um resíduo de Glicina adicional conservado entre os resíduos de Glicinas normalmente encontrados em P-loop NTPases, cujo consenso é estendido para GxxGxGKT/S. Outra característica desta classe, descrita por Leipe e colaboradores, em 2002, é a presença de um resíduo de ácido aspártico na região N-terminal da fita-beta que contém o motivo Walker B, equivalente à região N-terminal da terceira fita-beta da Figura 2, logo após o terceiro *loop*, que está posicionado de forma oposta aos motivos que interagem com a molécula de nucleotídeo fosfato e que motivou a sugestão de uma função de estabilização estrutural para este resíduo.

Também é possível observar que os resíduos presentes no motivo G4 (geralmente xKxD), responsável por prover especificidade a base da guanina e muito conservado entre as proteínas da classe TRAFAC, são bem mais variáveis na classe SIMIBI, principalmente nas famílias que atuam como ATPases, como observado na família MinD/ParA que compreende ATPases, onde a atividade hidrolítica e a alternância entre o estado ligado a ATP e a ADP influenciam

diretamente o comprimento de polímeros que atuam na membrana celular ou no particionamento de cromossomos (Lutkenhaus, 2012).

1.2.1.3 Classe TRAFAC

Membros das famílias de proteínas da classe TRAFAC possuem uma Treonina/Serina extremamente conservada que é responsável por fazer uma ponte de hidrogênio com um íon Mg^{+2} necessário para a atividade de GTPase (Figura 1). Além disso, quase todas as famílias da classe TRAFAC possuem um resíduo de Serina posicionado no *loop* da última fita-beta do núcleo enzimático GTPase, envolvido na ligação com a Guanina. As proteínas TRAFAC também possuem uma estrutura plesiomórfica com uma fita-beta em sentido anti-paralelo a todas as outras do núcleo catalítico das GTPases, característica que as diferencia das GTPases da classe SIMIBI, e de outros grupos de proteínas do *fold P-loop* NTPases (Figura 2). Diversas famílias da classe TRAFAC também possuem uma extensão α -hélica na região C-terminal (Figura 2), geralmente responsável por interagir com outras proteínas ou associarem-se com atividades na membrana celular.

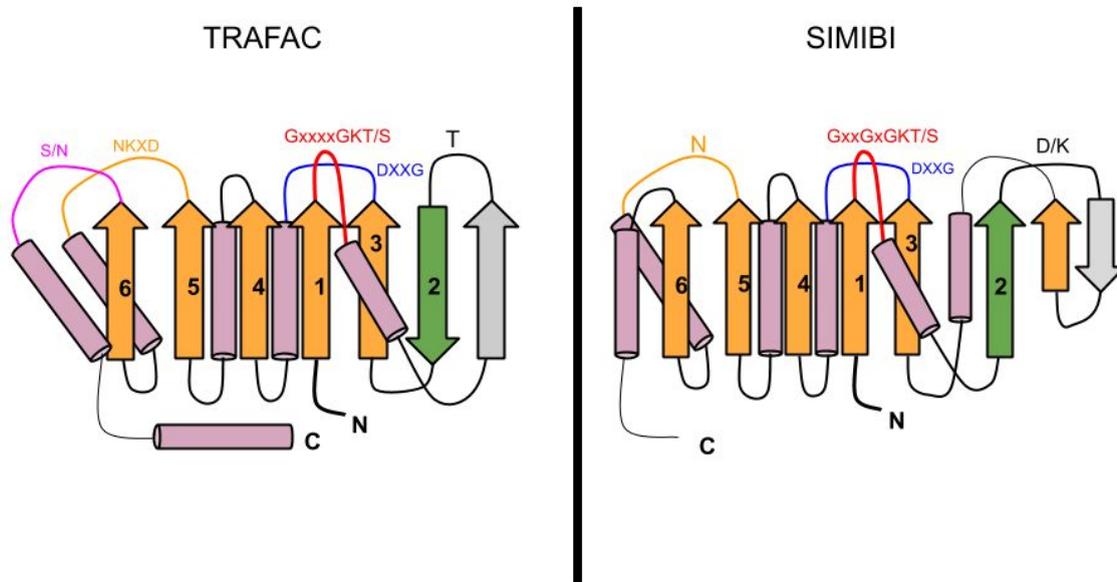


Figura 2. Representação gráfica da estrutura secundária plesiomórfica das duas classes de GTPases. À esquerda está uma representação da classe TRAFAC, à direita da classe SIMIBI, ambas com ênfase (verde) na fita-beta que precede a fita-beta do motivo Walker B (DxxG, conhecido como G3, em ambas). O ancestral TRAFAC possuía uma fita-beta antiparalela precedente a fita-beta do motivo Walker B (G3), enquanto o ancestral SIMIBI possui a fita-beta equivalente em sentido paralelo a fita-beta que precede o motivo Walker B. Em laranja, e representadas como fitas, estão as fitas- β presentes na estrutura compartilhada entre todos os integrantes de ambas as classes, em cinza estão representadas fitas-betas numa posição variável entre as proteínas de cada classe, ou seja, algumas famílias não conservam esta região ou possuem inserções de outras estruturas secundárias nesta posição. Em lilás, e com representação em cilindros, estão as α -hélices conservadas entre as fitas-betas da estrutura $\alpha\beta$ compartilhada entre todas essas proteínas. Os loops estão representados como linhas de diferentes cores de acordo com os motivos conservados nestas posições, o tamanho da linha demonstra a posição do loop no espaço dividido pela posição das fitas-betas. Em vermelho está o motivo Walker A ou G1; Em azul, Walker B ou G3; Em laranja o motivo de especificidade à molécula de GTP (xKxD) ou G4; Em magenta o motivo menos conservado entre todas as GTPases (G5), mas que também estabiliza a molécula de GTP durante a atividade catalítica dessas proteínas; Em preto, após a segunda fita- β ancestral está o resíduo conservado de Treonina (TRAFAC) ou Ácido Aspártico/Lisina (SIMIBI) também conhecido como G2. Um “N” em negrito aponta para a região N-terminal das proteínas, enquanto um “C” indica a região C-terminal.

1.3 Evolução das GTPases

A classificação mais recente de P-loop GTPases foi introduzida em 2002 por Leipe e colaboradores. Esses autores descreveram 8 superfamílias de GTPases TRAFAC e 5 da classe SIMIBI. Compreender a existência de todos esses grupos foi importante para a análise proposta por este trabalho, para delimitar o espaço entre as sequências e seus respectivos processos evolutivos.

Após a publicação da classificação introduzida por Leipe e colaboradores, milhares de novos genomas de eucariotos e procariotos foram depositados nos bancos de dados públicos e a análise de novas hipóteses sobre as relações das GTPases se tornaram viáveis. O sequenciamento de novas linhagens de arqueas e bactérias, a partir de amostras metagenômicas, como as Asgard archaea (Zaremba-Niedzwiedzka et al. 2017), as DPANN (Eme et al. 2018) e a nova divisão de bactérias conhecida como CPR (Méheust et al. 2019), têm aquecido o debate sobre a diversidade de procariotos e a origem do grupo dos Eucariotos, afetando a forma como compreendemos a relação evolutiva dos três super reinos e reforçando a hipótese de um ancestral comum de arqueas e eucariotos, à exclusão das bactérias.

Sendo assim, este trabalho tem como objetivo desenvolver um método de classificação de famílias altamente divergentes de proteínas e aplicar este método na análise das relações entre proteínas da superfamília TRAFAC. Nossa abordagem é baseada na geração semi-automática de perfis de sequências, ou seja, modelos Ocultos de Markov, que representam cada família de TRAFAC, e na aplicação de métricas de distâncias e algoritmos de agrupamento para definir as relações evolutivas entre as famílias descritas. A aplicação da nossa abordagem à classe TRAFAC permitiu a revisão da classificação evolutiva introduzida por Leipe e colaboradores, a atualização desta análise no contexto do imenso volume de dados atualmente disponíveis e evidenciou novas famílias ou subgrupos que poderão ser alvos de análises futuras.

1.4.1 Septinas

As septinas foram descobertas em leveduras durante o processo de divisão celular e receberam seu nome em função de sua localização na região do septo celular, onde fazem parte do colar septal formado durante a separação das células mãe e filha (Byers et al. 1976; Haarer e Pringle 1987; Hartwell, 1971; Longtine et al. 1996). Essas proteínas formam filamentos essenciais para o processo de divisão celular de leveduras, recrutando outras proteínas para formar polímeros proteicos (McMurray et al. 2011).

Atualmente sabe-se que as septinas se associam com a membrana celular, actina e microtúbulos e que atuam na formação de vesículas e em processos de secreção, características que fez das septinas mais um dos componentes essenciais do citoesqueleto em eucariotos (Mostowy e Cossart 2012).

Após a descoberta das septinas em leveduras, estudos filogenéticos dessas proteínas revelaram uma grande diversidade de subfamílias, amplamente distribuídas em Opisthokonta (supergrupo no qual se inserem Animais e Fungos) (Pan et al. 2007; Cao et al. 2007), mas ausentes em linhagens de plantas mais recentes e com representação restrita em ciliados, algas verdes e algas marrons (Nishihama et al. 2011).

Embora outras famílias de GTPases da classe TRAFAC estejam presentes em diversas linhagens de eucariotos e procariotos, as septinas parecem restritas a certos grupos de Eucariotos. Um grupo relacionado, porém ainda pouco conhecido de GTPases TRAFAC, conhecido como paraseptinas, está presente em Bactérias mas não foi, até o momento, encontrado em genomas de Arqueas (Leipe et al. 2002).

O tamanho das septinas varia entre 275-539 aminoácidos e a ligação do GTP ao núcleo catalítico é necessária para a formação de oligômeros, incluindo desde dímeros a longas cadeias de polímeros com simetria aberta (cabeça-cauda). A conversão para GDP, quando presente, é necessária para ambos, polimerização e despolimerização (Saraste et al. 1990). Hélices em *coiled-coil* próximas à região

C-terminal podem ser encontradas na maioria das septinas, ao passo que a região N-terminal possui características únicas em cada subfamília (Longtine et al. 1996) (Figura 3a).

As septinas possuem características típicas de GTPases da classe TRAFAC (translation factor-related) (Figura 3b), incluindo o motivo “GxxxxGKS/T”, presente no *loop* após a primeira fita beta, que corresponde à assinatura característica do domínio catalítico de todas as *P-loop NTPases*. Esta assinatura é também conhecida como motivo Walker A ou G1, o primeiro dos cinco motivos que caracterizam todas as GTPases (Walker et al. 1982).

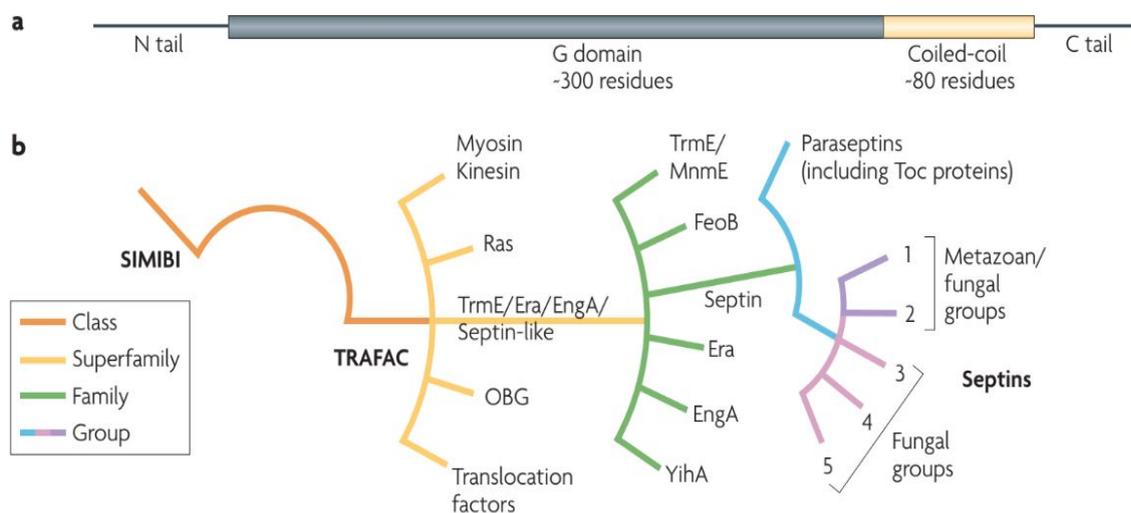


Figura 3. a) Arquitetura geral das Septinas com ênfase no domínio GTPase (G domain) na região N-terminal (*N tail*, ou cauda N), e o domínio “*Septin Unique Element*” (SUE, ou Elemento Único das Septinas), uma região constituída por uma presilha-beta e alfas-hélices em *Coiled-coil*, representado em amarelo na extremidade C-terminal (*C tail*, ou cauda C) da representação. O domínio GTPase é formado por entorno de 300 aminoácidos, enquanto o segundo domínio é formado por aproximadamente 80 resíduos. b) Diagrama da classificação das Septinas com base nos estudos evolutivos de Leipe, et al. 2002, mostrando que as Septinas são pertencentes a classe TRAFAC, mais proximamente relacionadas a outras famílias de GTPases como TrmE/MnmE, FeoB, ERA, EngA e YihA. (Leipe et al. 2002; Weirich, et al. 2008).

Um resíduo conservado de treonina, também conhecido por G2 ou “xTx”, localizado entre a segunda e terceira fitas betas do núcleo catalítico, é responsável por fornecer o segundo grupo hidroxila que coordena o íon Mg^{+2} , necessário para a

hidrólise de GTP (Bourne et al. 1991). O motivo “*hhhhDxxG*”, também conhecido por G3, faz parte da assinatura denominada Walker B, localizada na 4ª fita beta, cuja glicina conservada faz uma ponte de hidrogênio com o oxigênio terminal do fosfato gama da molécula de guanosina trifosfato (Leipe et al. 2002).

O motivo “NKxD” diferencia o domínio catalítico “G” de outras NTPases que interagem com outros nucleotídeos trifosfatos, é também conhecido como G4, é o responsável pela especificidade pela guanina a outras bases e está localizado depois da sexta fita beta. O motivo “NKxD” e o motivo “xCAX” (G5), localizado na fita 7, constituem as sinapomorfias mais conservadas no domínio catalítico das TRAFAC GTPases, e estão presentes nos membros da superfamília que inclui as septinas e paraseptinas. No entanto, a asparagina (N) no primeiro aminoácido do motivo “NKxD” não é completamente conservada, ou seja é raramente observada entre os membros da superfamília das septinas, incluindo nas subfamílias Pnut, YfjP-like, C18B2.5-like, Aig1/Toc34/Toc159-like e VC1806-like.

As três últimas subfamílias são membros de um subgrupo de TRAFAC que inclui todas as proteínas conhecidas como septinas e paraseptinas na literatura (Figura 3b). É possível que os genes ancestrais das septinas/paraseptinas tenham sido transferidos horizontalmente do ancestral pré-mitocondrial endossimbionte para dar origem as septinas e paraseptinas eucarióticas (Leipe et al. 2002).

1.4.2 Paraseptinas

Estudos anteriores (Leipe et al. 2002) dividiram as paraseptinas em dois subgrupos, Aig1-Toc34/Toc159 e paraseptinas bacterianas. O primeiro subgrupo compreende as Aig1-like, presentes em plantas e vertebrados, e as Toc34/Toc159-like encontradas em plantas. O segundo subgrupo está presente em bactérias e incluem os homólogos das proteínas codificadas pelos genes VC1806 (*Vibrio*), NMA0132 (*Neisseria*), HP0744 (*Helicobacter*), e slr1428 (*Synechocystis*) (Leipe et al. 2002).

As proteínas Aig1-like são conhecidas por suas funções no desenvolvimento do sistema imune, na regulação de respostas imunes em plantas e vertebrados e por regularem a apoptose celular em mamíferos através de vias mediadas pela mitocôndria (*mitochondria-mediated pathway*). Os membros dessa família podem estar localizados em diferentes compartimentos celulares, incluindo o citoplasma, retículo endoplasmático, complexo golgiense e mitocôndria (Poirier et al. 1999; Wang et al. 2009).

Toc34/Toc159-like são conhecidas por se localizarem na membrana de cloroplastos e importarem peptídeos para esta organela (Agne et al. 2009; Gutensohn et al. 2000). Membros desta família podem formar homo e heterodímeros, sendo que a alternância entre esses estados de oligomerização, como ocorre em muitas família de GTPases, define sua atividade enzimática (Koenig et al. 2008).

A descoberta do gene AIG1 (**AvrRpt2-induced gene 1**), o primeiro membro do clado das Aig1-like teve início com a análise da resposta de *Arabidopsis thaliana* à infecção por cepas de *Pseudomonas syringae* que carregavam diferentes genes de virulência (Reuber et al. 1996; Poirier et al. 1999). Estudos posteriores demonstraram que homólogos dessas proteínas também estão envolvidos na regulação da apoptose em mamíferos, levando a sua unificação na família GIMAP (*GTPases of Immunity-Associated Protein*), inicialmente conhecidas por IAN (*Immunity Associated Nucleotide binding proteins*).

Estudos mais recentes demonstraram que as GIMAP compõem um grupo monofilético com as Tocs e que exclui as demais paraseptinas, cuja relação com as GIMAP/Tocs e as septinas ainda não é clara. Membros do clado GIMAP/Tocs são capazes de dimerizar e atuam em diversos processos, tais como transporte de peptídeos e formação de vesículas, sugerindo que o último ancestral comum desta linhagem de GTPases já atuava em processos relacionados à membrana celular (Figura 4) (Schwefel et al. 2010).

Além dos motivos característicos das septinas presentes no domínio “G” (G1, G2, G3, G4 e G5), as proteínas GIMAP e Toc apresentam uma região conservada (CB ou Conserved Box) entre os motivos G3 e G4 (Figura 5) e um aminoácido com carga positiva, em geral lisina ou arginina, incorporado ao motivo G5 Krücken et al. (2004 e Nitta et al. (2006). Algumas GIMAP podem apresentar como extensão C-terminal uma sétima alfa-hélice e, em alguns casos, uma sequência de resíduos hidrofóbicos (Arasu, B.S. 2015).

Até a década passada as GIMAP/IAN haviam sido identificadas apenas em vertebrados e angiospermas, e caracterizadas experimentalmente em mamíferos placentários e plantas. Foi proposto que a diversificação das subfamílias de GIMAP ocorreu, por duplicação gênica, antes da divergência das linhagens atuais de mamíferos e de plantas (Krücken et al. 2004; Liu et al. 2008). A presença de ortólogos de GIMAP em plantas e *Danio rerio* (Filén et al. 2010) sugere uma distribuição mais ampla das GTPases associadas a processos de imunidade e pode implicar que a origem da família GIMAP remonta a linhagens eucarióticas precursoras de metazoários e plantas.

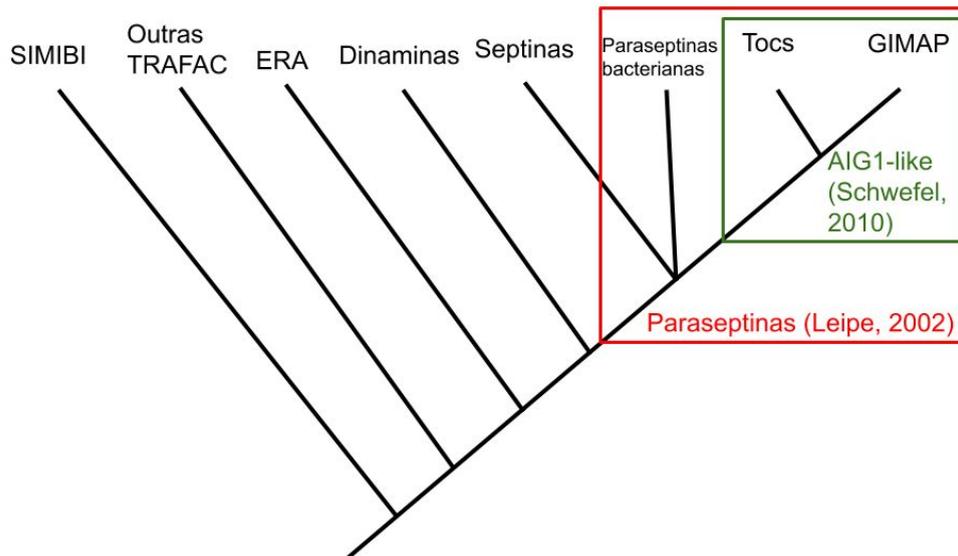


Figura 4. Filogenia das famílias de Septinas e a topologia simplificada dos membros representantes de cada clado cujas estruturas estão disponíveis no PDB (Berman et al. 2000; Schwefel et al. 2010).

Em mamíferos, GIMAP são predominantemente expressas nos linfócitos T e B, onde são necessárias para a maturação e sobrevivência dessas células (Nitta et al. 2006; Saunders et al. 2010). Além disso, membros do grupo GIMAP interagem com membros da família bcl-2 para regular a permeabilização da mitocôndria e o transporte de fatores apoptóticos da mitocôndria para o citoplasma.

O nocaute do gene GIMAP3 leva a severa linfopenia de linfócitos T (Pandarpurkar et al. 2003). GIMAP4 acelera a execução do processo de morte celular programada (Schnell et al. 2006), enquanto GIMAP6 e GIMAP8 inibem outras vias apoptóticas (Ho et al. 2017; Pandarpurkar et al. 2003).

Descobertas mais recentes revelaram que membros da família GIMAP participam da resposta imune em outras linhagens de animais além dos vertebrados, como o cnidário que vive em corais *Acropora millepora*. Esse animal, quando exposto a MDP (*muramyl dipeptide*), uma substância derivada da parede celular de bactérias, ativa a expressão de homólogos de três GIMAPs que participam da resposta imune em vertebrados (Weiss et al. 2013). Outros grupos identificaram GIMAPs envolvidas em respostas imunológicas complexas tanto em

Crassostrea virginica (ostra do leste), como em *Biomphalaria glabrata* (caramujo) (McDowell et al. 2016; Zhang et al. 2016).

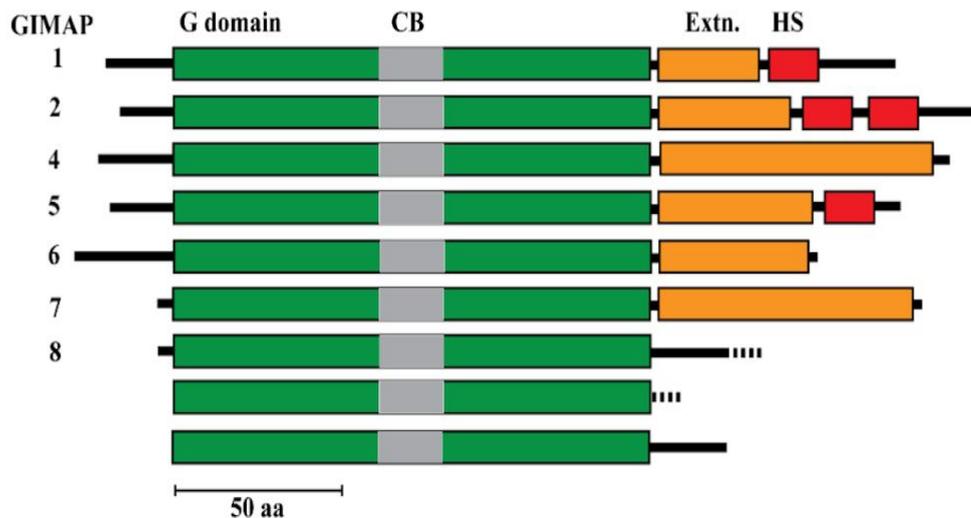


Figura 5. Arquitetura de domínio das GIMAPs presentes em humanos. O domínio “G” (G domain) intercala a região conservada CB (Conserved box) seguido de uma alfa-hélice que se estende na região C-terminal nas GIMAPs 1 a 7, enquanto GIMAP8 não possui essa extensão, mas possui três cópias fusionadas do domínio GTPase, e algumas GIMAP (1,2 e 5) terminam com sequências hidrofóbicas “HS” (Hydrophobic sequences) que podem, assim como as Septinas, formar uma região em coiled-coil (Conforme é possível ser observado na Figura 10 - Estrutura de GIMAP1). A barra corresponde a 50 aminoácidos (aa) (Arasu, B.S. 2015).

É notável uma grande importância na imunologia dos organismos citados para as proteínas as parasseptinas, e portanto compreender a história evolutiva dessa família e como ela se relaciona com a história de outras GTPases da classe TRAFAC é bastante relevante.

1.5 Mudanças no paradigma das relações evolutivas entre Eucariotos e Procariotos

Em 1977, Woese e Fox, descreveram os três domínios da vida que mais tarde foram nomeados: Eukarya (Eucaria ou Eucariotos), Bacteria e Archaea (Arquea). Estes domínios foram definidos primeiramente a partir da análise de árvores filogenéticas baseadas na sequência do gene do RNA da subunidade menor do ribossomo (subunidade 16S em procariotos e 18S em eucariotos), que demonstraram que os eucariotos e as arqueas formam clados separados das bactérias (Woese et al. 1990; Woese et al. 1977). Esse novo paradigma biológico foi consolidado por meio de análises moleculares das RNA polimerases do grupo das Arqueas, que se demonstraram mais complexas do que as bacterianas e mais similares às subunidades eucarióticas (Zillig et al. 1989).

Estudos recentes sugeriram que os Eucariotos descendem de uma linhagem relacionada a algum subgrupo de Arqueas, inicialmente apontando as TACK (Thaumarchaeota, Aigarcheota, Crenarchaeota e Korarchaeota) como o grupo mais próximo do último ancestral comum dos eucariotos (Figura 6). A recuperação de genomas completos a partir do sequenciamento em larga escala de amostras metagenômicas, porém, permitiu identificar novos grupos de arqueas com associação muito mais estreita com as linhagens eucarióticas. O primeiro desses novos grupos foi denominado Lokiarchaeota (Spang et al. 2015), e carrega, entre as arqueas conhecidas quando da sua publicação, o maior conteúdo de genes procarióticos que codificam proteínas consideradas específicas aos eucariotos (eukaryotic signature proteins, ou “ESPs”) (Spang et al. 2015; Hartman et al. 2002).

Estudos posteriores revelaram diversas outras linhagens aparentadas às Lokiarchaeotas, como Odinararchaeota, Thorarchaeota e Heimdallarchaeota, que juntos compõem o superfilo das Asgard Arqueas (Figura 6) (Zaremba-Niedzwiedzka et al. 2017).

Dentre as estruturas presentes no último ancestral comum dos eucariotos ou LECA (“Last Eukaryotic Common Ancestor”) é possível encontrar um sistema bem sofisticado de controle de membrana, complexo de Golgi, endossomos, lisossomos e peroxissomos, além de vias secretoras e endocíticas. Essas características podem ter sido herdadas verticalmente pelo ancestral dos eucariotos (FECA para “First Eukaryotic Common Ancestor”) ou, alternativamente, podem ter sido adquiridas por transferência horizontal de genes (Eme et al. 2018).

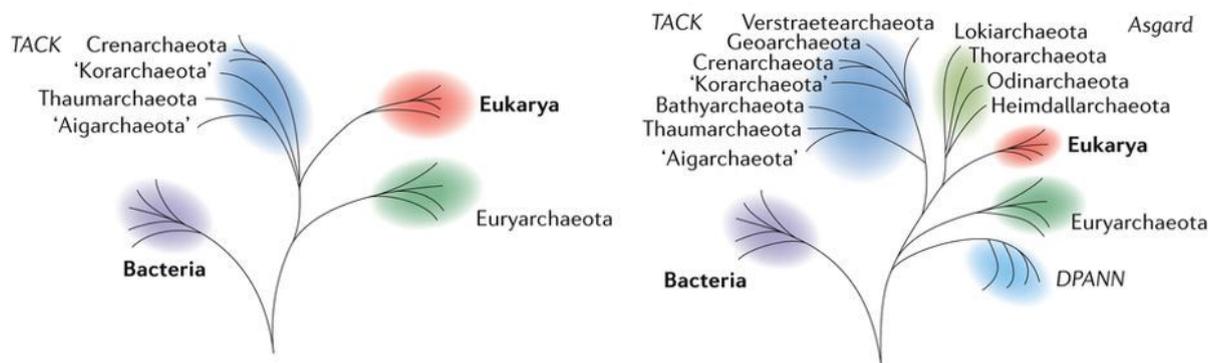


Figura 6. À esquerda está a árvore filogenética dos eucariotos (Eukarya) como clado-irmão das TACK arqueas, divergindo após o ancestral Euryarchaeota, hipótese construída antes da descoberta das Asgard Arqueas. À direita está a descrição filogenética das Asgard como clado-irmão do ancestral eucarioto (Eukarya), que divergiu após os diversos grupos dentro das TACK arqueas, Euryarchaeota e o recente descrito clado DPANN (Eme et al. 2018).

Os novos genomas das Asgard e sua relação com o grupo dos Eucariotos, quebrando o paradigma previamente estabelecido da relação entre Eucariotos e Arqueas, até então consideradas linhagens irmãs, porém distintas, dirigiu este trabalho a investigar a expansão e diversidade de GTPases nesses organismos, objetivando encontrar principalmente homólogos ainda não reportados da família das Septinas/Parasseptinas.

Nesse contexto, outras famílias de GTPases foram também detectadas nos genomas de Asgard Archaea. A forma como as famílias foram descritas e uma proposta de explicação evolutiva a respeito da história dessas famílias é descrita ao longo do trabalho.

2 MATERIAL E MÉTODOS

Para buscar homólogos da classe TRAFAC de GTPases foram selecionadas proteínas com estruturas resolvidas (Tabela 1) e cuja relação filogenética com o clado das septinas/paraseptinas tinha sido apontada por estudos anteriores. As buscas foram implementadas utilizando o software jackhmmer, do pacote HMMER (Potter et al. 2018; Eddy et al. 2011), no banco de dados de sequências de proteínas não-redundantes do NCBI (GenPept). Os resultados das buscas iterativas foram convertidos para um formato tabular e filtrados utilizando um *script* produzido pelo nosso laboratório, escrito na linguagem de programação Perl, que selecionou apenas os resultados com valores de *e-value* menores ou iguais a 10^{-3} .

A arquitetura de domínios das proteínas foi inferida usando um *script* desenvolvido em nosso grupo, que identifica as regiões correspondentes aos melhores resultados em buscas por domínios conservados em cada proteína, feitas com o software *hmmsearch* (Eddy, 2011) e o banco de dados Pfam (Finn et al. 2014). Nesse procedimento, os resultados das buscas são filtrados para valores de *e-value* $\leq 10^{-3}$ e sobreposições entre alinhamentos de modelos menores que 40% do comprimento da região alinhada são ignorados, resultado, na maioria dos casos, em arquiteturas de domínios sem sobreposição.

Os pHMMs construídos neste trabalho também foram utilizados, em vários estágios, na busca por homólogos no nosso banco de dados com o software *hmmsearch* (Eddy, 2011). O download dos genomas do NCBI também foi feito com auxílio deste mesmo pacote caseiro que inclui uma série de programas necessários para o fluxo da análise de sequências, assim como colher informações relacionadas aos genomas dos organismos com as proteínas encontradas, como a taxonomia (Figura 7).

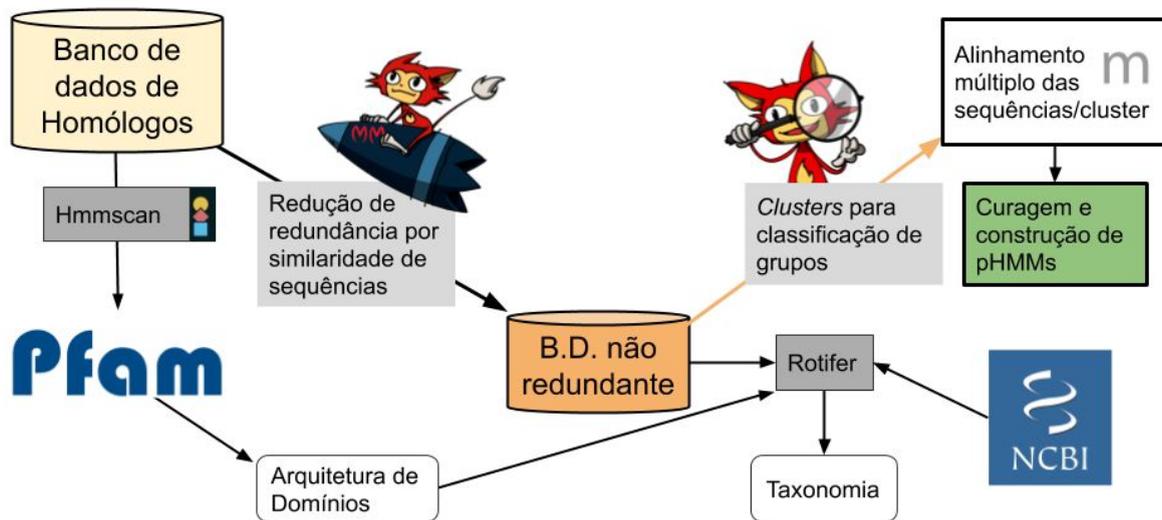


Figura 7. Workflow simplificado da análise desenvolvida por este trabalho. O desenho da raposa representa o MMseqs2, pois é a logomarca do pacote. “B.D.” é uma sigla para Banco de Dados. O “m” representa o logo do Mafft, software que constrói alinhamentos. Pfam é o banco de dados de famílias baseados em modelos de HMM e NCBI é o *National Center for Biotechnology Information*.

Para reduzir a complexidade dos dados e classificar os grupos encontrados foram utilizados algoritmos de *clustering* baseados na similaridade de sequências implementado no módulo *easy-clust*, do pacote MMseqs2 (Steinegger et al. 2017). Para obter a relação par-a-par entre as sequências representantes de clusters foi utilizado o protocolo de criação de banco de dados, descrito no manual do pacote MMseqs2, após a criação do banco de dados é possível obter a relação par-a-par das sequências representantes dos clusters por meio do módulo *search*, com as opções em modo padrão, e as medidas obtidas pelo módulo em formato de tabela (*tab-separated values*, valores separados por tabulação) com o módulo *ConvertAlis*.

Os alinhamentos dos grupos foram construídos com auxílio dos softwares Mafft (Kato et al. 2005) e Muscle (Edgar, 2004), editados e visualizados utilizando os softwares Aliview (Larsson, 2014) e Jalview (Clamp et al. 2004). Os modelos Ocultos de Markov (perfis de HMMs, ou pHMMs) foram construídos com o software HHmake, e os alinhamentos de todos os pares de modelos foram obtidos utilizando o *software* HHsearch, ambos presentes do pacote HHsuite (Steinegger et al. 2019).

As relações par-a-par entre sequências e modelos foram representadas como redes e visualizadas no software Cytoscape (Shannon et al. 2003). As estruturas disponíveis no Protein Data Bank foram visualizadas e analisadas usando o software PyMol (Schrodinger, L. L. C. 2020). Árvore filogenéticas, quando necessárias, foram construídas com auxílio do FastTree (Price et al. 2010) e visualizadas com o programa FigTree (Rambault, A. 2012).

Para construção do agrupamento hierárquico (dendrograma) de similaridade entre os componentes conectados de GTPases da classe TRAFAC, implementamos um programa na linguagem R que transforma os valores de *e-value* em estimativas de distâncias entre pHMMs. Para a construção, análise e visualização dos diferentes tipos de agrupamento avaliados usamos as bibliotecas *cluster* (Hadi et al. 1992); *FactoExtra* (Kassambara, 2020) e *Tidyverse* (Wickham et al. 2019).

3 JUSTIFICATIVA

3.1 Porquê compreender a evolução dessas famílias de proteínas?

Em 2002, Leipe e colaboradores, fizeram a primeira análise em larga escala das GTPases, classificando-as em diversas famílias. Dentre as duas grandes divisões propostas por estes autores, a classe TRAFAC, inclui muitas GTPases, associadas a diferentes processos de extrema importância para a manutenção da vida dos organismos, como os fatores de iniciação e alongação que atuam no processo de tradução, as proteínas Ras, que cumprem funções de sinalização, tráfico e direcionamento de proteínas entre compartimentos celulares e remodelamento da membrana, a família ERA, de proteínas que atuam na maturação de subunidades ribossomais, as dinaminas, que atuam no tráfego de vesículas nas células e as Septinas, que atuam no remodelamento de membrana e coordenação da divisão celular.

Todas as famílias citadas possuem homólogos em eucariotos e procariotos. A única família da classe TRAFAC que até o momento não possui homólogos descritos em procariotos (Arqueias e Bactérias) é a das septinas. Com o objetivo de explorar a possível origem procariótica das septinas e, a partir disso, lançar mais uma luz sobre o processo de eucariogênese, o objetivo inicial deste trabalho foi atualizar a classificação das TRAFAC e identificar famílias de proteínas em procariotos que fossem filogeneticamente relacionadas às septinas ou que tivessem potencial para realizar funções semelhantes nos processos de reorganização de membranas, citoesqueleto e na divisão celular.

Ainda dentro da proposta delineada acima, buscamos estender nossa análise para dois grupos de proteínas eucarióticas intimamente relacionadas às septinas, a família das GIMAPs e Tocs, anteriormente apontadas como um grupo parafilético relacionado às septinas, mas cuja própria história evolutiva ainda não foi satisfatoriamente explorada na literatura.

4 OBJETIVOS

Este trabalho objetiva classificar GTPases do *fold* P-loop NTPases e da classe TRAFAC com base em suas ancestralidades e divergências para compreender a relação filogenética dessas proteínas e construir hipóteses evolutivas a respeito dos grupos classificados.

4.1 Objetivos Gerais

- Gerar uma classificação natural e atualizada das GTPases TRAFAC;
- Identificar e classificar membros da família das septinas em genomas de procariotos (Bactérias e Arqueas);
- Compreender a história evolutiva das septinas e parasseptinas.

4.2 Objetivos específicos

- Delimitar o espaço de sequências das GTPases no banco de dados de proteínas não redundantes do NCBI (Classificação de alto nível das GTPases);
- Desenvolver um protocolo semi-automático para classificação de famílias de proteínas amplamente divergentes;
- Aplicar o protocolo acima na classificação e identificação das P-loop NTPases das classes TRAFAC e SIMIBI;
- Comparar os grupos de GTPases homólogos encontrados para compreender a relação evolutiva entre as famílias observadas;
- Estudar os detalhes da história evolutiva de septinas e parasseptinas;
- Identificar famílias, ainda não caracterizadas experimentalmente, no universo de proteínas TRAFAC, com ênfase naquelas com potencial de envolvimento em processos de remodelagem de membrana;

5 RESULTADOS E DISCUSSÃO

5.1 Coleta de homólogos e características das famílias de interesse

Para compreender o universo de sequências das GTPases relacionadas ao clado ERA/Dinamina/Septina fizemos buscas iterativas, por membros desses grupos, no banco de dados de sequências não redundantes do NCBI (descritos com detalhes na seção Materiais e Métodos). Como semente nessas buscas, usamos membros desses grupos já caracterizados experimentalmente e com estrutura tridimensional conhecidas, listadas na Tabela 1.

Tomando como base trabalhos anteriores que apontaram as septinas, GIMAPs e Tocs como membros de um clado que partilha ancestralidade comum com as proteínas da família ERA e as dinaminas, clado este que diverge das demais famílias de TRAFAC muito cedo na evolução do grupo, buscamos selecionar um conjunto mínimo e representativo de proteínas nessas famílias para usar em nossas buscas. Tanto as dinaminas como as septinas possuem como sinapomorfia a capacidade de formar homo- e heteroligômeros e de associarem-se com a membrana celular, sendo ambos os processos controlados pela dimerização via interface G dessas proteínas, que corresponde à região em que se localizam os motivos G1-G5, destacados nas figuras a seguir (Ferguson et al. 2012).

Uma segunda estratégia de busca e coleta de sequências homólogas usada neste projeto, aplicou modelos ocultos de Markov para famílias de GTPases, obtidos no banco de dados Pfam (Protein Family Database) (Bateman, 2002), cuja alta sensibilidade poderia detectar homólogos distantes das famílias desejadas. As buscas baseadas em modelos do Pfam, porém, convergiram para um subconjunto dos resultados obtidos com nossa abordagem iterativa baseada em múltiplas sementes.

Como as sequências usadas para iniciar as buscas iterativas eram altamente divergentes entre si, essa estratégia foi capaz de recuperar sequências homólogas dos grupos desejados mais divergentes do que a capacidade de detecção dos modelos do Pfam.

Tabela 1. Proteínas do *Protein Data Bank* (PDB) utilizadas como sementes para as buscas por homólogos distantes.

PDB	Cadeia	Descrição	Referência	Artigo (DOI)
3IEV	A	Estrutura em cristal de ERA em complexo com MgGNP e a região 3' final do rRNA 16S.	Tu, C. et al. 2009	https://www.pnas.org/content/106/35/14843
1TPZ	A	Estrutura em cristal de IIGP1: um paradigma para GTPases p47 de resistência induzidas por interferon.	Ghosh, A. et al. 2004	10.1016/j.molcel.2004.07.017
3V70	A	Estrutura em cristal da GTPase humana IMAP membro da família 1.	Nedyalkova, L. et al. 2011	10.2210/pdb3V70/pdb
5IRR	A	Estrutura em cristal do domínio GTPase da Septina de <i>Chlamydomonas reinhardtii</i> .	Pinto, A.P.A. et al. 2017	10.1074/jbc.M116.762229
3BB3	A	Estrutura em cristal de Toc33 de <i>Arabidopsis thaliana</i> em complexo com GDP e Mg ²⁺ .	Koenig, P. et al. 2008	10.1016/j.str.2008.01.008

A Tabela 1 lista as proteínas usadas como sementes em nossas buscas iterativas. Em alguns casos, a região homóloga ao domínio catalítico precisou ser isolada do resto da sequência original, para evitar a recuperação de falsos positivos. Esse foi o caso, por exemplo, da sequência com identificador PDB “3IEV_A”, para a qual foi necessário excluir a região C-terminal, que corresponde ao domínio KH (*K-Homology*, ou K-Homólogo). Este domínio é capaz de ligar e interagir com RNA, atividade necessária para que essa proteína cumpra sua função de maturação do RNA ribossomal 16S e montagem da subunidade ribossomal 30S (Tu et al. 2011; Razi et al. 2019).

A proteína 3IEV_A, representante da família ERA, foi selecionada para o início das buscas devido a sua suposta ancestralidade com as septinas, proposta nos trabalhos de Leipe (2002), e Schwefel (2012). A família ERA foi assim nomeada devido sua primariamente observada similaridade com as proteínas Ras

encontradas em *Escherichia coli*, *E. coli* *Ras-like protein*, ou proteína de *E. coli* similar a Ras, outra família de GTPases. Nossas análises indicam que o domínio ERA-GTPase é formado por seis fitas betas flanqueadas por cinco alfas hélices, que possuem os clássicos motivos G1-G5 (Figura 8), como descrito no início deste trabalho (ver introdução). Este domínio e sua atividade de ligação a molécula de GTP é essencial para o reconhecimento da estrutura de RNA ribossomal, que por sua vez promove a atividade de hidrólise de GTP (Tu et al. 2009).

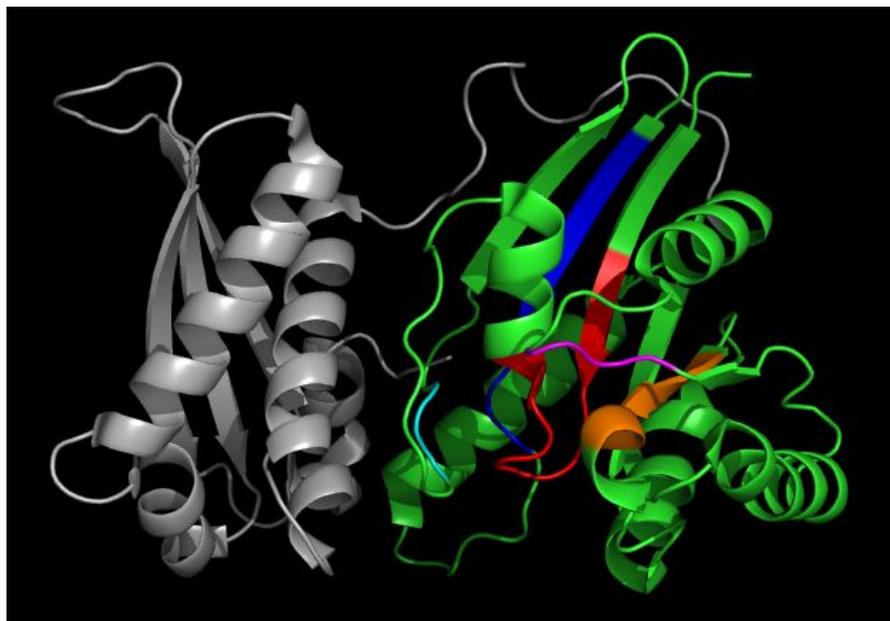


Figura 8. Estrutura da GTPase da família ERA em verde e do domínio KH 2 em cinza, com ênfase nos motivos que descrevem as GTPases, G1-G5, com os *loops* direcionados a molécula de GTP não mostrada. Vermelho: G1 ou Walker-A (*hhhhGXXXXGKT/S*), Ciano: G2 (T), Azul: G3 ou Walker-B (*hhhhDXXG*), Laranja: G4 (NKXD), Magenta: G5 (SA). Cinza: domínio KH do tipo 2.

Outra sequência selecionada para o início das buscas foi a da proteína induzida por interferon IIGP (1TPZ_A), atualmente membro da família “Mx” de proteínas de resistência e imunidade e representante da grande família das Dinaminas (Haller et al. 2015). Essas proteínas são conhecidas por mediar respostas imunes autônomas a diversos patógenos intracelulares por meio da indução de interferons, assim como a atividade de alguns genes da família GIMAP (ver introdução).

A septina de *Chlamydomonas reinhardtii*, que se localiza na base da estrutura flagelar deste organismo, foi utilizada para o início das buscas por possuir a sequência mais divergente entre as Septinas com estrutura disponível até o momento. Embora seja encontrada uma única cópia no genoma do organismo, foi observado por microscopia eletrônica que estas proteínas são capazes de dimerizar e formar filamentos assim como as septinas de Opisthokonta (clado de Animais, Fungos e grupos relacionados).

Nossas análises demonstraram que além da estrutura comum a todas as GTPases, com fitas-betas alternadas com alfas-hélices, as septinas possuem a inserção de uma presilha-beta na extensão do núcleo do domínio GTPases, na região C-terminal, necessária para a polimerização e estabilização da interação proteína-proteína entre as septinas (Figura 9, em roxo claro). Essa presilha-beta indica o início da região chamada Elemento Único das Septinas, ou *Septin Unique Element*, SUE, onde estão os resíduos responsáveis pela interação entre monômeros de septinas durante a formação de polímeros (Figura 9) (Pinto et al. 2017; Sirajuddin et al. 2007).

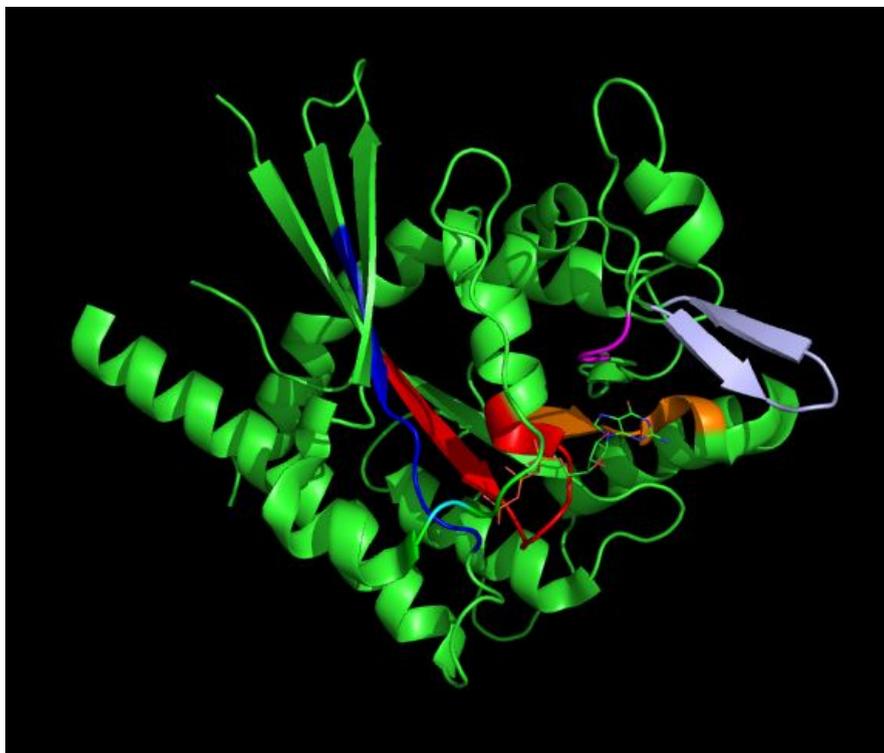


Figura 9. Estrutura do domínio GTPase da proteína 5IRR_A, de *Chlamydomonas reinhardtii*, ligada a guanosina 5'-3- O-(thio)trifosfato (GSP), análogo de GTP, com ênfase

nos motivos que caracterizam as GTPases e na sinapomorfia característica da família das septinas, que é caracterizada por uma presilha-beta, mostrada em roxo claro. Vermelho: G1 ou Walker-A (*hhhhGXXXXGKT/S*), Ciano: G2 (T), Azul: G3 ou Walker-B (*hhhhDXXG*), Laranja: G4 (XKAD), Magenta: G5 (SND).

As proteínas 3BB3_A (Toc33/34) e 3V70_A (GIMAP) pertencem à família *AIG1-like*, e compartilham uma série de características conservadas a nível de sequência, que são exclusivas destes dois grupos e vão além dos domínios clássicos das GTPases (G1-G5). Denominada AIG1 por causa do primeiro membro da família encontrado (ver introdução), envolvido na resposta imune contra bactérias patogênicas em *Arabidopsis thaliana* (Reuber et al 1996).

As proteínas da família AIG1-like cumprem diversas funções, a maioria delas associadas à membrana das células e organelas, e podem ser encontradas em genomas de diversos grupos de eucariotos. Por exemplo, a proteína XP_653817, apresenta expressão aumentada em cepas virulentas de *Entamoeba histolytica* (Biller, et al. 2010). Notavelmente, um estudo recente de homólogos de AIG1, também em *Entamoeba histolytica* revelou o papel dessas proteínas na virulência deste patógeno, com indícios de que as AIG1 podem atuar na regulação da adesão da membrana das células parasitas às células hospedeiras (Nakada-Tsukui et al. 2018).

As GIMAPs, aqui representadas pela proteína de PDB 3V70_A (Figura 10), atuam em diversas funções como a maturação de linfócitos T e B e a regulação da via mitocondrial de apoptose celular. Análises da estrutura dessas proteínas evidenciam que elas possuem uma alfa-hélice adicional na região C-terminal de seu domínio GTPase (Figura 10, em roxo claro), que auxilia no processo de polimerização e formação de estruturas mais complexas, como anéis, formados por múltiplos oligômeros, da proteína GIMAP2 (Schwefel, 2012).

Em contrapartida às demais proteínas da família AIG1, as proteínas TOCs atuam na formação do complexo Tic/Toc, uma estrutura de transporte necessária para o fluxo de peptídeos entre a célula e o cloroplasto. A sigla TOC vem de *translocase of the outer membrane of the chloroplast*, ou translocases da membrana externa de cloroplastos, um complexo de proteínas especializado no

reconhecimento e translocação pela membrana externa de proteínas cuja destino final é a matriz do cloroplasto.

Nota-se que, embora o sistema Toc e o complexo de translocases da membrana externa de mitocôndrias (TOM para *translocase of the outer membrane of mitochondria*), empreguem porinas da superfamília 85 de barris-beta, como principal componente do canal na membrana externa, esses dois complexos não são relacionados, sendo que a maioria dos seus componentes pertencem a famílias e *folds* distintos. A divergência em termos de composição se reflete em variações na forma como ambos os sistemas regulam a translocação: enquanto as TOM são dirigidas por afinidade ao substrato, as TOC são reguladas pela ligação e hidrólise de GTP, e por eventos de fosforilação (Schleiff et al. 2010).

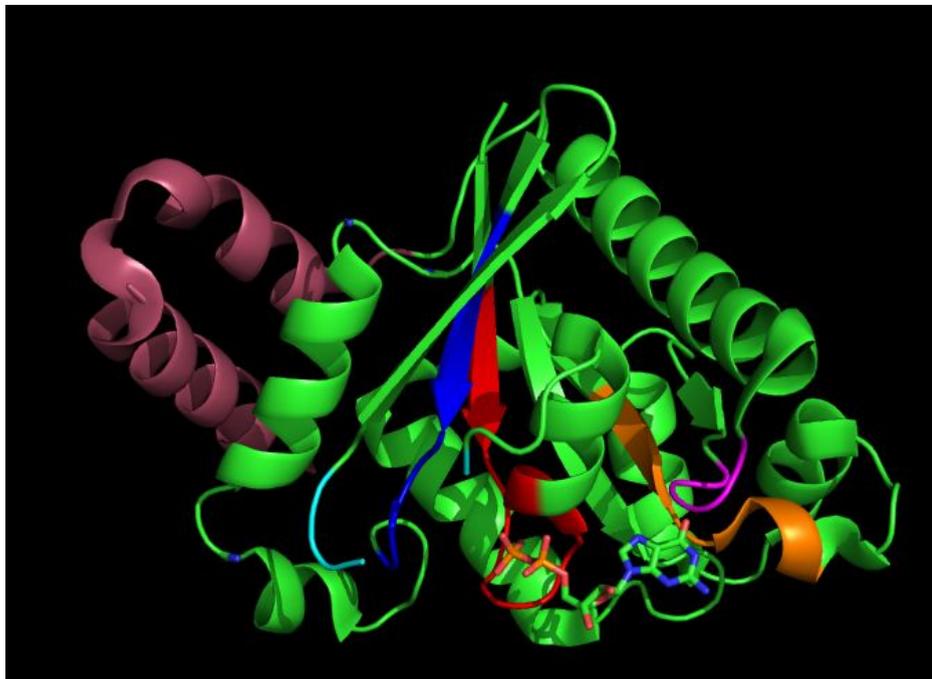


Figura 10. Representação da estrutura em 3D da proteína 3V70_A, ou GTPase IMAP 1 (GIMAP1) de *Homo sapiens*, ligada a GDP (representado em *sticks*, ou palitos), com ênfase nos motivos que caracterizam as P-loop GTPases da classe TRAFAC e na última alfa-hélice (em roxo) que caracteriza uma sinapomorfia do grupos das GIMAP. Vermelho: G1 ou Walker-A (*hhhhGXXXXGKT/S*), Ciano: G2 (T), Azul: G3 ou Walker-B (*hhhhDXXG*), Laranja: G4 (TXXE), Magenta: G5 (DNR).

O complexo Tic/Toc é fundamental para a maturação e homeostase das organelas de diversos organismos (Richardson et al. 2014; Schleiff et al. 2010). As proteínas Toc33/34 (Figura 11) juntamente com outras Toc (não GTPases), formam o componente da membrana externa, exposto ao citoplasma, do mecanismo de importação de peptídeos e proteínas para o tráfego do cloroplasto. As análises das estruturas dessas GTPases mostraram que embora tenha sido destacada anteriormente a presença de uma presilha-beta, comum a Tocs e septinas, e formada por fitas localizada após a α -hélice que sucede a última fita do núcleo catalítico conservado, as Tocs (Figura 11) diferem quanto às regiões que contribuem as fitas-betas que formam essa presilha. A primeira fita-beta componente da presilha é encontrada logo após a segunda alfa-hélice, que sucede a terceira fita-beta do núcleo catalítico, e a segunda fita-beta se localiza na mesma posição da presilha observada nas septinas, conforme descrito nas Figuras 9 e 11. Tal característica foi o que levou os primeiros estudos a indicarem que as proteínas TOCs seriam ancestrais do clado das septinas (Leipe, 2002; Schwefel, 2012), porém com novas estruturas depositadas nos bancos de dados, fica evidente que as duas famílias diferem nos detalhes da estrutura em presilha-beta das septinas.

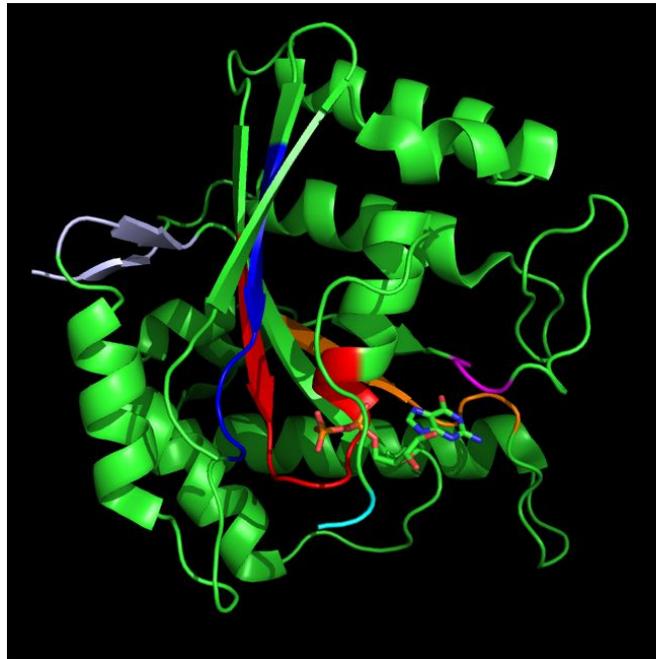


Figura 11. Representação da estrutura em 3D da proteína 3BB3_A, da família Toc33/34, GTPases que compõem a estrutura externa do sistema tic/toc que controla o transporte de peptídeos no cloroplasto. de *Arabidopsis thaliana*, em complexo com uma molécula de GDP (representado em *sticks*), com ênfase nos motivos que caracterizam as P-loop GTPases da classe TRAFAC e na presilha-beta-like (em roxo claro), característica observada como sinapomorfia do grupo das Septin-like (Schwefel, 2012). Vermelho: G1 ou Walker-A (*hhhhGXXXXGKT/S*), Ciano: G2 (*XPFX*), Azul: G3 ou Walker-B (*hhhhDXXG*), Laranja: G4 (*XHXQ*), Magenta: G5 (*ENS*).

As GTPases Toc33/34 também partilham os clássicos motivos G1 (Walker A) e G3 (Walker B) com as demais GTPases, no entanto os motivos G4 e G5 são considerados sinapomorfias do clado que as une com as GIMAPs e demais AIG1-like, sendo uma das características que sugerem que as TOCs divergiram da linhagem das septinas antes da emergência de outros grupos de proteínas AIG1-like, com os quais podem, talvez, constituir um grupo monofilético.

Com resultado de nossa estratégia de buscas iterativas, foram identificadas 2.326.843 sequências com similaridade significativa aos domínios GTPase. O uso de múltiplas sementes foi essencial para garantir a sensibilidade de nossa busca do amplo universo de sequências altamente divergentes do clado ERA/Dinaminas/Septinas.

5.2 Classificação das GTPases

5.2.1 Identificação e separação das famílias de P-loop GTPases

Nosso primeiro objetivo foi construir uma classificação evolutiva abrangente dos grupos de septinas/paraseptinas e outras GTPases da classe TRAFAC. Essa classificação foi necessária para compreendermos os limites entre os grupos e famílias de sequências já conhecidas ou ainda não identificadas na literatura.

Para reduzir a redundância e a complexidade dos dados, o universo de proteínas coletadas foi dividido em grupos (*clusters*) com base na similaridade das sequências, que foi calculada a partir de alinhamentos par-a-par, com auxílio do pacote MMseqs (Tabela 2).

Tabela 2. Etapas de clusterização, com os parâmetros utilizados representados na colunas e o número total de clusters após cada etapa nas linhas, a clusterização foi realizada com auxílio do pacote MMseqs.

agrupamento	identidade	cobertura	e-value	clusters	--max-seqs
1	80%	80%	10^{-3}	231,412	1,155,370
2	0%	80%	10^{-3}	19,874	231,412
3	50%	0	10^{-3}	5,775	19,874
4	80%	80%	10^{-3}	239,994	padrão
5	0%	80%	10^{-3}	29,947	padrão

Os parâmetros usados para o agrupamento de sequências são listados na Tabela 2, e diferem quanto à identidade e cobertura mínima permitidas entre os membros de cada grupo. Numa primeira etapa, visando apenas a redução da redundância em nossos dados, foram exigidos 80% de cobertura das sequências alinhadas e identidade mínima de 80%. As sequências representantes desses *clusters* foram selecionadas para uma nova etapa de agrupamento, cujo objetivo era

isolar proteínas da mesma família, ou seja, proteínas caracterizadas pela mesma arquitetura de domínios conservados mesmo que a similaridade desses domínios estivesse no limite de detecção de métodos de alinhamento par-a-par. Os parâmetros que usamos para atingir esse objetivo foram um mínimo de 80% de cobertura e valores de *e-value* menores ou iguais a 10^{-3} , independentemente do valor de identidade.

O segundo agrupamento (cobertura $\geq 80\%$ e *e-value* $\leq 10^{-3}$) permitiu uma primeira avaliação da abundância e distribuição do tamanho das famílias de GTPases obtidas nas buscas, no entanto muitas proteínas ainda ficaram isoladas como *singletons* (grupos com um único membro) (Tabela 3).

Tabela 3. Intervalos que demonstram os tamanhos de cluster na etapa 3 de agrupamento, evidenciando os intervalos a cada 1000 seqüências por cluster.

total seqs	seqs/cluster	tamanho cluster	total	intervalo	
				min	max
147032	0.635369	0.99823	19839	1	999
21267	0.091901	0.00080	16	1000	1999
28624	0.12369	0.00060	12	2000	2999
10948	0.04731	0.00015	3	3000	3999
4726	0.02042	0.00005	1	4000	4999
5199	0.02246	0.00005	1	5000	5999
6011	0.02597	0.00005	1	6000	6999
7605	0.03286	0.00005	1	7000	7999

O principal fator que contribuiu para essa fragmentação foi a estringência do requisito de cobertura pois, como exemplificado no alinhamento exibido na Figura

12, muitos dos membros de uma mesma família tinham seus alinhamentos par-a-par limitados às assinaturas mais conservadas do núcleo catalítico das GTPases, e muitas GTPases podem possuir fusões com outros domínios, assim como inserções e extensões. Por isso mais uma estratégia de agrupamento foi utilizada, na qual exigimos 50% de identidade mínima, mas sem um valor de corte para a cobertura do alinhamento, para agrupar GTPases da mesma família, mas com tamanhos diferentes como mostrado na Figura 12, alinhamento onde a *query* é uma GIMAP conhecida cujo núcleo GTPase se inicia após 24 aminoácidos, enquanto *subject* é uma GIMAP de *Tupaia chinensis*, cujo núcleo do domínio GTPases se inicia após 79 aminoácidos, ambas possuem 50% de identidade entre si, mas diferem em cobertura, também foram mantidos os valores de *e-value* de até 10^{-3} .



Figura 12. Alinhamento de duas sequências da família GIMAP construído com o algoritmo de alinhamento local do blastp Altschul et al. (1997), como Query ou semente foi utilizada uma GIMAP com estrutura resolvida (3V70_A), de *Homo sapiens*, e como Subject ou alvo foi utilizada uma sequência que faz parte da família GIMAP (número de acesso no NCBI: ELW64117.1), de *Tupaia chinensis*, uma espécie de musaranho arborícola da ordem Scandentia, grupo de mamíferos relacionado a primatas, insectívoros e roedores. As assinaturas conservadas das GTPases estão marcadas (G1 a G4) e correspondem a loops que fazem contato e contribuem para a ligação da molécula de nucleotídeo trifosfato. Alguns elementos da estrutura secundária da proteína utilizada como *query* são destacados, as flechas representando fitas-betas enquanto os cilindros representam alfas-hélices. A visualização do alinhamento foi gerada na plataforma web do BlasterJS Blanco-Míguez et al. (2018).

Embora o agrupamento baseado em valores mínimos de cobertura de 50% tenha reduzido o número de grupos, esse número era ainda muito maior que o esperado com base nas famílias de gtpases descritas na literatura (Leipe, 2002).

Uma inspeção manual dos grupos, que incluiu a análise individual de algumas sequências, coletadas ao acaso em grupos menores e maiores, indicou que diversas famílias já descritas na literatura permaneciam fragmentadas, o que implicava a necessidade de implementar métodos de agrupamento menos divisivos. Com o objetivo de avaliar os resultados dos agrupamentos já construídos e buscando identificar relações entre os grupos, decidimos construir e explorar redes de similaridade de sequências para nosso conjunto de GTPases, focando apenas nas relações de similaridade dos domínios catalíticos isolados.

Para a construção das redes de similaridade, calculamos novamente o alinhamento par-a-par das sequências representativas dos *clusters* do agrupamento com mínimo de 50% de cobertura (agrupamento 3). A função *search* do pacote *MMseqs2* (descrito em Materiais e métodos) foi aplicada a todos os representantes do agrupamento 3, contra um banco de dados composto pelo mesmo conjunto de sequências. A partir dos alinhamentos par-a-par de um conjunto de sequências é possível obter diferentes estatísticas que podem ser usadas como estimador do grau de similaridade entre as sequências. Valores comumente usados incluem o *bitscore* normalizado, o *e-value*, o percentual de identidade entre as sequências, ou grandezas derivadas dessas medidas, como o *bitscore* por posição (usado no software *blastclust*). Uma vez que tenhamos estimativas de similaridade para todos os pares de sequências no conjunto sob análise, é possível construir uma rede de similaridade de sequências onde os nós são as sequências e as arestas são o estimador de similaridade escolhido.

As redes de similaridade de sequência, *sequence similarity networks* ou SSNs, são ferramentas de visualização populares na área de análise de sequências porque permitem a visualização e análise global de todas as relações de similaridade e da diversidade de sequências no banco de dados, habilitando a identificação de sequências experimentalmente caracterizadas em grupos relacionados aos de sequências não-caracterizadas, o que permite a classificação e inferência de função para estas últimas (Copp et al. 2018).

A visualização da rede dos representantes do agrupamento 3 foi construída com o *software* Cytoscape (Shannon et al. 2003), aplicando o algoritmo de força ponderado pelos valores do percentual de identidade entre as sequências. A rede resultante se mostrou bastante densa e, a princípio, pouco informativa, com quase a totalidade das sequências aglomeradas numa malha muito densa dentro do maior componente conectado da rede (Figura 12). Apesar disso, ao compararmos o número máximo de conexões possível $((n-1)*n/2)$ com o número observado, a discrepância, excluindo ou não as sub-redes periféricas, sugeria que havia informação suficiente na rede para dividir o componente principal em subredes com significado biológico.

Com o objetivo de encontrar sub-redes significativas na rede de similaridades dos representantes de *clusters* de GTPases, decidimos aplicar um algoritmo de inferência de comunidades, grupos de nós com maior probabilidade de estarem conectados entre si do que com qualquer outros nós de outras comunidades (Barabási, 2013) (Figura 12). Nesta análise, escolhemos aplicar o algoritmo de Leiden, que é uma versão aprimorada do algoritmo de Louvain, e utiliza modularidade para encontrar grupos de nós na rede de forma rápida e eficiente (Traag et al. 2019). Este método procura maximizar a diferença entre o número absoluto de arestas numa comunidade e o número esperado dessas mesmas arestas, sendo que o valor esperado é estimado pelo quadrado da soma dos graus dos nós de uma comunidade, dividido por duas vezes o número total de arestas na rede, baseado no chamado modelo de configuração, um método de geração de redes aleatórias. Para aplicar o algoritmo de Leiden em nossos dados implementamos um *script*, na linguagem R, utilizando as bibliotecas *igraph* e *leiden* (Csardi & Nepusz, 2006).

Para delimitar os grupos conhecidos e classificar as sequências com famílias ou funções ainda desconhecidas, os clusters das comunidades foram anotados manualmente. Para processar rapidamente os membros das comunidades definidas pelo algoritmo Leiden, iniciamos a anotação manual identificando, nestas comunidades, P-loop NTPases cuja estrutura tridimensional fora resolvida experimentalmente e cuja atividade tivesse sido alvo de experimentos descritos em artigos publicados em revistas indexadas (Anexo). Em paralelo, inferimos

arquitecturas de domínios conservados, ou seja, a ordem e identidade de domínios protéicos nos membros das comunidades. Esses domínios foram detectadas pelo software *hmmscan* do pacote Hmmer, usando os modelos do banco de dados do Pfam ($e\text{-value} \leq 10^{-3}$). A inspeção e análise manual das proteínas com estrutura conhecida e a presença, em um grupo ou comunidade, de uma alta frequência de proteínas contendo regiões com alta similaridade a certos modelos de GTPases, nos permitiu atribuir a esses grupos e comunidades o nome correspondente dessas superfamílias. Nesse processo, baseamos nossas inferências em modelos cuja especificidade para algumas famílias e superfamílias conhecidas de GTPases verificamos durante nossa análise. Como exemplo desse procedimento, a maior comunidade em nosso conjunto de dados agrupou vários dos membros da família Ras com estrutura tridimensional resolvida experimentalmente (Anexo 1) e incluiu uma frequência elevada (>90%) de proteínas cujo domínio GTPases alinhava com máximo $e\text{-value}$ ao modelo Pfam para a família Ras (PF00071) dos quais 5% possuem a configuração Ras+Ras (Figura 13).

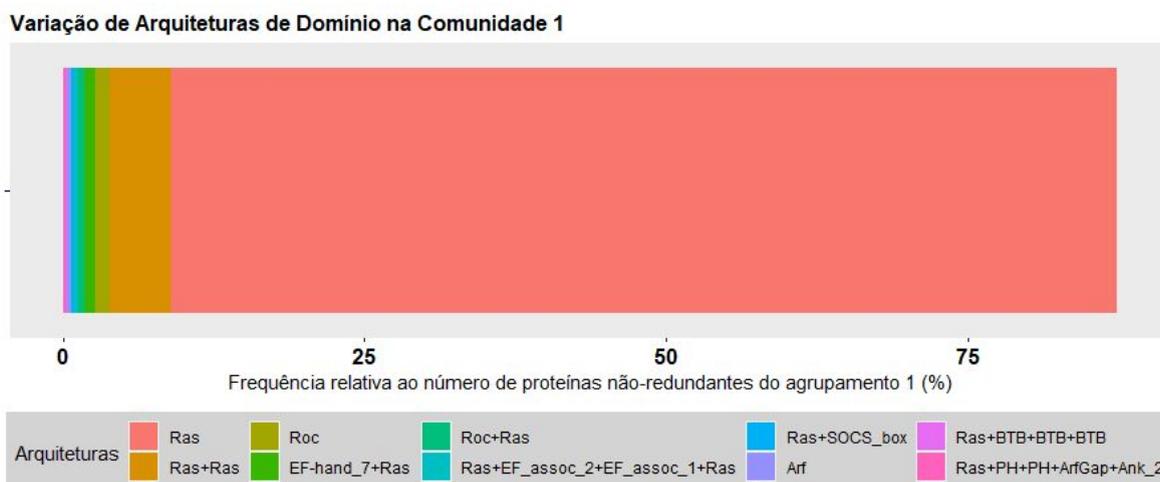


Figura 13. Gráfico em barras mostrando a frequência relativa de arquitecturas de domínios da comunidade 1. De 30,480 membros (após a redução de redundância aplicada no agrupamento 1), 78.4% combinaram com o modelo de HMM da família Ras do banco de dados do Pfam, enquanto 5% combinaram para duas cópias do mesmo HMM, Ras+Ras; 1.24% para Roc; 0.86% EF-hand+Ras; 0.6% Roc+Ras; 0.2% Ras+EF_assoc_2+EF_assoc_1+Ras; 0.2% Ras+SOCS_box; 0.2% Arf; 0.2% Ras+BTB*3; e 0.2% com Ras+PH+PH+ArfGap+Ank_2. Mostrando apenas as 10 arquitecturas de domínios mais frequentes na comunidade. As arquitecturas foram construídas com um *script* na linguagem Perl, desenvolvido por nosso laboratório.

Para validar a identificação inicial de famílias e superfamílias em nossos grupos e comunidades, até aqui baseada apenas na inspeção das proteínas com estrutura conhecida, presentes nesses grupos, e da frequência de arquiteturas de domínios, analisamos os alinhamentos múltiplos de cada comunidade. Tais alinhamentos foram construídos pelo software Mafft.

Nosso objetivo neste momento foi o de verificar a presença e conservação, das sinapomorfias conhecidas (Leipe, 2002), de cada família, nos grupos e comunidades anotados e assim verificar se a anotação baseada na frequência de arquitetura de domínios podia ser generalizada para os membros e grupos que não fossem detectados pelo modelo mais frequente ou por nenhum modelo.

Durante a análise das sinapomorfias, notamos que algumas famílias distintas, previamente descritas na literatura sem detalhamento de sua relação de similaridade ou filogenética, foram reunidas em grupos maiores. A comunidade 4, por exemplo, uniu as famílias FeoB, que atua no transporte de Ferro(+2) (Williams et al. 1985), e ERA, discutida na seção anterior, mostrando que estes grupos são mais similares entre si do que às demais GTPases. A presença de sinapomorfias de uma dessas duas famílias nos membros de grupos classificados pertencendo à outra família, porém, dificultou a correta classificação e distinção dos membros das famílias ERA e FeoB, sugerindo que a especificidade do agrupamento 3 era insuficiente para a classificação das GTPases e motivando a implementação de uma classificação adicional, que discutiremos mais adiante.

Na Figura 14 é possível visualizar a projeção das comunidades detectadas pelo algoritmo de Leiden sobre o componente principal da rede de similaridade de sequências das P-loop NTPases da família das GTPases. Nessa representação, cada nó é um grupo de GTPases definido pelo agrupamento 3 e o *layout* da rede foi definido pelo algoritmo de força. É possível observar que as comunidades não são mapeadas aleatoriamente na rede, mas se distribuem de forma congruente com o posicionamento definido pelo algoritmo de força. Esse resultado era esperado, pois tanto o algoritmo de força como o Leiden visam destacar sub-redes distintas a partir da análise das conexões na rede. Discordâncias entre os dois métodos de análise, porém, poderiam indicar erros de classificação. Da mesma forma, congruências

entre os dois métodos e a visualização dos resultados serviriam para obtermos mais *insights* sobre a estrutura geral do espaço de sequências.

Examinando a rede na Figura 12, identificamos, na região inferior da SSN, grupos contendo membros da superfamília Ras (tons de vermelho e vinho na SSN, Figura 14). É possível identificar, nesse *cluster* de sub-redes, proteínas das famílias Ran, Rho, Rab, proteínas-G, Rho, Roc, Arf, Ral, Rap, Ypt1 e outras variações de genes da superfamília Ras, todos já caracterizados experimentalmente (Anexo), e com estrutura tridimensional resolvida por difração de raios X. Os membros da superfamília Ras partilham um núcleo catalítico do *fold* GTPase caracterizado por uma folha composta de seis fitas- β , cercada por cinco alfas-hélices (Figura 1), correspondendo, portanto, a uma versão simplificada do núcleo conservado das TRAFAC.

Na parte superior da SSN, grupos de proteínas semelhantes às famílias ERA (*ERA-like*), septinas e paraseptinas (*Septin-like*) e dinaminas (*Dynammin-like*) (Figura 14), tendem a formar sub-redes distintas, mas com centros de massa que se aproximam, numa configuração que exclui outras GTPases. Na parte superior, à esquerda, é possível observar dois *clusters*, um à esquerda, identificado como septinas e proteínas similares, como GIMAP, Toc33/34 e AIG1, e outro, à direita, identificado como *Dynammin-like*, que inclui todas as proteínas relacionadas a dinaminas, incluindo a proteína Mitofusina-1, que se localiza na membrana externa das mitocôndrias e é essencial para o processo de fusão de mitocôndrias adjacentes (Cao et al. 2017).

A rede de similaridade e a análise de comunidades baseada no agrupamento 3 foram importantes para gerar nossa primeira classificação, de larga escala, dos homólogos de GTPases encontrados, além de permitir a primeira visualização do espaço de sequências dessas P-loop NTPases. Essa primeira classificação nos permitiu começar a entender o universo dessas famílias, porém os limites das comunidades formadas e as relações entre elas não ficaram claros, o que nos impediu de identificar quais grupos de proteínas de procariotos eram claramente mais próximos, filogeneticamente, de septinas e paraseptinas eucarióticas. Diante dessas limitações, construímos um novo agrupamento de sequências (agrupamento 5) e, para esse agrupamento, aplicamos a comparação de perfis de sequências

representados por modelos Ocultos de Markov (*profile Hidden Markov Models* ou pHMMs), gerando uma nova SSN. Os *e-values* calculados nessas comparações de pHMMs serviram como pesos para a ponderação das arestas da SSN entre grupos. Finalmente, ao invés de aplicar o algoritmo de construção de comunidades, optamos por isolar apenas as subredes correspondentes aos componentes conectados (CC) dessa rede. Um componente conectado é uma sub-rede composta por nós que possuem pelo menos uma conexão com pelo menos um nó da mesma sub-rede. Os CCs tendem a ser, portanto, mais aglomerativos que as sub-redes geradas pelo algoritmo de Leiden o que, como veremos adiante, acabou se mostrando mais eficiente para isolar corretamente os membros de famílias e superfamílias conhecidas do universo de GTPases.

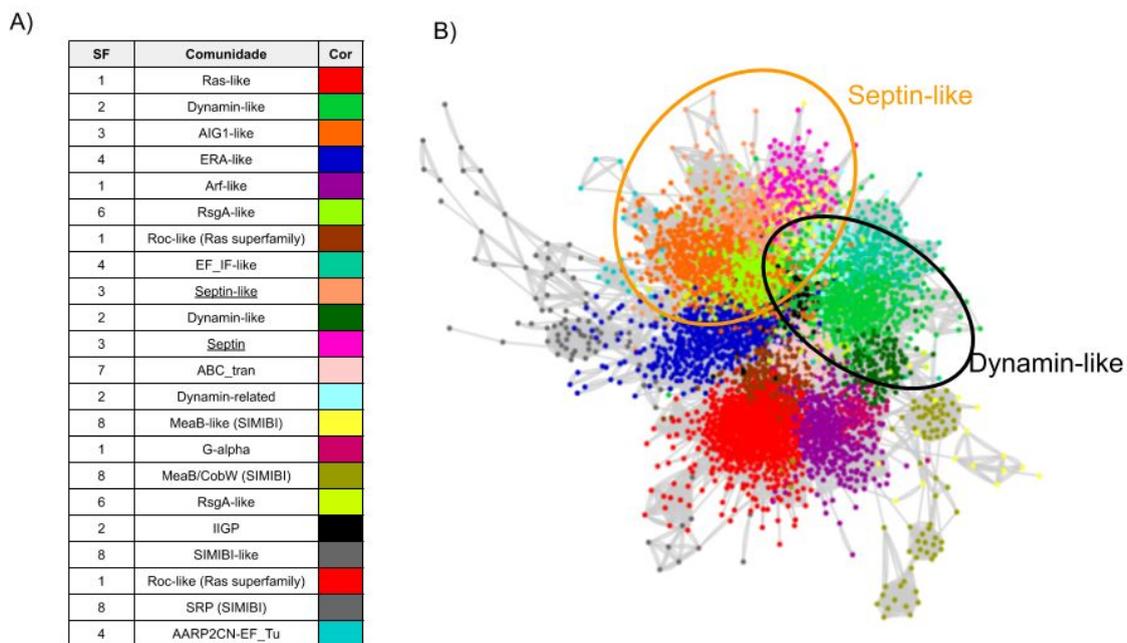


Figura 14. Rede de similaridade de seqüências do supergrupo das P-loop GTPases. **A)** Legenda de cores para as comunidades, refletindo a anotação manual. A anotação manual foi baseada em membros de cada comunidade que já foram caracterizados experimentalmente. A coluna SF inclui uma classificação preliminar das comunidades em grupos maiores, baseada na similaridade de seqüências e na anotação. **B)** Rede de similaridade, onde destacamos os *clusters* das famílias *Septin-like*, que incluem septinas e paraseptinas e *Dynamin-like*, que incluem dinaminas pequenas e outros grupos de dinaminas. As arestas representam *e-values* menores que 10^{-3} . A visualização acima foi gerada no Cytoscape, utilizando o algoritmo *Prefuse force-directed* com base no valor do percentual da identidade entre os nós. São mostrados 5775 nós, cada um representando um conjunto de seqüências que compartilham mais do que 50% de identidade. Clusters isolados e singletons foram retirados para melhorar a visualização.

Para construção da nova SSN, um novo agrupamento de todas as proteínas TRAFAC coletadas (agrupamento 5, na Tabela 2, cobertura $\geq 80\%$, $e\text{-value} \leq 10^{-3}$), foi construído, gerando 29947 grupos de sequências. Em seguida, construímos alinhamentos múltiplos, usando o software Mafft, das sequências de todos os *clusters* deste agrupamento com no mínimo 10 sequências, totalizando 2418 *clusters*. Os pHMMs derivados de cada um desses alinhamentos pelo software HHmake foram comparados uns com os outros usando o *software* HHsearch. Ambos os programas são distribuídos no pacote HH-suíte e incluem estados para modelagem de estruturas secundárias em seus pHMMs, o que lhes confere maior especificidade e sensibilidade nas buscas de similaridade (Steinegger et al. 2019). O procedimento aplicado em nossa segunda classificação é, portanto, análogo ao usado na primeira construção de comunidades, mas a estimativa de similaridade, antes baseada em alinhamentos par-a-par apenas entre sequências representativas, é substituída pelos valores de $e\text{-value}$ derivados do alinhamento de pares de modelos ocultos de Markov.

Existem diversos métodos capazes de comparar modelos estatísticos baseados no alinhamento de sequências homólogas contra um banco de dados de sequências, como o PSI-BLAST (Altschul et al. 1997), que utiliza perfis de PSSM para aumentar a sensibilidade das buscas por homólogos distantes, ou o *software* jackhmmer do pacote HMMER, que foi usado neste trabalho para as buscas de homólogos descritas no capítulo anterior, e se baseia em perfis de HMM construídos ao longo das iterações, sobre um banco de dados de sequências, para capturar homólogos distantes. Estes perfis obtêm mais informações sobre as famílias de sequências do que de uma sequência única, dado que são construídos com base em alinhamentos múltiplos de sequências homólogas. Sendo assim, os perfis permitem distinguir entre posições conservadas e variáveis de resíduos num alinhamento que são importantes para definir famílias de proteínas, além disso os perfis descrevem a exata variação possível para cada posição do alinhamento já que nesses perfis ou modelos está a informação da probabilidade da ocorrência de cada resíduo em cada posição de um alinhamento múltiplo. Os modelos de HMM são similares aos modelos de sequências que levam em consideração a variação dos aminoácidos por posição do alinhamento (PSSMs), mas além das frequências

de resíduos estes modelos também contém a informação de regiões específicas com inserções ou deleções no alinhamento. Foi mostrado que o alinhamento entre perfis de HMM resulta em alinhamentos de melhor qualidade e maior sensibilidade (Söding, 2005).

A SSN baseada na comparação de pHMMs (Figura 15) foi selecionada após um processo de refinamento, durante o qual foram geradas diversas visualizações de redes definidas por diferentes valores de corte para o peso das arestas. Como exemplo, a primeira rede gerada por esse procedimento mostrava a relação par-a-par de todos os modelos, sem filtro para as arestas, e era similar à SSN da Figura 14. Diferentes combinações dos valores de corte para os parâmetros *e-value*, *bitscore*, probabilidade, identidade e cobertura do alinhamento foram testados mas optamos, finalmente, por explorar apenas variações no *e-value* e usar os valores padrão de cobertura (80%), sem limites para *bitscore* e identidade.

Analisando todas as redes geradas a partir de valores máximos de *e-value* entre 10^{-140} e 10^{-3} , observamos a formação de um número moderado de componentes conectados quando o *e-value* máximo permitido para as comparações entre pHMMs foi fixado em 10^{-35} . Abaixo desse valor, o número de componentes conectados explodia, impedindo a detecção de relações entre os grupos. Acima desse valor, formavam-se componentes conectados gigantes e uma nuvem de grupos isolados, destruindo a estrutura da rede. Fixando o *e-value* máximo em 10^{-35} , a SSN se divide em um conjunto de 148 componentes conectados, sendo que 77 destes componentes são formados por pHMMs isolados, sem conexão aos outros componentes. Isolando os componentes conectados com mais de 5 pHMMs, aplicamos os seguintes passos de processamento para cada componente: (i) para todos os grupos em cada componente conectado, selecionamos uma única sequência representativa por grupo. A sequência selecionada foi sempre a medóide de cada grupo, ou seja, a sequência cuja distância para as outras sequências no mesmo grupo era mínima. (ii) As sequências de um mesmo componente conectado foram alinhadas pelo programa Mafft, no modo “--local-pair”. Nesse modo, a distância entre os pares de sequências durante o cálculo da árvore guia, é calculada usando o algoritmo Smith-Waterman, que gera alinhamentos par-a-par com o valor máximo de *score*, garantindo as melhores estimativas de distância entre os pares. O

uso do modo "--local-pair", em nossa experiência, resulta nos melhores alinhamentos, obtidos sem intervenção manual, para sequências que podem não ser homólogas em todo seu comprimento. (iii) Modelos Oculto de Markov foram construídos a partir dos alinhamentos de representativos de cada componente conectado e os modelos foram novamente comparados usando o programa HHsearch, resultando na rede da Figura 15.

Nessa nova rede, conseguimos separar *clusters* de famílias das classes TRAFAC e SIMIBI de GTPases. Usando os componentes conectados desse novo agrupamento, foi enfim possível isolar os homólogos de SIMIBI presentes em nossos dados que eram, na verdade, uma contaminação e, até esse ponto, não tínhamos conseguido isolar das TRAFAC. Alguns componentes conectados (ou CC's) agregaram muitos nós que representam a diversidade de modelos do banco de dados de GTPases construído pela abordagem deste trabalho, porém outros CCs eram menores, como as famílias da classe SIMIBI de GTPases. Isso provavelmente se deve à utilização de sementes somente da classe TRAFAC para a busca de homólogos remotos e não reflete diretamente o espaço de sequências da classe SIMIBI, embora tenha sido possível coletar os grupos dessa classe já conhecidos e descritos na literatura.

Repetindo o procedimento usado na análise de comunidades, analisamos os alinhamentos e arquiteturas de domínios dos 46 maiores componentes conectados da nova SSN, identificando as sinapomorfias conhecidas da estrutura terciária e os resíduos conservados do núcleo catalítico de cada família selecionada da classe TRAFAC. Como visto na Figura 15, todas as famílias de TRAFAC conhecidas puderam ser identificadas nos maiores componentes conectados.

Para auxiliar na anotação dos CC's foi empregada a comparação desses grupos com proteínas já caracterizadas estruturalmente, utilizando novas buscas com pHMMs que representavam a população de sequências em cada CC, contra o banco de dados do PDB, com o *software* HHsearch. Para cada CC, construímos um modelo Oculto de Markov a partir do alinhamento curado do domínio GTPase dos representantes de cada grupo no agrupamento 5. Estes modelos foram comparados entre eles mesmos com o *software* HHsearch (ver material e métodos), resultando em valores de *e-value* para os alinhamentos dos pares de pHMMs, os quais foram

usados para estimar a similaridade entre os CCs e formular hipóteses evolutivas a respeito da diversificação das GTPases da classe TRAFAC.

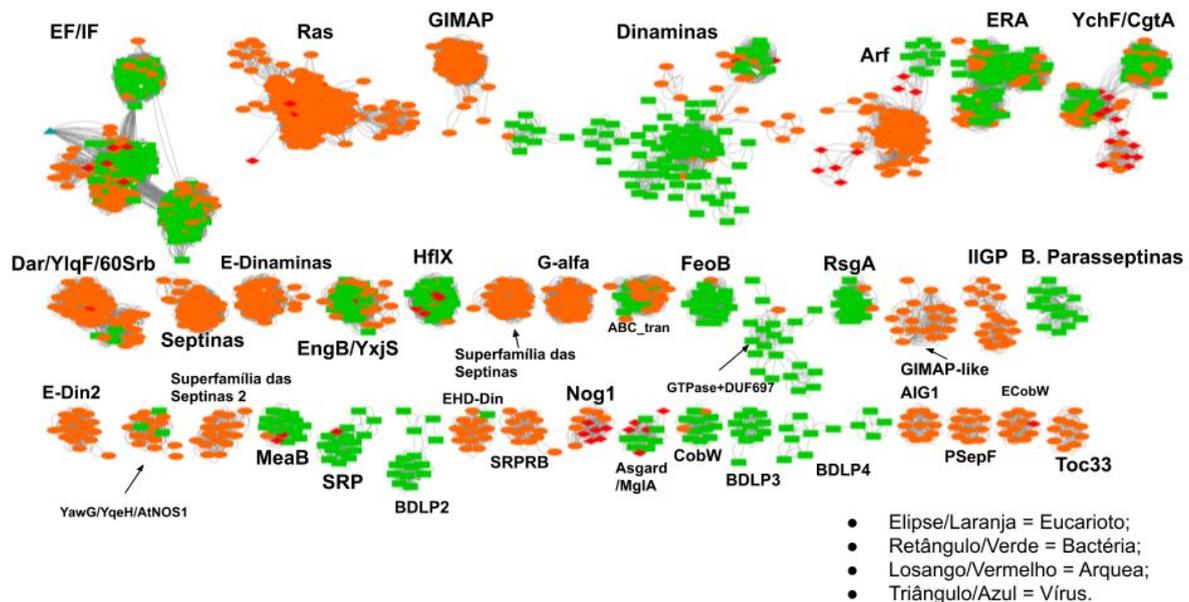


Figura 15. Representação do espaço de modelos do nosso banco de dados de homólogos de GTPases construído com base na comparação par-a-par de Modelos Ocultos de Markov (HMMs) cuja visualização foi gerada com o algoritmo *Force-directed* baseado no *e-value* das arestas no software Cytoscape. Cada conjunto representa um componente conectado, aos quais foram assinalados nomes das famílias de GTPases conhecidas. Componentes muito pequenos foram retirados para melhorar a visualização. Os nós representam HMMs, sendo que cada um dos HMMs é derivado de um alinhamento de sequências representativas dos grupos de sequências gerados pelo MMseqs usando o algoritmo *set cover* e exigindo 80% de cobertura e *e-value* máximo de 10^{-3} para os alinhamentos par-a-par de um mesmo grupo. As arestas representam a conexão entre nós para HMMs cujo alinhamento par-a-par, gerado pelo software HHblits Steinegger et al. (2019) resulte em *e-value* máximo de $1E-35$. Nas cores/formas estão representados os representantes de cada modelo, construídos com o algoritmo *Set Cover* de clustering de similaridade do pacote MMseqs2.

As superfamílias mais abundantes e diversas formaram componentes conectados grandes, que poderiam ser subdivididos em subgrupos compostos por famílias de GTPases relacionadas. Esse foi o caso, por exemplo, da superfamília dos Fatores de Tradução (EF/IF), encontrada em todos os grupos de seres vivos (bactérias, eucariotos e arqueas) e também em vírus. Essa superfamília compreende as famílias de fatores de iniciação eucarióticos, como eIF2 (*eukaryotic*

Initiation Factor 2, ou Fator de iniciação eucariótico 2), cuja fosforilação controla sua atividade de regulador da velocidade de síntese de proteínas no ribossomo, como parte da resposta ao estresse em humanos (Kenner et al. 2019). Inclui também os fatores de iniciação procarióticos, como IF-1 e IF-2, que são necessários para a montagem do complexo de iniciação ribossomal em *E. coli* (Kaledhonkar et al. 2019), e as famílias de fatores de elongação eucarióticos como o EF1A (*Elongation factor 1A*, ou Fator de Alongamento/Elongação 1A) de *Saccharomyces cerevisiae*, responsáveis por carregar aminoacil tRNAs até o sítio-A do complexo ribossomal de tradução 80S (Wirth et al. 2017); e fatores de elongação procarióticos como os EF-Tu's (*Elongation Factor Tu*, ou Fatores de Elongação Tu), que também são regulados por modificações pós-traducionais, como metilação e fosforilação, e atuam na resposta de bactérias à falta de nutrientes e na entrada na fase de crescimento estacionária (Talavera et al. 2018).

As famílias do componente conectado dos fatores de alongamento e iniciação são produto da diversificação de um grupo muito antigo de GTPases TRAFAC especializado na regulação ou execução da tradução. Nossa análise desse componente conectado aproxima os dados descritos na literatura: a superfamília de Fatores de Tradução compreende duas famílias principais, a família Bms1 e a família de fatores de tradução clássicos, sendo esta última dividida em quatro subfamílias muito antigas, IF2/eIF5B, eIF2g/SeIB, EF-Tu/EF-1a e EF-G/EF-2, que já estavam presentes no último ancestral comum de todas as formas de vida celular (LUCA). A congruência de nossos resultados com a literatura, para esse grupo, é ilustrada no *layout* de força da Figura 12, que evidencia a existência de sub-redes distintas de Fatores de Tradução, sub-redes que correspondem às famílias referidas acima.

Além de possibilitar a distinção entre as famílias de GTPases encontradas, a rede da Figura 5 permitiu identificar e isolar os membros de diversas famílias relacionadas às dinaminas, presentes em procariotos, conhecidas como BDLP's, ou *Bacterial Dynammin-Like Proteins*, ou seja, proteínas bacterianas similares às dinaminas. Esse grupo, ainda pouco explorado na literatura, inclui várias famílias envolvidas em processos de reorganização, fissão e fusão de membranas durante a divisão celular e a formação de vesículas (Low et al. 2006).

Os métodos de classificação descritos acima, baseados na comparação par-a-par entre as sequências e entre os perfis de HMM foram essenciais para: i) compreender o espaço entorno das sequências adquiridas do nosso banco de dados de homólogos de GTPases; ii) conseguir identificar e separar as famílias de GTPases das classes TRAFAC e SIMIBI; iii) mapear sequências ainda não anotadas no banco de dados em famílias de GTPases já conhecidas; iv) Comparar as famílias encontradas de modo a compreender a relação evolutiva entre elas; v) Criar modelos de HMM para cada família e vi) Permitir identificar os genomas que possuem essas GTPases e descobrir ainda mais sobre a origem e evolução dessas famílias de proteínas.

Embora nossa análise dos componentes conectados na SSN dos pHMMs (Figura 15) tenha demonstrado que eles reproduzem ou agrupam as principais famílias de TRAFAC descritas na literatura, membros de algumas famílias importantes, como septinas e dinaminas, acabaram classificadas em mais de um componente conectado, sugerindo que conexões significativas entre os CCs estavam sendo perdidas pelo método de particionamento ou pelo valor de corte escolhido para o *e-value*. Como a exploração de parâmetros de corte, nesta análise e na SSN da Figura 14, indicou que seria difícil ou impossível encontrar um conjunto ótimo de parâmetros que permitisse reproduzir as famílias descritas na literatura, optamos por aplicar um método de agrupamento aglomerativo, com o objetivo de produzir um agrupamento hierárquico que evidenciasse as relações entre componentes conectados.

5.2.2 Agrupamento hierárquico e a relação entre as famílias de GTPases da classe TRAFAC

Para melhor entender as relações entre as famílias de TRAFACs contidas nos CCs, construímos um agrupamento hierárquico desses elementos usando, como medida de distância, uma transformação dos *e-values* de alinhamentos par-a-par entre pHMMs construídos para representar cada CC. Nessa análise, coletamos as sequências representativas de cada grupo dentro dos CC (agrupamento 5) e alinhamos esses representantes usando o programa Mafft e a opção “--local-pair”. Modelos de Markov em perfil foram gerados a partir de cada alinhamento múltiplo e o *software* HHsearch foi usado para calcular os alinhamentos par-a-par entre os pHMMs que representavam a população de sequências em cada CC. Os *e-values* dessas comparações par-a-par foram transformados utilizando a função $-\log_{10}$, de modo a reduzir as diferenças dos valores originais, que compreendiam múltiplas ordens de grandeza e são, em geral, muito próximos de zero (10^{-4} a 10^{-180}). Os valores transformados de *e-value* foram então normalizados, isto é, submetidos a uma transformação linear que fez a média da distribuição ficar igual a zero e o desvio padrão igual a um. Essa segunda transformação tornou desnecessária a escolha de valores arbitrários para o máximo *e-value*. A matriz de distâncias, resultante das transformações acima, foi usada como entrada para o algoritmo *Agnes* (*Agglomerative Nesting*), da biblioteca *cluster* da linguagem R, e foi utilizado o método de variância mínima de Ward para selecionar, a cada iteração do algoritmo aglomerativo, o próximo par de grupos a ser fusionado (Hadi et al. 1992). A visualização da hierarquia de grupos, em formato de árvore (dendrograma), foi construída com auxílio da biblioteca *FactoExtra*, presente no CRAN.

O dendrograma visualizado na Figura 16 mostra a similaridade entre os modelos construídos para os CCs e permitiu a definição de novos níveis de agrupamento para as principais superfamílias da classe TRAFAC de GTPases.

Embora a análise dos resultados do método de agrupamento proposto aqui ainda esteja em andamento, é importante ressaltar que conseguimos, pela primeira vez, superar a grande dificuldade na inferência das relações evolutivas entre essas superfamílias, consequência do alto grau de divergência dessas famílias muito antigas e que, em alguns casos que veremos mais adiante, experimentam episódios de diversificação acelerada, com altas taxas de evolução e duplicação. A Figura 16 é, portanto, a primeira árvore de distâncias, derivada semi-automaticamente e por um processo completamente reprodutível, capaz de revelar as relações de similaridade de todas as grandes superfamílias de GTPases. Nossa expectativa é de que essa abordagem possa ser aplicada a outras famílias de P-loop NTPases e outros grandes grupos de proteínas divergentes, complementando o processo de curadoria manual, via exame de sequências e estruturas, que é usado atualmente na classificação dessas e outras superfamílias e *folds* de proteínas.

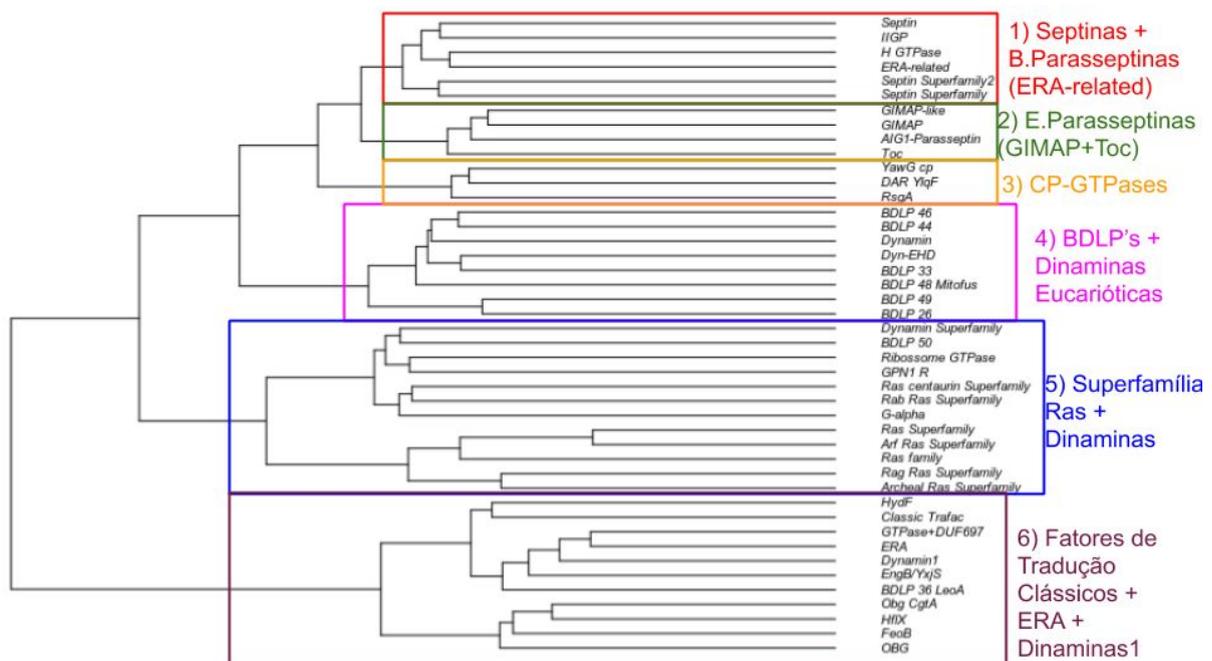


Figura 16. Árvore das distâncias entre os modelos de pHMMs construídos e subsequentemente das famílias de GTPases encontradas evidenciando a relação evolutiva entre diversas famílias altamente divergentes já descritas na literatura e novas famílias de GTPases ainda não notadas. Em ordem, de cima a baixo, e de acordo com os grupos definidos na árvore: 1) Septinas+B.Parasseptinas: Septin (Septinas); IIGP; H GTPase; ERA-related (Relacionadas a ERA, receberam essa *tag* inicialmente por sua alta similaridade com 3IEV_A, uma estrutura de ERA depositada no PDB, mas foram vistas

serem as *B. Parasseptinas* como mostrado na Figura 15 mostrando os componentes conectados); Septin Superfamily 2 (Superfamília das Septinas 2); Septin Superfamily (Superfamília das Septinas). 2) E. Parasseptinas: GIMAP-like (Similar a GIMAP); GIMAP; AIG1-Paraseptin (Parasseptinas de AIG1); Toc. 3) CP-GTPases ou GTPases Circularmente Permutadas: YawG; DAR YlqF; RsgA. 4) BDLP's + Dinaminas Eucarióticas: BDLP45; BDLP44; Dynamin; Dyn-EHD; BDLP33; BDLP48 Mitofus; BDLP49; BDLP26. 5) Superfamília Ras + Dinaminas: Dynamin Superfamily (Superfamília das Dinaminas); BDLP50; Ribosome GTPase (GTPase ribossomal); GPN1 R; Ras Centaurin Superfamily (Superfamília Ras - Centaurin); Rab Ras Superfamily (Superfamília Ras - Rab); G-alpha (G-alfa); Ras Superfamily (Superfamília Ras); Arf Ras Superfamily (Superfamília Ras - Arf); Ras family (Família Ras); Rag Ras Superfamily (Superfamília Ras - Rag); Archeal Ras Superfamily (Superfamília Ras - Arqueal); 6) Fatores de Tradução Clássicos + Dinaminas1: HydF; Classic Trafac (Trafac clássico, que inclui EF/IF citados anteriormente); GTPase+DUF697; ERA; Dynamin1 (Dinamina1); EngB/YxjS; BDLP 36 LeoA; Obg CgtA; HflX; FeoB; OBG.

Ao avaliarmos a topologia do dendrograma na Figura 16, seis grupos principais podem ser identificados que incluem as diferentes superfamílias da classe TRAFAC. O primeiro desses *clusters* agrupa as septinas e parasseptinas com as IIGP (supergrupo 1, parte superior da figura). A proximidade dessas famílias com as IIGP diverge das relações descritas na literatura pois, anteriormente, as IIGP haviam sido classificadas como membros do grupo das dinaminas, em virtude da presença de uma extensão alfa-hélica similar ao domínio de “Tromba” das dinaminas. Por ter sido baseado exclusivamente na similaridade do núcleo catalítico das GTPases, nosso dendrograma sugere que o ganho das extensões C-terminais ocorreu múltiplas vezes ao longo da evolução das GTPases e sugere que essas famílias são mais próximas entre si do que de quaisquer outras famílias de GTPases e formam um grupo monofilético com outros dois grupos identificados como *H GTPases* e *B. Parasseptinas*. As *H GTPases* são um grupo composto por diversas GTPases não caracterizadas experimentalmente e que estão presentes, principalmente, em genomas de Fungos. O componente conectado primeiramente denominado ERA-related, devido sua alta similaridade com proteínas da família ERA, na Figura 16, corresponde às *B. Parasseptinas* da Figura 15 e inclui algumas das parasseptinas bacterianas descritas por Leipe e colaboradores em 2002, tais como VC1806, NMA0132, slr1428 e as proteínas YeeP, YkfA, e YfjP, exemplos de uma expansão linhagem específica de parasseptinas em *Escherichia coli* Leipe et al. (2002). A outra parasseptina descrita por esses autores, HP0744, porém, é uma

fusão do domínio GTPase com DUF697 e pertence, em nossa análise, a outro clado (ver abaixo). O supergrupo 1 inclui ainda dois outros grupos da superfamília das septinas, *Septin superfamily* e *Septin superfamily 2*, que são encontrados apenas em fungos e cujas características descrevemos em detalhes mais à frente.

A topologia da árvore na Figura 16 ratifica a árvore proposta por Schwefel, em 2012 (Figura 4), que identificou as GIMAP's e Tocs como parasseptinas, pois estas famílias, juntamente com as septinas, formaram um clado separado das demais famílias de GTPases da classe TRAFAC (agrupamentos 1 e 2, em vermelho e verde na Figura 16). Como linhagem irmã deste grupo é possível observar a presença de um *cluster* formado pelas GTPases circularmente permutadas, a superfamília CP-GTPases (agrupamento 3, laranja), que inclui as famílias YawG, DAR/YIqF e RsgA. Membros dessas famílias atuam na maturação de subunidades ribossomais e são necessárias no estágio final de montagem do ribossomo 30S (Campbell et al. 2008; Goto et al. 2011; Thurlow et al. 2016). A topologia do dendrograma sugere que o evento de permutação circular característico desse grupo ocorreu após a separação desta linhagem e a linhagem das septinas, a partir do ancestral comum deste grupo com as proteínas semelhantes a septinas (*Septin-like*, grupos 1 e 2 na Figura 16).

Logo abaixo do cluster que une as *Septin-like* e as CP-GTPases é possível encontrar um grupo monofilético que inclui as dinaminas eucarióticas, como as ATPases da família EHD, e procarióticas, como diversas famílias de BDLP's (*Bacterial Dynammin-like Proteins*) (Lowe, 2008), cada uma identificada pelo nome atribuído ao componente conectado correspondente e descrito na seção anterior. Existem diversas famílias de BDLP's neste clado que ainda não foram exploradas na literatura e novos trabalhos experimentais serão necessários para compreender os diversos processos dos quais podem participar as GTPases deste grupo de proteínas.

Em concordância com a literatura, os componentes conectados contendo proteínas semelhantes às Ras GTPases formaram um agrupamento separado das demais superfamílias de TRAFACs. Esse clado uniu as as famílias Arf, Rag, Rab, Rho e as GTPases heterotriméricas, que formam o complexo conhecido como proteína G. Todos esses grupos compartilham considerável similaridade de

sequência, além de assinaturas características no motivo conservado G5 e são alvos de modificações pós-traducionais que determinam sua localização subcelular, elemento crítico para seu papel na célula, pois essas proteínas tipicamente atuam em redes regulatórias associadas a sensores localizados na membrana, interagindo com moduladores e efetores citoplasmáticos que determinam a transição entre os estados ativo e inativo dessas GTPases, compondo circuitos que regulam diversos processos celulares em organismos procariotos e eucariotos (Wennerberg, 2005). Dentre os moduladores que controlam sua atividade, uma característica conservada na superfamília Ras é a regulação da atividade catalítica dessas proteínas por um repertório diverso de proteínas conhecidas por GAPs (*GTPase-Activating Proteins*, ou Proteínas de Ativação da atividade de GTPase) e GEFs (*Guanine nucleotide Exchange Factor*, ou Fatores de troca de nucleotídeo de Guanina). Como a atividade de GTPase é muito lenta nas famílias de proteínas que compõem este agrupamento (Bos et al. 2007), a ação de GAPs e GEFs permite uma regulação fina de sua atividade, característica que é efetivamente uma sinapomorfia das proteínas da Superfamília Ras.

O último agrupamento monofilético na Figura 16 compreende: os fatores de alongamento e iniciação, citados anteriormente; a família HydF, que contém um domínio adicional com um núcleo de ferro/enxofre e atua na translocação e maturação de núcleos de ferro de [Fe-Fe]-hidrogenases (Maso et al. 2015); o componente conectado contendo os membros da família ERA, aparece como clado irmão da linhagem que inclui a proteína HP0744, previamente descrita como parasseptina, e caracterizada pela fusão da GTPase a um domínio de função desconhecida conhecido por DUF697; uma família de dinaminas eucarióticas; a família EngB/YxjS; e uma família de BDLPs conhecida como LeoA, que atua na secreção de vesículas com toxinas e por isso é frequentemente descrita como fator de virulência (Michie et al. 2014). O outro clado dentro do supergrupo 6 inclui pelo menos duas famílias com funções associadas à montagem e manutenção do ribossomo. Dentre estas, a família OBG atua na maturação de ribossomos e *checkpoint* da montagem da subunidade 50S (Chatterjee et al. 2019). Esta família é conservada em bactérias e amplamente distribuída em eucariotos, onde inclui a subfamília hOLA1, cujos membros participam da regulação de processos de

resposta a estresses ambientais e oncogênese em mamíferos (Balasingam et al. 2020; Czyz, et al. 2005). A família HflX é outro membro deste clado que inclui GTPases que interagem fisicamente com a subunidade 50S do ribossomo e que atuam na maturação ribossomal e, possivelmente, também na síntese proteica (Blombach et al. 2011; Polkinghorne et al. 2008). O último membro deste clado, a família FeoB, se destaca por ser o único grupo que não está envolvido no processamento do ribossomo e, ao invés disso, exerce importante papel no transporte de íons ferrosos (Fe^{+2}) em Euryarchaeota e bactérias (Hung et al. 2010). A atividade de FeoB no transporte de ferro depende da ampla região transmembrana, presente na extremidade carboxi terminal dos membros dessa família. Embora a proteína FeoB de *Escherichia coli* demonstre atividade hidrolítica específica para GTP, um estudo recente demonstrou que o ortólogo de *Vibrio cholera* é capaz de hidrolisar tanto GTP quanto ATP e que mutações do motivo G4 (xNKxD) abolem a atividade de GTPase mas preservam a de ATPase e que esta atividade é suficiente para promover o transporte de íons Fe^{+2} (Shin et al. 2019). Esses resultados são compatíveis com a ausência de conservação do motivo G4 nesse grupo e sugerem que essas GTPases atuam de forma análoga às ATPases dos transportadores ABC, provendo energia para os rearranjos estruturais da permease C-terminal, necessários para a translocação dos íons (Leipe, 2002). Esse conjunto de adaptações, específicas dessa família, contribuem para sua divergência em relação aos outros membros do supergrupo 6 e pode explicar a discordância da posição dessa família com relação à classificação de Leipe *et al.* (2002), que coloca, com base em análises de similaridade, FeoB como grupo irmão das ERA GTPases que, em nosso dendrograma, pertencem a outro clado. O exemplo de FeoB, além de ilustrar a dificuldade de classificar essas famílias em um nível tão elevado, sugerem que nossa análise seria beneficiada por métodos que permitissem avaliar a incerteza na posição das famílias no dendrograma de similaridade, análogos ao uso do *bootstrap* não-paramétrico em filogenias.

5.3 Superfamília das Septinas

A análise aglomerativa dos pHMMs do agrupamento 5, descrita acima, revelou que que septinas e parasseptinas se agrupam em uma superfamília de proteínas *Septin-like* que, pelo menos no caso das septinas de eucariotos, são demonstradamente capazes de interagir com a membrana celular e atuam em atividades de reorganização de compartimentos celulares, como a divisão celular e o remodelamento da membranas. Para análise e caracterização do grupo *Septin-like*, portanto, excluimos a família de parasseptinas inicialmente descrita em *H. pylori* (HP0744), que se revelou um grupo distinto e mais semelhante à superfamília ERA, do supergrupo 6. Para caracterizar as *Septin-like*, descrevemos primeiro as sinapomorfias desta família, que serve de referência para este clado, e em seguida caracterizamos sua distribuição.

5.3.1 Sinapomorfias das septinas

As septinas formaram um grupo distinto das outras famílias na nossa análise de CC's: além das sinapomorfias clássicas das proteínas da classe TRAFAC, as septinas possuem, em geral, um resíduo de Asparagina/Treonina conservado na região N-terminal da primeira fita- β , que precede o motivo Walker A. Essa sinapomorfia está presente em todas as septinas exceto duas subfamílias, uma presente em organismos dos filos Basidiomycota e Ascomycota, que possui um resíduo de cisteína conservado na mesma posição, e a outra presente no grande grupo Opisthokonta (clado que une animais, fungos e grupos relacionados da base dos Metazoa, como Choanoflagellata e *Dictyostelium*), representada no alinhamento pela proteína ORD96361.1, de *Hepatospora eriocheir*, do grupo Microsporidiae de fungos parasitas unicelulares, que perdeu completamente essa região.

Outra característica importante das septinas é uma incidência maior de substituições da treonina conservada das TRAFAC's, presente no *loop* que precede a fita- β adicional desse grupo, antiparalela à fita- β conservada que termina no motivo Walker B. Esse resíduo de treonina estabiliza um dos íon Mg^{+2} necessários para a hidrólise de GTP, mas é perdido em cerca de 80% das septinas, sugerindo que a atividade hidrolítica de parte dos membros dessa família pode ser abolida ou menor que a de outras GTPases. Uma região de aminoácidos negativamente carregados precede a fita- β do motivo Walker B, e outra região polar também é encontrada entre o Walker B e o motivo G4, no *loop* após a quinta fita-beta, dessa família de proteínas. Um resíduo de histidina, precedendo um ou dois resíduos de cisteína (motivo HCC, HXC ou HCX), é conservado em quase todas as septinas e delimita o início da quarta fita- β . Esse motivo está localizado na região da cauda, ou seja, na face do monômero oposta ao sítio catalítico. Essa região é responsável pelas interações proteína-proteína envolvidas na formação de polímeros de septinas (Angelis et al 2016). Outra região de aminoácidos carregados é encontrada entre as duas últimas fitas- β do domínio G. A última fita- β possui um resíduo de prolina conservado em sua região N-terminal e o motivo G5 geralmente possui uma glicina conservada antes do resíduo de serina típico das TRAFAC.

A maior parte das septinas possui uma extensão C-terminal, denominada elemento único das septinas (SUE para *Septin Unique Element*), que não tem similaridade detectável com nenhuma região ou domínio presente em outras GTPases mas é conservado entre as septinas. Esse elemento consiste em uma presilha- β seguida de duas alfas-hélices anfipáticas que formam estruturas *coiled-coil*. O SUE atua diretamente na interação entre monômeros de septinas, com papel determinante na formação de polímeros, filamentos e outros agrupamentos maiores de proteínas (Auxier et al. 2019). Apesar da semelhança no núcleo catalítico de septinas e parasseptinas, que produz o agrupamento *Septin-like*, as parasseptinas não possuem uma região similar ou análoga ao SUE, o que sugere que a aquisição deste elemento foi decisiva na divergência das septinas.

5.3.2 Distribuição da superfamília das septinas nos eucariotos

Para avaliar a distribuição das septinas, agregamos aos resultados da classificação aglomerativa os homólogos identificados, no nosso banco de dados de GTPases, usando como o modelo construído para o componente conectados das septinas como semente para buscas baseadas em hmmsearch. Esses genes estão amplamente distribuídos entre diversos grupos de Opisthokonta, com presença abundante em Metazoa e Fungi. Homólogos desse grupo de proteínas são encontrados nos genomas de anelídeos, como Poliquetos e Clitelados, cujo recente trabalho de genômica comparativa mostrou que seus genomas são mais similares ao de alguns deuterostômios como estrelas-do-mar e hemicordados do que de protostômios como artrópodes (Simakov et al. 2013); Diversos grupos de Artrópodes também possuem septinas e outros protostômios dos filos Tardigrada, Rotifera, Priapulida e Nematoda também apresentam ortólogos de septinas.

O grupo dos deuterostômios é o que apresenta maior diversidade de subfamílias de septinas e a presença destes genes pode ser observada entre os equinodermos, hemicordados e cordados (Tabela 4). A presença das septinas também pôde ser detectada entre os ramos mais antigos de metazoários, incluindo Placozoa, Orthonectida e Porifera (Tabela 4), sugerindo que esta família já cumpria um papel importante na construção do citoesqueleto nas células do ancestral metazoário. Ainda no clado Opisthokonta é possível observar essa família de proteínas em diversos fungos como o já citado Microsporidia, e outros filos como Basidiomycota, Blastocladiomycota, Chytridiomycota, Cryptomycota, Mucoromycota e Zoopagomycota (Tabela 4).

Tabela 4. Grupos de eucariotos com pelo menos um gene identificado como membro da família das septinas. Na primeira coluna está representado o filo dos organismos, na segunda o subfilo ou clado específico de cada grupo encontrado e na terceira o supergrupo definido pela mais recente filogenia de eucariotos (Burki, 2020). As células de fundo cinza claro indicam clados que incluem pelo menos um subgrupo no qual a presença de septinas só foi reconhecida no curso deste trabalho.

Filo	Grupo	Supergrupo	
Evosea	Amebozoa	Amorphea	
Chlorophyta	Chloroplastidae	Archaeplastida	
Rhodophyta			
Streptophyta			
Haptophyta	Haptista	Haptista	
Ciliophora	Alveolata	TSAR	
Annelida	Animal	Opisthokonta	
Brachiopoda			
Chordata			
Cnidaria			
Echinodermata			
Hemichordata			
Mollusca			
Nematoda			
Orthonectida			
Placozoa			
Platyhelminthes			
Porifera			
Priapulida			
Rotifera			
Tardigrada			
Arthropoda			
Ascomycota			Fungo
Basidiomycota			
Blastocladiomycota			
Chytridiomycota			
Cryptomycota			
Microsporidia			
Mucoromycota			
Zoopagomycota			
Choanoflagellata	Coanoflagelado		

As septinas também são encontradas em outras linhagens de eucariotos, incluindo Evosea (Figura 17), um grupo de Amebozoa do clado Amorphea, que une o grupo Amebozoa com o clado Opisthokonta, o que sugere que o ancestral do clado Amorphea já possuía genes de septinas em seu genoma. Notavelmente, essa é a primeira detecção de homólogos de septina nesse grupo de amebas (Momány et al. 2008; Nishihama et al. 2011; Onishi, 2016), o que fortalece a noção de que genes desta família já estavam presentes no ancestral mais recente de todo o grupo.

A distribuição das septinas, porém, não está restrita ao clado Opisthokonta, pois homólogos de septinas podem ser reconhecidos em algas verdes, como a proteína de *Chlamydomonas reinhardtii*, utilizada como uma das sementes em nossas buscas (ver seção 5.1, Tabela 1), em algas do clado Archaeplastida, como as algas vermelhas do filo Rhodophyta, e também em plantas multicelulares terrestres. A presença de septinas nas plantas terrestres não havia sido indicada anteriormente na literatura, mas é provável que estes genes participem da montagem do anel de divisão, na região do septo celular, como descrito anteriormente para algas verdes por (Yamazaki et al. 2013).

Os organismos do grupo Haptophyta, como *Chrysochromulina tobinii* e *Emiliania huxleyi*, também possuem genes de septinas. Esses organismos já foram classificados como membros do grupo Chromalveolata, com base na premissa de que os cloroplastos secundários presentes nestes organismos se originaram de um único episódio de assimilação, por endossimbiose, de uma alga eucariótica, que teria ocorrido na linhagem ancestral de todos esses organismos. Trabalhos mais recentes, porém, usaram o genoma nuclear desses organismos para demonstrar que o clado Chromalveolata é polifilético e reclassificaram Haptophyta no clado Haptista, separando Alveolata e Stramenopila no clado TSAR e Cryptophyta em Cryptista (Burki et al. 2020) (Figura 17).

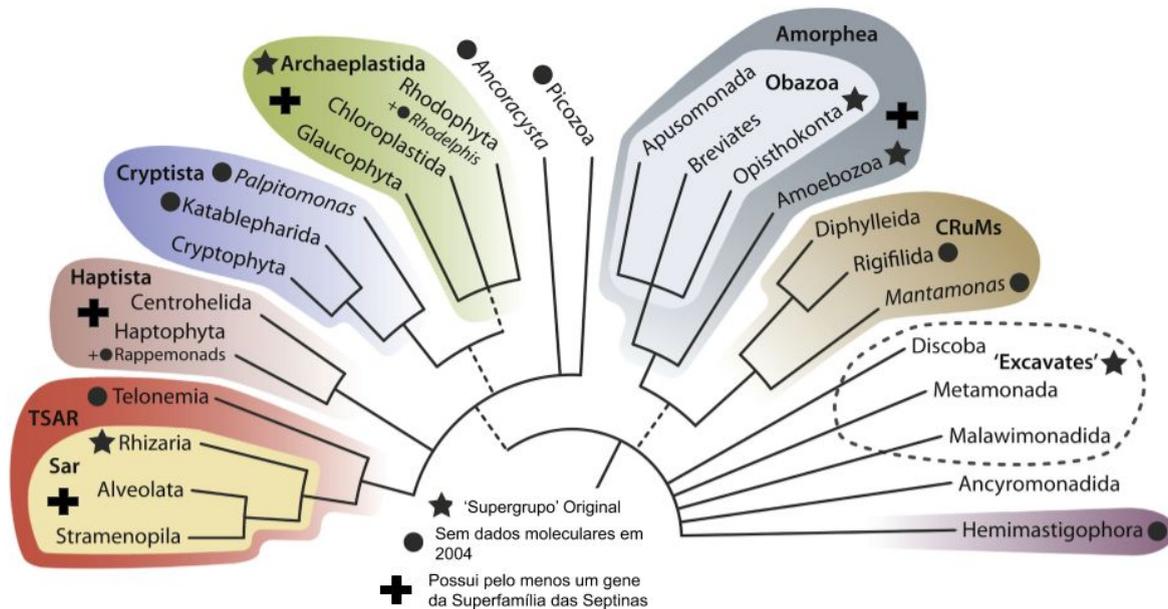


Figura 17. Árvore filogenética dos eucariotos construída por Burki e colaboradores em 2020, baseada no consenso de diversos estudos filogenômicos, mostrando as principais linhagens de eucariotos que atualmente temos conhecimento. Os grupos coloridos correspondem aos supergrupos atuais. As ordens não resolvidas das ramificações entre as linhagens estão mostradas como multifurcações. Linhas pontilhadas representam incertezas a respeito da origem monofilética de alguns grupos. Estrelas representam taxa que eram considerados supergrupos nas primeiras versões do modelo de supergrupo definidos por Bauldauf, em 2004; todos os supergrupos “originais” exceto Archaeplastida foram desmontados ou instituídos em novas taxas. Os círculos mostram grandes linhagens para as quais não havia dados moleculares quando o modelo de supergrupos surgiu e/ou não haviam sido descobertos até aquele momento. As cruzes indicam grupos nos quais foram identificados homólogos de septinas por este trabalho. Adaptado de Burki, et al. 2020.

Embora tenham sido encontradas GTPases de outras famílias em genomas de membros do antigo grupo Excavata, como proteínas da família Arf, outras linhagens da superfamília Ras e Fatores de Tradução, como as proteínas EKF29986.1 e XP_806835.1 (Arf e Fatores de Tradução Clássico, respectivamente) de *Trypanosoma cruzi*, não foi possível identificar nenhuma septina no genoma dos organismos desses grupos filogenéticos. O grupo Excavata foi recentemente revisado, com base em dados moleculares, e dividido em três linhagens distintas: Discoba, Metamonada e Malawimonadida (Burki et al. 2020). A ausência de Septinas na maioria dos grupos de basais eucariotos sugere duas hipóteses alternativas: (i) esses grupos divergiram antes do ancestral eucariótico adquirir o

primeiro gene homólogo da família das septinas, ou seja, o ancestral das septinas teria sido adquirido posteriormente, provavelmente por transferência lateral do último ancestral comum das mitocôndrias, como proposto por Leipe em 2002; ou (ii) esses grupos sofreram deleções nos genes que codificam para septinas. Considerando a filogenia da Figura 17 e a distribuição de membros da família das septinas, derivada nesse trabalho, um mínimo de cinco e um máximo de 10 perdas precisam ser postuladas para explicar a distribuição das septinas, dependendo de quais linhagens irmãs podem ser resolvidas em grupos monofiléticos. Por outro lado, um total de quatro eventos de ganho de homólogos de septinas já explicam a distribuição da família na mesma árvore dos eucariotos. Um único evento de ganho, seguido de cinco eventos de perda é outra hipótese aproximadamente equivalente para explicar a distribuição desses genes. Todas as alternativas, porém, com exceção da hipótese de presença de homólogos de septina em LECA antes da aquisição da mitocôndria, oscilam entre igualmente parcimoniosas ou são apenas marginalmente superiores, indicando que mais dados relacionados às linhagens basais de eucariotos são necessários para definirmos a hipótese que melhor explica a história evolutiva dessa família.

Nossa abordagem de classificação e definição das famílias identificou as parasseptinas bacterianas como o grupo mais intimamente relacionado às septinas eucarióticas. Embora amplamente distribuída entre a maioria dos eucariotos, a ausência de septinas ou parasseptinas em eucariotos que divergiram na base do clado (Baldauf, 2003) e nos grupos órfãos eucarióticos definidos por Burki, em 2020, implica que a análise de outros grupos de proteínas *Septin-like* e uma eventual amostragem mais ampla das linhagens basais de eucariotos serão necessárias para aprofundar a compreensão das origens das septinas eucarióticas.

5.4 GIMAP

O segundo grande grupo de proteínas *Septin-like* é a família GIMAP, que inclui diversas GTPases que, como as septinas, interagem com a membrana celular e com outras proteínas. As GIMAPs são necessárias no processo de maturação de linfócitos T e B em mamíferos (ver Introdução, seção 1.5.2.) e estão amplamente distribuídas entre as linhagens de eucariotos e em vírus de eucariotos, como Picornaviridae e Herpesviridae, observado por este trabalho e por Aravind et al. (2012).

Essa família de proteínas possui um resíduo de arginina conservado na região N-terminal do motivo Walker A, que contém um motivo semelhante ao das proteínas da classe SIMIBI, GxxGxGKS (consenso 80%). Além disso, um resíduo de glicina seguido por uma prolina e uma histidina são altamente conservados no *loop* que antecede a quarta fita- β no núcleo conservado dessa família. Essa histidina marca o início de uma região de aminoácidos hidrofóbicos conservada neste grupo, conhecida como *conserved box*, e que é uma sinapomorfia comum às GIMAPs e Tocs.

Algumas subfamílias de GIMAP de peixes ósseos (Actinopterygii) sofreram deleções nos resíduos que compõem o motivo Walker A, sugerindo a perda da capacidade de interação com o fosfato- γ da molécula de GTP e talvez da atividade hidrolítica nessas proteínas. É possível que a hidrólise de GTP não seja necessária para a função biológica desse grupo divergente de GIMAPs, uma vez que a deleção desses resíduos coincide com uma enorme expansão de parálogos, desse subgrupo de GIMAPs, em peixes ósseos, constituindo uma verdadeira expansão linhagem-específica que ocorre independentemente das múltiplas duplicações do genoma completas observadas em teleósteos (Meyer et al. 2005). A expansão das GIMAPs em teleósteos é um fenômeno único entre vertebrados e pode ser visualizada na filogenia da Figura 18. Expansões de linhagem-específica são frequentemente estimuladas por pressões seletivas que aumentam a probabilidade de fixação e divergência dos novos parálogos, não raro como resposta a uma

pressão ambiental ou interação ecológica, como a competição e/ou predação (Jordan, 2001).

No caso específico das GIMAPs de peixes, um estudo recente com o organismo modelo paulistinha (*zebra fish*) demonstrou que GIMAPs presentes em picornavirus e GIMAPs do hospedeiro têm sua expressão induzida durante a infecção viral (Balla et al. 2020). Embora o papel exato desses genes durante a infecção seja conhecido, o envolvimento de GIMAPs na maturação do sistema imune de vertebrados e na resposta inata a vírus já foi demonstrado (ver Introdução, seção 1.5.2). É possível que as múltiplas cópias de GIMAPs do hospedeiro poderiam atuar como dominantes negativos contra a interferência das GIMAPs virais na resposta imune do hospedeiro.

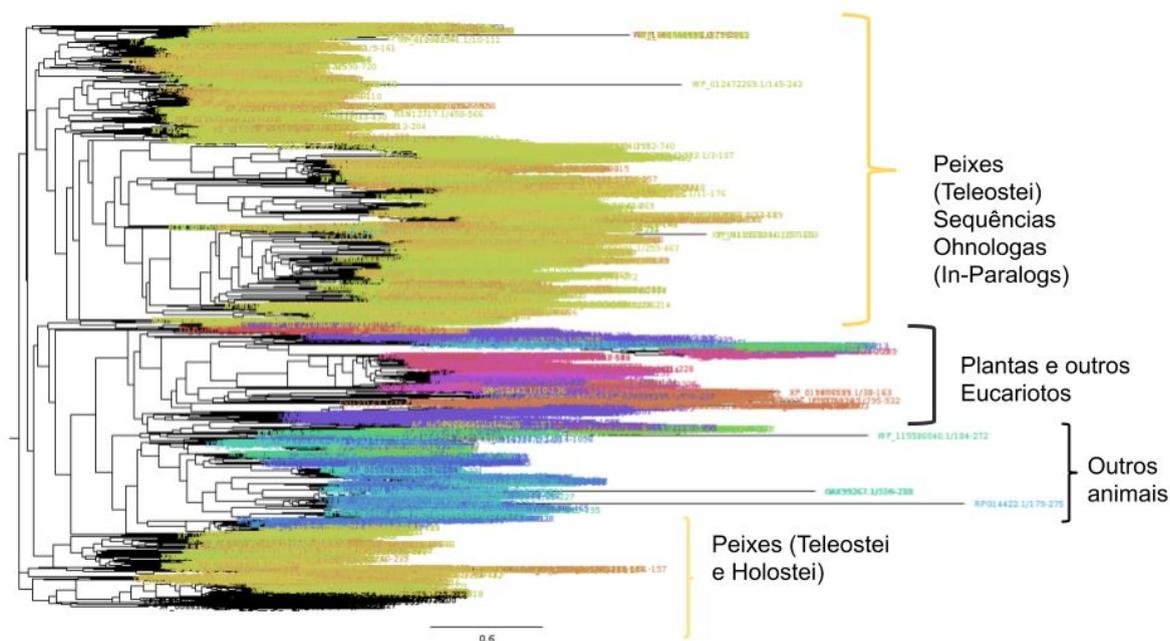


Figura 18. Árvore filogenética de uma das famílias de GIMAP, incluindo diversos grupos de eucariotos, com ênfase na ampla distribuição dessas proteínas em genomas de peixes que parecem ter evoluído de uma série de duplicações de genoma inteiro. A árvore de máxima verossimilhança foi obtida de um alinhamento curado de um dos clusters da família GIMAP utilizando o software FastTree (versão: 2.1.10).

6 CONCLUSÃO

Neste estudo, aplicamos técnicas de agrupamento a sequências e pHMMs, que representavam populações de sequências, de P-loop NTPases da classe TRAFAC. Esses métodos permitiram identificar famílias novas e conhecidas, derivando uma classificação natural baseada na similaridade e na história evolutiva das proteínas analisadas. Foram desenvolvidos e implementados dois protocolos semi-automáticos de análise e construção de grupos, que combinam métodos de agrupamento com algoritmos de análise de redes de similaridade de sequências.

O primeiro desses protocolos se baseia na aplicação do algoritmo de Leiden, para inferência de comunidades, na rede de relações de similaridade entre pares de GTPases representativas de grupos da classe TRAFAC. O segundo protocolo foi baseado na detecção de componentes conectados em uma rede de similaridade de pHMMs derivados dos alinhamentos de membros de cada grupo de sequências, automaticamente definido pelo programa MMSeqs com base na similaridade entre as sequências de GTPases. O segundo protocolo revelou maior especificidade na classificação das diferentes famílias, ao reproduzir os agrupamentos já estabelecidos na literatura do grupo e isolar P-loop NTPases que não pertenciam à classe TRAFAC. Esse foi um resultado surpreendente pois todos os estudos anteriores que classificaram P-loop NTPases dependeram da análise manual das estruturas tridimensionais dessas proteínas para identificar sinapomorfias que permitissem separar membros das TRAFAC de outras classes, como SIMIBI e as ATPases de transportadores ABC. Sendo assim acreditamos que o uso de perfis de HMMs, que modelam, explicitamente, a variação de composição e a distribuição de inserções ou deleções observadas em cada grupo de sequências, seja o responsável pelo aumento da especificidade do segundo protocolo.

A aplicação de medidas de distância entre pHMMs na inferência de relações entre grupos altamente divergentes, aqui ilustrada com membros da classe TRAFAC, pode vir a constituir uma abordagem geral e rigorosa para inferência de relações entre grupos extremamente divergentes de proteínas, eventualmente complementar à construção de filogenias, especialmente quando lidando com os problemas de saturação em substituições de resíduos. Adicionalmente, os perfis de HMM desenhados neste trabalho servirão para a busca e anotação dessas famílias de GTPases em novos genomas.

Embora não tenhamos tratado aqui de muitas das famílias das GTPases que foram identificadas no curso dessa análise, as famílias de dinaminas presentes em bactérias (grupo 4), conhecidas como BDLP's, são particularmente interessantes por sua associação com processos de remodelagem de membrana e citoesqueleto e porque não terem sido, até o momento, amplamente exploradas na literatura experimental, fazendo com que a função de muitas subfamílias permaneça desconhecida. A posição de BDLP36 (LeoA) em nosso dendrograma, como linhagem irmã de ERA e fatores de tradução, é particularmente interessante, pois pode implicar um viés do segundo protocolo ou uma relação real. Pretendemos futuramente explorar essas possibilidades suplementando a análise de similaridade com dados de vizinhança gênica, que permitem a inferência, *in-silico*, de funções biológicas e interações entre genes.

Dentre os seis grupos de TRAFAC definidos neste trabalho, o grupo das septinas/parasseptinas inclui as parasseptinas bacterianas, resultado que estabelece, definitivamente, que estas proteínas são os melhores candidatos para estudo das atividades similares às da septinas eucarióticas, dada sua ampla distribuição entre as bactérias.

7 REFERÊNCIAS

AGNE, B.; KESSLER, F. Protein transport in organelles: The Toc complex way of preprotein import. **The FEBS journal**, v. 276, n. 5, p. 1156–1165, 2009.

ALTSCHUL, S. F.; MADDEN, T. L.; SCHÄFFER, A. A.; et al. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. **Nucleic acids research**, v. 25, n. 17, p. 3389–3402, 1997.

ANGELIS, D.; SPILIOTIS, E. T. Septin Mutations in Human Cancers. **Frontiers in Cell and Developmental Biology**, v. 4, 2016. Frontiers. Disponível em: <<https://www.frontiersin.org/articles/10.3389/fcell.2016.00122/pdf>>. Acesso em: 8/7/2020.

ARASU, B.S. Structural and Biochemical Analysis of GTPases of IMMunity-Associated Proteins (GIMAPs) and their Interaction Partners. Ph.D thesis Max-Delbrück-Centrum für Molekulare Medizin Berlin-Buch. Fachbereich Biologie, Chemie, Pharmazie der Freien Universität Berlin. Berlin; 2015.

ARAVIND, L.; ANANTHARAMAN, V.; ZHANG, D.; DE SOUZA, R. F.; IYER, L. M. Gene flow and biological conflict systems in the origin and evolution of eukaryotes. **Frontiers in cellular and infection microbiology**, v. 2, 2012. Frontiers. Disponível em: <<https://www.frontiersin.org/articles/10.3389/fcimb.2012.00089/pdf>>. Acesso em: 27/8/2020.

AUXIER, B.; DEE, J.; BERBEE, M. L.; MOMANY, M. Diversity of opisthokont septin proteins reveals structural constraints and conserved motifs. **BMC evolutionary biology**, v. 19, n. 1, p. 1–11, 2019. BioMed Central. Acesso em: 8/7/2020.

BALASINGAM, N.; BRANDON, H. E.; ROSS, J. A.; WIEDEN, H.-J.; THAKOR, N. Cellular roles of the human Obg-like ATPase 1 (hOLA1) and its YchF homologs. **Biochemistry and cell biology = Biochimie et biologie cellulaire**, v. 98, n. 1, p. 1–11, 2020.

BALLA, K. M.; RICE, M. C.; GAGNON, J. A.; ELDE, N. C. Linking Virus Discovery to Immune Responses Visualized during Zebrafish Infections. **Current biology: CB**, v. 30, n. 11, p. 2092–2103.e5, 2020. Cell Press. Acesso em: 27/8/2020.

BALDAUF, S. L. The Deep Roots of Eukaryotes. **Science**, 2003. Disponível em: <<http://dx.doi.org/10.1126/science.1085544>>.

BARABÁSI, A.-L. Network science. **Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences**, 2013. Disponível em: <<http://dx.doi.org/10.1098/rsta.2012.0375>>.

BATEMAN, A. The Pfam Protein Families Database. **Nucleic Acids Research**, 2002. Disponível em: <<http://dx.doi.org/10.1093/nar/30.1.276>>. .

BERMAN, H. M. The Protein Data Bank. **Nucleic acids research**, v. 28, n. 1, p. 235–242, 2000.

BILLER, L.; DAVIS, P. H.; TILLACK, M.; MATTHIESEN, J.; LOTTER, H.; STANLEY JR, S. L.; TANNICH, E.; BRUCHHAUS, I. Differences in the transcriptome signatures of two genetically related *Entamoeba histolytica* cell lines derived from the same isolate with different pathogenic properties. **BMC Genomics**. 2010 11:63.

BLANCO-MÍGUEZ, A.; FDEZ-RIVEROLA, F.; SÁNCHEZ, B.; LOURENÇO, A. BlasterJS: A novel interactive JavaScript visualisation component for BLAST alignment results. **PloS one**, v. 13, n. 10, p. e0205286, 2018.

BLOMBACH, F.; LAUNAY, H.; ZORRAQUINO, V.; et al. An HflX-type GTPase from *Sulfolobus solfataricus* binds to the 50S ribosomal subunit in all nucleotide-bound states. **Journal of bacteriology**, v. 193, n. 11, p. 2861–2867, 2011.

BOS, J. L.; REHMANN, H.; WITTINGHOFER, A. GEFs and GAPs: critical elements in the control of small G proteins. **Cell**, v. 129, n. 5, p. 865–877, 2007.

BURKI, F.; ROGER, A. J.; BROWN, M. W.; SIMPSON, A. G. B. The New Tree of Eukaryotes. **Trends in Ecology & Evolution**, 2020. Disponível em: <<http://dx.doi.org/10.1016/j.tree.2019.08.008>>. .

BYERS, B.; GOETSCH, L. A highly ordered ring of membrane-associated filaments in budding yeast. **The Journal of cell biology**, v. 69, n. 3, p. 717–721, 1976.

CAMPBELL, T. L.; BROWN, E. D. Genetic interaction screens with ordered overexpression and deletion clone sets implicate the *Escherichia coli* GTPase YjeQ in late ribosome biogenesis. **Journal of bacteriology**, v. 190, n. 7, p. 2537–2545, 2008.

CAO, L.; DING, X.; YU, W.; et al. Phylogenetic and evolutionary analysis of the septin protein family in metazoan. **FEBS letters**, v. 581, n. 28, p. 5526–5532, 2007.

CAO, Y.-L.; MENG, S.; CHEN, Y.; et al. MFN1 structures reveal nucleotide-triggered dimerization critical for mitochondrial fusion. **Nature**, v. 542, n. 7641, p. 372–376, 2017.

CHATTERJEE, A.; ACHARJEE, A.; DAS, S.; DATTA, P. P. Deletion analyses reveal insights into the domain specific activities of an essential GTPase CgtA in *Vibrio cholerae*. **Archives of biochemistry and biophysics**, v. 665, p. 143–151, 2019. Academic Press. Acesso em: 11/8/2020.

CLAMP, M.; CUFF, J.; SEARLE, S. M.; BARTON, G. J. The Jalview Java alignment editor. **Bioinformatics**, v. 20, n. 3, p. 426–427, 2004. Oxford Academic. Acesso em: 26/8/2020.

COPP, J. N.; AKIVA, E.; BABBITT, P. C.; TOKURIKI, N. Revealing Unexplored Sequence-Function Space Using Sequence Similarity Networks. **Biochemistry**, v. 57, n. 31, p. 4651–4662, 2018.

CZYZ, A.; WEGRZYN, G. The Obg subfamily of bacterial GTP-binding proteins: essential proteins of largely unknown functions that are evolutionarily conserved from bacteria to humans. **Acta biochimica Polonica**, v. 52, n. 1, p. 35–43, 2005.

EDGAR, R. C. MUSCLE: multiple sequence alignment with high accuracy and high throughput. **Nucleic acids research**, v. 32, n. 5, p. 1792–1797, 2004. Oxford Academic. Acesso em: 26/8/2020.

EDDY, S. R. "Accelerated Profile HMM Searches." **PLoS computational biology**. 7(10): e1002195. 2011.

EME, L.; SPANG, A.; LOMBARD, J.; STAIRS, C. W.; ETTEMA, T. J. G. Archaea and the origin of eukaryotes. **Nature reviews. Microbiology**, v. 16, n. 2, p. 120, 2018.

FERGUSON, S. M.; DE CAMILLI, P. Dynamin, a membrane-remodelling GTPase. **Nature reviews. Molecular cell biology**, v. 13, n. 2, p. 75–88, 2012.

FILÉN, S.; LAHESMAA, R. GIMAP Proteins in T-Lymphocytes. **Journal of signal transduction**, v. 2010, p. 268589, 2010.

FINN, R. D.; BATEMAN, A.; CLEMENTS, J.; et al. Pfam: the protein families database. **Nucleic acids research**, v. 42, n. Database issue, p. D222–30, 2014.

GORBALENYA, A. E.; KOONIN, E. V. Superfamily of UvrA-related NTP-binding proteins. Implications for rational classification of recombination/repair systems. **Journal of molecular biology**, v. 213, n. 4, p. 583–591, 1990.

GOTO, S.; KATO, S.; KIMURA, T.; MUTO, A.; HIMENO, H. RsgA releases RbfA from 30S ribosome during a late stage of ribosome biosynthesis. **The EMBO journal**, v. 30, n. 1, p. 104–114, 2011.

GUTENSOHN, M.; SCHULZ, B.; NICOLAY, P.; FLÜGGE, U. I. Functional analysis of the two Arabidopsis homologues of Toc34, a component of the chloroplast protein import apparatus. **The Plant journal: for cell and molecular biology**, v. 23, n. 6, p. 771–783, 2000.

HAARER, B. K.; PRINGLE, J. R. Immunofluorescence localization of the *Saccharomyces cerevisiae* CDC12 gene product to the vicinity of the 10-nm filaments in the mother-bud neck. **Molecular and cellular biology**, v. 7, n. 10, p. 3678–3687, 1987.

HADI, A. S.; KAUFMAN, L.; ROUSSEEUW, P. J. Finding Groups in Data: An Introduction to Cluster Analysis. **Technometrics**, 1992. Disponível em: <<http://dx.doi.org/10.2307/1269576>>.

HALLER, O.; STAEHELI, P.; SCHWEMMLE, M.; KOCHS, G. Mx GTPases: dynamin-like antiviral machines of innate immunity. **Trends Microbiol.** 2015; 23(3):154-163. doi:10.1016/j.tim.2014.12.003.

HARTMAN, H.; FEDOROV, A. The origin of the eukaryotic cell: a genomic investigation. **Proceedings of the National Academy of Sciences of the United States of America**, v. 99, n. 3, p. 1420–1425, 2002.

HARTWELL, L. H. Genetic control of the cell division cycle in yeast. IV. Genes controlling bud emergence and cytokinesis. **Experimental cell research**, v. 69, n. 2, p. 265–276, 1971.

HAYASHI, I.; OYAMA, T.; MORIKAWA, K. Structural and functional studies of MinD ATPase: implications for the molecular recognition of the bacterial cell division apparatus. **The EMBO journal**, v. 20, n. 8, p. 1819–1828, 2001.

HO, C.-H.; TSAI, S.-F. Functional and biochemical characterization of a T cell-associated anti-apoptotic protein, GIMAP6. **The Journal of biological chemistry**, v. 292, n. 22, p. 9305–9319, 2017.

HUNG, K.-W.; CHANG, Y.-W.; ENG, E. T.; et al. Structural fold, conservation and Fe(II) binding of the intracellular domain of prokaryote FeoB. **Journal of structural biology**, v. 170, n. 3, p. 501–512, 2010.

IYER, L. M.; LEIPE, D. D.; KOONIN, E. V.; ARAVIND, L. Evolutionary history and higher order classification of AAA+ ATPases. **Journal of structural biology**, v. 146, n. 1-2, p. 11–31, 2004.

JORDAN, I. K. Lineage-Specific Gene Expansions in Bacterial and Archaeal Genomes. **Genome Research**, 2001. Disponível em: <<http://dx.doi.org/10.1101/gr-gr-1660r>>.

KALEDHONKAR, S.; FU, Z.; CABAN, K.; et al. Late steps in bacterial translation initiation visualized using time-resolved cryo-EM. **Nature**, v. 570, n. 7761, p. 400–404, 2019.

KATOH, K.; KUMA, K.-I.; TOH, H.; MIYATA, T. MAFFT version 5: improvement in accuracy of multiple sequence alignment. **Nucleic acids research**, v. 33, n. 2, p. 511–518, 2005. Oxford Academic. Acesso em: 26/8/2020.

KENNER, L. R.; ANAND, A. A.; NGUYEN, H. C.; et al. eIF2B-catalyzed nucleotide exchange and phosphoregulation by the integrated stress response. **Science**, v. 364, n. 6439, p. 491–495, 2019.

KOENIG, P.; OREB, M.; HÖFLE, A.; et al. The GTPase cycle of the chloroplast import receptors Toc33/Toc34: implications from monomeric and dimeric structures. **Structure** , v. 16, n. 4, p. 585–596, 2008.

KRÜCKEN, J.; SCHROETEL, R. M. U.; MÜLLER, I. U.; et al. Comparative analysis of the human gimap gene cluster encoding a novel GTPase family. **Gene**, v. 341, p. 291–304, 2004.

LARSSON, A. AliView: a fast and lightweight alignment viewer and editor for large datasets. **Bioinformatics** , v. 30, n. 22, p. 3276–3278, 2014. Oxford Academic. Acesso em: 26/8/2020.

LEIPE, D. D.; WOLF, Y. I.; KOONIN, E. V.; ARAVIND, L. Classification and evolution of P-loop GTPases and related ATPases. **Journal of molecular biology**, v. 317, n. 1, p. 41–72, 2002.

LIU, C.; WANG, T.; ZHANG, W.; LI, X. Computational identification and analysis of immune-associated nucleotide gene family in Arabidopsis thaliana. **Journal of plant physiology**, v. 165, n. 7, p. 777–787, 2008.

LONGO, L. M.; JABŁOŃSKA, J.; VYAS, P.; et al. On the Emergence of P-Loop NTPase and Rossmann Enzymes from a Beta-Alpha-Beta Ancestral Fragment. , 2020. Disponível em: <<http://dx.doi.org/10.1101/2020.07.15.204123>>. .

LONGTINE, M. S.; DEMARINI, D. J.; VALENCIK, M. L.; et al. The septins: roles in cytokinesis and other processes. **Current opinion in cell biology**, v. 8, n. 1, p. 106–119, 1996.

LOW, H. H.; LÖWE, J. A bacterial dynamin-like protein. **Nature**, 2006. Disponível em: <<http://dx.doi.org/10.1038/nature05312>>. .

LUTKENHAUS, J. The ParA/MinD family puts things in their place. **Trends in microbiology**, v. 20, n. 9, p. 411–418, 2012.

MASO, L.; GALAZZO, L.; VALLESE, F.; et al. A conformational study of the GTPase domain of [FeFe]-hydrogenase maturation protein HydF by PELDOR spectroscopy. **Applied Magnetic Resonance**, 2015. Disponível em: <<http://dx.doi.org/10.1007/s00723-015-0641-z>>. .

MCDOWELL, I. C.; MODAK, T. H.; LANE, C. E.; GOMEZ-CHIARRI, M. Multi-species protein similarity clustering reveals novel expanded immune gene families in the eastern oyster Crassostrea virginica. **Fish & shellfish immunology**, v. 53, p. 13–23, 2016.

MCMURRAY, M. A.; BERTIN, A.; GARCIA, G., 3rd; et al. Septin filament formation is essential in budding yeast. **Developmental cell**, v. 20, n. 4, p. 540–549, 2011.

MÉHEUST, R.; BURSTEIN, D.; CASTELLE, C. J.; BANFIELD, J. F. The distinction of CPR bacteria from other bacteria based on protein family content. **Nature communications**, v. 10, n. 1, p. 4173, 2019.

MEYER, A.; VAN DE PEER, Y. From 2R to 3R: evidence for a fish-specific genome duplication (FSGD). **BioEssays: news and reviews in molecular, cellular and developmental biology**, v. 27, n. 9, p. 937–945, 2005.

MICHIE, K. A.; BOYSEN, A.; LOW, H. H.; MØLLER-JENSEN, J.; LÖWE, J. LeoA, B and C from Enterotoxigenic Escherichia coli (ETEC) Are Bacterial Dynamins. **PLoS ONE**, 2014. Disponível em: <<http://dx.doi.org/10.1371/journal.pone.0107211>>. .

MILNER-WHITE, E. J.; COGGINS, J. R.; ANTON, I. A. Evidence for an ancestral core structure in nucleotide-binding proteins with the type A motif. **Journal of molecular biology**, v. 221, n. 3, p. 751–754, 1991.

MONTOYA, G.; TE KAAT, K.; MOLL, R.; SCHÄFER, G.; SINNING, I. The crystal structure of the conserved GTPase of SRP54 from the archaeon *Acidianus ambivalens* and its comparison with related structures suggests a model for the SRP–SRP receptor complex. **Structure**, 2000. Disponível em: <[http://dx.doi.org/10.1016/s0969-2126\(00\)00131-3](http://dx.doi.org/10.1016/s0969-2126(00)00131-3)>. .

MOSTOWY, S.; COSSART, P. Septins: the fourth component of the cytoskeleton. **Nature reviews. Molecular cell biology**, v. 13, n. 3, p. 183–194, 2012.

NAKADA-TSUKUI K.; SEKIZUKA T.; SATO-EBINE E.; ESCUETA-DE CADIZ, A.; JI D.; TOMII K. AIG1 affects in vitro and in vivo virulence in clinical isolates of *Entamoeba histolytica*. **PLoS Pathogens** 14. 2018.

NISHIHAMA, R.; ONISHI, M.; PRINGLE, J. R. New insights into the phylogenetic distribution and evolutionary origins of the septins. **Biological chemistry**, v. 392, n. 8-9, p. 681–687, 2011.

NITTA, T.; NASREEN, M.; SEIKE, T.; et al. IAN family critically regulates survival and development of T lymphocytes. **PLoS biology**, v. 4, n. 4, p. e103, 2006.

PANDARPURKAR, M.; WILSON-FRITCH, L.; CORVERA, S.; et al. IAN4 is required for mitochondrial integrity and T cell survival. **Proceedings of the National Academy of Sciences**, v. 100, n. 18, p. 10382–10387, 2003.

PAN, F.; MALMBERG, R. L.; MOMANY, M. Analysis of septins across kingdoms reveals orthology and new motifs. **BMC evolutionary biology**, v. 7, p. 103, 2007.

PINTO, A. P. A.; PEREIRA, H. M.; ZERAIK, A. E.; et al. Filaments and fingers: Novel structural aspects of the single septin from. **The Journal of biological chemistry**, v. 292, n. 26, p. 10899–10911, 2017.

POIRIER, G. M.; ANDERSON, G.; HUVAR, A.; et al. Immune-associated nucleotide-1 (IAN-1) is a thymic selection marker and defines a novel gene family conserved in plants. **Journal of immunology**, v. 163, n. 9, p. 4960–4969, 1999.

POLKINGHORNE, A.; ZIEGLER, U.; GONZÁLEZ-HERNÁNDEZ, Y.; et al. Chlamydomonas reinhardtii HflX belongs to an uncharacterized family of conserved GTPases and associates with the Escherichia coli 50S large ribosomal subunit. **Microbiology**, 2008. Disponível em: <<http://dx.doi.org/10.1099/mic.0.2008/022137-0>>. .

POTTER, S. C.; LUCIANI, A.; EDDY, S. R.; et al. HMMER web server: 2018 update. **Nucleic acids research**, v. 46, n. W1, p. W200–W204, 2018.

PRICE, M. N.; DEHAL, P. S.; ARKIN, A. P. FastTree 2 – Approximately Maximum-Likelihood Trees for Large Alignments. **PLoS ONE**, 2010. Disponível em: <<http://dx.doi.org/10.1371/journal.pone.0009490>>. .

RAMBAUT, A.; Figtree, versão 1.4.3. Programa computacional distribuído pelo autor. <<http://tree.bio.ed.ac.uk/software/figtree/>> acesso em: 08/09/2020.

RAZI, A.; DAVIS, J. H.; HAO, Y.; et al. Role of Era in assembly and homeostasis of the ribosomal small subunit. **Nucleic acids research**, v. 47, n. 15, p. 8301–8317, 2019.

REUBER, T. L.; AUSUBEL, F. M. Isolation of Arabidopsis genes that differentiate between resistance responses mediated by the RPS2 and RPM1 disease resistance genes. **The Plant cell**, v. 8, n. 2, p. 241–249, 1996.

RICHARDSON, L. G. L.; PAILA, Y. D.; SIMAN, S. R.; et al. Targeting and assembly of components of the TOC protein import complex at the chloroplast outer envelope membrane. **Frontiers in plant science**, v. 5, 2014. Frontiers. Disponível em: <<https://www.frontiersin.org/articles/10.3389/fpls.2014.00269/pdf>>. Acesso em: 11/5/2020.

SARASTE, M.; SIBBALD, P. R.; WITTINGHOFER, A. The P-loop — a common motif in ATP- and GTP-binding proteins. **Trends in biochemical sciences**, v. 15, n. 11, p. 430–434, 1990.

SAUNDERS, A.; WEBB, L. M. C.; JANAS, M. L.; et al. Putative GTPase GIMAP1 is critical for the development of mature B and T lymphocytes. **Blood**, v. 115, n. 16, p. 3249–3257, 2010.

SCHINDELIN, H.; KISKER, C.; SCHLESSMAN, J. L.; HOWARD, J. B.; REES, D. C. Structure of ADP·AIF4—stabilized nitrogenase complex and its implications for signal transduction. **Nature**, 1997. Disponível em: <<http://dx.doi.org/10.1038/387370a0>>. .

SCHLEIFF, E.; BECKER, T. Common ground for protein translocation: access control for mitochondria and chloroplasts. **Nature reviews. Molecular cell biology**, v. 12, n. 1, p. 48–59, 2010. Nature Publishing Group. Acesso em: 11/5/2020.

SCHNELL, S.; DÉMOLLIÈRE, C.; VAN DEN BERK, P.; JACOBS, H. Gimap4 accelerates T-cell death. **Blood**, v. 108, n. 2, p. 591–599, 2006.

SCHRODINGER, L. L. C.; The PyMOL Molecular Graphics System, Version 2.0. 2020. Software distribuído digitalmente. <<https://www.schrodinger.com/pymol>>. Acesso em: 08/09/2020.

SCHULZ, G. E. Binding of nucleotides by proteins: Current opinion in structural biology 1992, 2: 61...-67. **Current opinion in structural biology**, v. 2, n. 1, p. 61–67, 1992.

SCHWEFEL, D.; FRÖHLICH, C.; EICHHORST, J.; et al. Structural basis of oligomerization in septin-like GTPase of immunity-associated protein 2 (GIMAP2). **Proceedings of the National Academy of Sciences of the United States of America**, v. 107, n. 47, p. 20299–20304, 2010.

SCHWEINS, T.; WITTINGHOFER, A. GTP-binding proteins. Structures, interactions and relationships. **Current biology: CB**, v. 4, n. 6, p. 547–550, 1994.

SHIN, M.; MEY, A. R.; PAYNE, S. M. Vibrio cholerae FeoB contains a dual nucleotide-specific NTPase domain essential for ferrous iron uptake. **Proceedings of the National Academy of Sciences of the United States of America**, v. 116, n. 10, p. 4599–4604, 2019. National Academy of Sciences. Acesso em: 14/9/2020.

SIMAKOV, O.; MARLETAZ, F.; CHO, S.-J.; et al. Insights into bilaterian evolution from three spiralian genomes. **Nature**, v. 493, n. 7433, p. 526–531, 2013.

SIRAJUDDIN, M.; FARKASOVSKY, M.; HAUER, F.; et al. Structural insight into filament formation by mammalian septins. **Nature**, v. 449, n. 7160, p. 311–315, 2007.

SÖDING, J. Protein homology detection by HMM-HMM comparison. **Bioinformatics**, v. 21, n. 7, p. 951–960, 2005.

SPANG, A.; SAW, J. H.; JØRGENSEN, S. L.; et al. Complex archaea that bridge the gap between prokaryotes and eukaryotes. **Nature**, v. 521, n. 7551, p. 173–179, 2015.

STEINEGGER, M.; MEIER, M.; MIRDITA, M.; et al. HH-suite3 for fast remote homology detection and deep protein annotation. **BMC bioinformatics**, v. 20, n. 1, p. 473, 2019.

STEINEGGER, M.; SÖDING, J. MMseqs2 enables sensitive protein sequence searching for the analysis of massive data sets. **Nature biotechnology**, v. 35, n. 11,

p. 1026–1028, 2017.

TALAVERA, A.; HENDRIX, J.; VERSÉES, W.; et al. Phosphorylation decelerates conformational dynamics in bacterial translation elongation factors. **Science advances**, v. 4, n. 3, p. eaap9714, 2018.

THURLOW, B.; DAVIS, J. H.; LEONG, V.; et al. Binding properties of YjeQ (RsgA), RbfA, RimM and Era to assembly intermediates of the 30S subunit. **Nucleic acids research**, v. 44, n. 20, p. 9918–9932, 2016.

TRAAG, V. A.; WALTMAN, L.; VAN ECK, N. J. From Louvain to Leiden: guaranteeing well-connected communities. **Scientific reports**, v. 9, n. 1, p. 1–12, 2019. Nature Publishing Group. Acesso em: 13/5/2020.

TU, C.; ZHOU, X.; TARASOV, S. G.; et al. The Era GTPase recognizes the GAUCACCUCC sequence and binds helix 45 near the 3' end of 16S rRNA. **Proceedings of the National Academy of Sciences of the United States of America**, v. 108, n. 25, p. 10156–10161, 2011.

TU, C.; ZHOU, X.; TROPEA, J. E.; et al. Structure of ERA in complex with the 3' end of 16S rRNA: implications for ribosome biogenesis. **Proceedings of the National Academy of Sciences of the United States of America**, v. 106, n. 35, p. 14843–14848, 2009.

VETTER, I. R.; WITTINGHOFER, A. Nucleoside triphosphate-binding proteins: different scaffolds to achieve phosphoryl transfer. **Quarterly reviews of biophysics**, v. 32, n. 1, p. 1–56, 1999.

WALKER, J. E.; SARASTE, M.; RUNSWICK, M. J.; GAY, N. J. Distantly related sequences in the alpha- and beta-subunits of ATP synthase, myosin, kinases and other ATP-requiring enzymes and a common nucleotide binding fold. **The EMBO journal**, v. 1, n. 8, p. 945–951, 1982.

WANG, Z.; LI, X. IAN/GIMAPs are conserved and novel regulators in vertebrates and angiosperm plants. **Plant signaling & behavior**, v. 4, n. 3, p. 165–167, 2009.

WEISS, Y.; FORÊT, S.; HAYWARD, D. C.; et al. The acute transcriptional response of the coral *Acropora millepora* to immune challenge: expression of GiMAP/IAN genes links the innate immune responses of corals with those of mammals and plants. **BMC genomics**, v. 14, p. 400, 2013.

WENNERBERG, K. The Ras superfamily at a glance. **Journal of Cell Science**, 2005. Disponível em: <<http://dx.doi.org/10.1242/jcs.01660>>. .

WICKHAM, H.; AVERICK, M.; BRYAN, J.; et al. Welcome to the Tidyverse. **Journal of Open Source Software**, v. 4, n. 43, p. 1686, 2019. Acesso em: 26/8/2020.

- WILLIAMS, P. H.; CARBONETTI, N. H. The Plasmid-Specified Aerobactin Iron Uptake System of Escherichia Coli. **Plasmids in Bacteria**, 1985. Disponível em: <http://dx.doi.org/10.1007/978-1-4613-2447-8_51>.
- WIRTH, C.; ANDERSEN, G. R.; HUNTE, C. Crystal structure analysis of the yeast elongation factor complex EEF1A:EEF1BA. , 2017. Disponível em: <<http://dx.doi.org/10.2210/pdb5o8w/pdb>>.
- WOESE, C. R.; FOX, G. E. Phylogenetic structure of the prokaryotic domain: The primary kingdoms. **Proceedings of the National Academy of Sciences**, v. 74, n. 11, p. 5088–5090, 1977.
- WOESE, C. R.; KANDLER, O.; WHEELIS, M. L. Towards a natural system of organisms: proposal for the domains Archaea, Bacteria, and Eucarya. **Proceedings of the National Academy of Sciences of the United States of America**, v. 87, n. 12, p. 4576–4579, 1990.
- YAMAZAKI, T.; OWARI, S.; OTA, S.; et al. Localization and evolution of septins in algae. **The Plant journal: for cell and molecular biology**, v. 74, n. 4, p. 605–614, 2013.
- ZAREMBA-NIEDZWIEDZKA, K.; CACERES, E. F.; SAW, J. H.; et al. Asgard archaea illuminate the origin of eukaryotic cellular complexity. **Nature**, v. 541, n. 7637, p. 353–358, 2017.
- ZHANG, S.-M.; LOKER, E. S.; SULLIVAN, J. T. Pathogen-associated molecular patterns activate expression of genes involved in cell proliferation, immunity and detoxification in the amebocyte-producing organ of the snail *Biomphalaria glabrata*. **Developmental and comparative immunology**, v. 56, p. 25–36, 2016.
- ZHOU, T.; RADAEV, S.; ROSEN, B. P.; GATTI, D. L. Structure of the ArsA ATPase: the catalytic subunit of a heavy metal resistance pump. **The EMBO journal**, v. 19, n. 17, p. 4838–4845, 2000.
- ZILLIG, W.; KLENK, H.-P.; PALM, P.; et al. The phylogenetic relations of DNA-dependent RNA polymerases of archaebacteria, eukaryotes, and eubacteria. **Canadian journal of microbiology**, v. 35, n. 1, p. 73–80, 1989.

Anexo 1 - Classificação das GTPases estruturalmente caracterizadas (PDBs classificados)

ID	Arquiteturas de domínios - Pfam	Arquiteturas de domínios - Prof. Aravind	Grupo	Classificação	CC	Família
5YST_A	Ras	sGTP.1	1	Ras-like	1	Ras Superfamily
3RAN_A	Ras	sGTP.1	1	Ras-like	1	Ras Superfamily
1QG4_A	Ras	sGTP.1	1	Ras-like	1	Ras Superfamily
1QBK_C	Ras	sGTP.1	1	Ras-like	1	Ras Superfamily
3EA5_A	Ras	sGTP.1	1	Ras-like	1	Ras Superfamily
5CIW_A	Ras	sGTP.1	1	Ras-like	1	Ras Superfamily
1BYU_A	Ras	sGTP.1	1	Ras-like	1	Ras Superfamily
5YTB_A	Ras	sGTP.1	1	Ras-like	1	Ras Superfamily
5YRO_A	Ras	sGTP.1	1	Ras-like	1	Ras Superfamily
1QG2_A	Ras	sGTP.1	1	Ras-like	1	Ras Superfamily
4C0Q_C	Ras	sGTP.1	1	Ras-like	1	Ras Superfamily
4PHG_A	Ras	sGTP.1	1	Ras-like	1	Ras Superfamily
4PHF_A	Ras	sGTP.1	1	Ras-like	1	Ras Superfamily
1KY2_A	Ras	sGTP.1	1	Ras-like	1	Ras Superfamily
5XC5_A	Ras	sGTP.1	1	Ras-like	1	Ras Superfamily
5XC3_A	Ras	sGTP.1	1	Ras-like	1	Ras Superfamily
3RAP_R	Ras	sGTP.1	1	Ras-like	1	Ras Superfamily
4QXA_A	Ras	sGTP.1	1	Ras-like	1	Ras Superfamily
1YZL_A	Ras	sGTP.1	1	Ras-like	1	Ras Superfamily
1WMS_A	Ras	sGTP.1	1	Ras-like	1	Ras Superfamily
1S8F_A	Ras	sGTP.1	1	Ras-like	1	Ras Superfamily
3CLV_A	Roc+Ras	sGTP.1	1	Ras-like	1	Ras Superfamily
3CPJ_B	Ras	sGTP.1	1	Ras-like	1	Ras Superfamily
3RWM_B	Ras	sGTP.1	1	Ras-like	1	Ras Superfamily
2GIL_A	Ras	sGTP.1	1	Ras-like	1	Ras Superfamily
6EKK_C	Ras	sGTP.1	1	Ras-like	1	Ras Superfamily
3TW8_B	Ras	sGTP.1	1	Ras-like	1	Ras Superfamily
4U5X_A	Ras	sGTP.1	1	Ras-like	1	Ras Superfamily
4DID_A	Ras	sGTP.1	1	Ras-like	1	Ras Superfamily
5UPL_B	Ras	sGTP.1	1	Ras-like	1	Ras Superfamily
1CF4_A	Ras	sGTP.1	1	Ras-like	1	Ras Superfamily
3GCG_A	Ras	sGTP.1	1	Ras-like	1	Ras Superfamily
4YC7_A	Ras	sGTP.1	1	Ras-like	1	Ras Superfamily
5ZVP_A	Ras	sGTP.1	1	Ras-like	1	Ras Superfamily
6E6H_A	Ras	sGTP.1	1	Ras-like	1	Ras Superfamily
5UHV_A	Ras	sGTP.1	1	Ras-like	1	Ras Superfamily
3CON_A	Ras	sGTP.1	1	Ras-like	1	Ras Superfamily
2EFC_B	Ras	sGTP.1	1	Ras-like	1	Ras Superfamily
2IF0_A	Ras	sGTP.1	1	Ras-like	1	Ras Superfamily
2IEY_A	Ras	sGTP.1	1	Ras-like	1	Ras Superfamily
2ZET_A	Ras	sGTP.1	1	Ras-like	1	Ras Superfamily
3QBT_A	Ras	sGTP.1	1	Ras-like	1	Ras Superfamily
4LHW_A	Ras	sGTP.1	1	Ras-like	1	Ras Superfamily
1D5C_A	Ras	sGTP.1	1	Ras-like	1	Ras Superfamily
2OCB_A	Ras	sGTP.1	1	Ras-like	1	Ras Superfamily
2HEI_A	Ras	sGTP.1	1	Ras-like	1	Ras Superfamily
1TU3_A	Ras	sGTP.1	1	Ras-like	1	Ras Superfamily
1TU4_A	Ras	sGTP.1	1	Ras-like	1	Ras Superfamily
4Q9U_B	Ras	sGTP.1	1	Ras-like	1	Ras Superfamily
1N6R_A	Ras	sGTP.1	1	Ras-like	1	Ras Superfamily
4KYI_B	Ras	sGTP.1	1	Ras-like	1	Ras Superfamily
1N6I_A	Ras	sGTP.1	1	Ras-like	1	Ras Superfamily
1N6H_A	Ras	sGTP.1	1	Ras-like	1	Ras Superfamily
1N6P_A	Ras	sGTP.1	1	Ras-like	1	Ras Superfamily
1N6N_A	Ras	sGTP.1	1	Ras-like	1	Ras Superfamily
1N6O_A	Ras	sGTP.1	1	Ras-like	1	Ras Superfamily
3MJH_A	Ras	sGTP.1	1	Ras-like	1	Ras Superfamily
1Z0D_A	Ras	sGTP.1	1	Ras-like	1	Ras Superfamily
1Z07_A	Ras	sGTP.1	1	Ras-like	1	Ras Superfamily
1HUQ_A	Ras	sGTP.1	1	Ras-like	1	Ras Superfamily
2G77_B	Ras	sGTP.1	1	Ras-like	1	Ras Superfamily

1Z06_A	Ras	sGTP.1	1	Ras-like	1	Ras Superfamily
5SZJ_A	Ras	sGTP.1	1	Ras-like	1	Ras Superfamily
2FG5_A	Ras	sGTP.1	1	Ras-like	1	Ras Superfamily
5YOZ_A	Ras	sGTP.1	1	Ras-like	1	Ras Superfamily
1Z0A_A	Ras	sGTP.1	1	Ras-like	1	Ras Superfamily
3M1I_A	Ras	sGTP.1	1	Ras-like	1	Ras Superfamily
3WYG_A	Ras	sGTP.1	1	Ras-like	1	Ras Superfamily
1EK0_A	Ras	sGTP.1	1	Ras-like	1	Ras Superfamily
1Z22_A	Ras	sGTP.1	1	Ras-like	1	Ras Superfamily
2O52_A	Ras	sGTP.1	1	Ras-like	1	Ras Superfamily
5OED_B	Ras	sGTP.1	1	Ras-like	1	Ras Superfamily
5OEC_B	Ras	sGTP.1	1	Ras-like	1	Ras Superfamily
6FF8_A	Ras	sGTP.1	1	Ras-like	1	Ras Superfamily
2WWW_A	Ras	sGTP.1	1	Ras-like	1	Ras Superfamily
4FMB_B	Ras	sGTP.1	1	Ras-like	1	Ras Superfamily
4FMC_B	Ras	sGTP.1	1	Ras-like	1	Ras Superfamily
4FMD_F	Ras	sGTP.1	1	Ras-like	1	Ras Superfamily
5YXH_A	Ras	sGTP.1	1	Ras-like	1	Ras Superfamily
1T91_A	Ras	sGTP.1	1	Ras-like	1	Ras Superfamily
3LAW_A	Ras	sGTP.1	1	Ras-like	1	Ras Superfamily
1VG1_A	Ras	sGTP.1	1	Ras-like	1	Ras Superfamily
5Z2M_A	Ras	sGTP.1	1	Ras-like	1	Ras Superfamily
5CLL_A	Ras	sGTP.1	1	Ras-like	1	Ras Superfamily
6A38_A	Ras	sGTP.1	1	Ras-like	1	Ras Superfamily
1RRP_A	Ras	sGTP.1	1	Ras-like	1	Ras Superfamily
3NC1_C	Ras	sGTP.1	1	Ras-like	1	Ras Superfamily
3ZJY_A	Ras	sGTP.1	1	Ras-like	1	Ras Superfamily
2BKU_A	Ras	sGTP.1	1	Ras-like	1	Ras Superfamily
1WA5_A	Ras	sGTP.1	1	Ras-like	1	Ras Superfamily
3NBY_C	Ras	sGTP.1	1	Ras-like	1	Ras Superfamily
5DIS_B	Ras	sGTP.1	1	Ras-like	1	Ras Superfamily
6O62_A	Ras	sGTP.1	1	Ras-like	1	Ras Superfamily
5XR4_A	Ras	sGTP.1	1	Ras-like	1	Ras Superfamily
5XR7_A	Ras	sGTP.1	1	Ras-like	1	Ras Superfamily
3TSO_A	Ras	sGTP.1	1	Ras-like	1	Ras Superfamily
4DJT_A	Ras	sGTP.1	1	Ras-like	1	Ras Superfamily
2F7S_A	Ras	sGTP.1	1	Ras-like	1	Ras Superfamily
2IEZ_A	Ras	sGTP.1	1	Ras-like	1	Ras Superfamily
4LI0_A	Ras	sGTP.1	1	Ras-like	1	Ras Superfamily
5SZI_A	Ras	sGTP.1	1	Ras-like	1	Ras Superfamily
2FU5_C	Ras	sGTP.1	1	Ras-like	1	Ras Superfamily
2ERY_A	Ras	sGTP.1	1	Ras-like	1	Ras Superfamily
2J0V_A	Ras	sGTP.1	1	Ras-like	1	Ras Superfamily
2BCG_Y	Ras	sGTP.1	1	Ras-like	1	Ras Superfamily
1YZN_A	Ras	sGTP.1	1	Ras-like	1	Ras Superfamily
2G6B_A	Ras	sGTP.1	1	Ras-like	1	Ras Superfamily
2YC2_C	Ras	sGTP.1	1	Ras-like	1	Ras Superfamily
1C1Y_A	Ras	sGTP.1	1	Ras-like	1	Ras Superfamily
4M8N_E	Ras	sGTP.1	1	Ras-like	1	Ras Superfamily
3X1Y_A	Ras	sGTP.1	1	Ras-like	1	Ras Superfamily
4DXA_A	Ras	sGTP.1	1	Ras-like	1	Ras Superfamily
6BA6_B	Ras	sGTP.1	1	Ras-like	1	Ras Superfamily
4KVG_A	Ras	sGTP.1	1	Ras-like	1	Ras Superfamily
6AXF_B	Ras	sGTP.1	1	Ras-like	1	Ras Superfamily
3BRW_D	Ras	sGTP.1	1	Ras-like	1	Ras Superfamily
3X1Z_A	Ras	sGTP.1	1	Ras-like	1	Ras Superfamily
3X1W_A	Ras	sGTP.1	1	Ras-like	1	Ras Superfamily
1GUA_A	Ras	sGTP.1	1	Ras-like	1	Ras Superfamily
5VCU_A	Ras	sGTP.1	1	Ras-like	1	Ras Superfamily
3OES_A	Ras	sGTP.1	1	Ras-like	1	Ras Superfamily
3BWD_D	Ras	sGTP.1	1	Ras-like	1	Ras Superfamily
2NTY_C	Ras	sGTP.1	1	Ras-like	1	Ras Superfamily
3KKO_A	Ras	sGTP.1	1	Ras-like	1	Ras Superfamily

3KKP_A	Ras	sGTP.1	1	Ras-like	1	Ras Superfamily
3PIR_A	Ras	sGTP.1	1	Ras-like	1	Ras Superfamily
1X1R_A	Ras	sGTP.1	1	Ras-like	1	Ras Superfamily
1Z0J_A	Ras	sGTP.1	1	Ras-like	1	Ras Superfamily
1YVD_A	Ras	sGTP.1	1	Ras-like	1	Ras Superfamily
2A78_A	Ras	sGTP.1	1	Ras-like	1	Ras Superfamily
1ZC3_A	Ras	sGTP.1	1	Ras-like	1	Ras Superfamily
1UAD_A	Ras	sGTP.1	1	Ras-like	1	Ras Superfamily
2KWI_A	Ras	sGTP.1	1	Ras-like	1	Ras Superfamily
2KE5_A	Ras	sGTP.1	1	Ras-like	1	Ras Superfamily
5CM8_B	Ras	sGTP.1	1	Ras-like	1	Ras Superfamily
3B13_B	Ras	sGTP.1	1	Ras-like	1	Ras Superfamily
3TH5_A	Ras	sGTP.1	1	Ras-like	1	Ras Superfamily
4GZM_A	Ras	sGTP.1	1	Ras-like	1	Ras Superfamily
4GZL_A	Ras	sGTP.1	1	Ras-like	1	Ras Superfamily
3SBD_A	Ras	sGTP.1	1	Ras-like	1	Ras Superfamily
2YIN_C	Ras	sGTP.1	1	Ras-like	1	Ras Superfamily
2W2T_A	Ras	sGTP.1	1	Ras-like	1	Ras Superfamily
2W2V_A	Ras	sGTP.1	1	Ras-like	1	Ras Superfamily
2IC5_A	Ras	sGTP.1	1	Ras-like	1	Ras Superfamily
2G0N_A	Ras	sGTP.1	1	Ras-like	1	Ras Superfamily
2BMD_A	Ras	sGTP.1	1	Ras-like	1	Ras Superfamily
1YU9_A	Ras	sGTP.1	1	Ras-like	1	Ras Superfamily
1Z0K_A	Ras	sGTP.1	1	Ras-like	1	Ras Superfamily
5WDS_A	Ras	sGTP.1	1	Ras-like	1	Ras Superfamily
2ATX_A	Ras	sGTP.1	1	Ras-like	1	Ras Superfamily
1YZT_A	Ras	sGTP.1	1	Ras-like	1	Ras Superfamily
1YZU_A	Ras	sGTP.1	1	Ras-like	1	Ras Superfamily
1Z08_A	Ras	sGTP.1	1	Ras-like	1	Ras Superfamily
1NF3_A	Ras	sGTP.1	1	Ras-like	1	Ras Superfamily
2DFK_B	Ras	sGTP.1	1	Ras-like	1	Ras Superfamily
5F11_B	Ras	sGTP.1	1	Ras-like	1	Ras Superfamily
2ODB_A	Ras	sGTP.1	1	Ras-like	1	Ras Superfamily
1A4R_B	Ras	sGTP.1	1	Ras-like	1	Ras Superfamily
2WM9_B	Ras	sGTP.1	1	Ras-like	1	Ras Superfamily
1AN0_A	Ras	sGTP.1	1	Ras-like	1	Ras Superfamily
2QRZ_A	Ras	sGTP.1	1	Ras-like	1	Ras Superfamily
1KZ7_B	Ras	sGTP.1	1	Ras-like	1	Ras Superfamily
3BC1_A	Ras	sGTP.1	1	Ras-like	1	Ras Superfamily
6HUF_A	Ras	sGTP.1	1	Ras-like	1	Ras Superfamily
1Z2C_A	Ras	sGTP.1	1	Ras-like	1	Ras Superfamily
5HPY_B	Ras	sGTP.1	1	Ras-like	1	Ras Superfamily
1KMQ_A	Ras	sGTP.1	1	Ras-like	1	Ras Superfamily
4XH9_B	Ras	sGTP.1	1	Ras-like	1	Ras Superfamily
3MSX_A	Ras	sGTP.1	1	Ras-like	1	Ras Superfamily
4XOI_A	Ras	sGTP.1	1	Ras-like	1	Ras Superfamily
4XSG_A	Ras	sGTP.1	1	Ras-like	1	Ras Superfamily
1XCG_B	Ras	sGTP.1	1	Ras-like	1	Ras Superfamily
1TX4_B	Ras	sGTP.1	1	Ras-like	1	Ras Superfamily
4D0N_A	Ras	sGTP.1	1	Ras-like	1	Ras Superfamily
3LW8_A	Ras	sGTP.1	1	Ras-like	1	Ras Superfamily
1S1C_A	Ras	sGTP.1	1	Ras-like	1	Ras Superfamily
1CXZ_A	Ras	sGTP.1	1	Ras-like	1	Ras Superfamily
2MSC_B	Ras	sGTP.1	1	Ras-like	1	Ras Superfamily
6QUU_A	Ras	sGTP.1	1	Ras-like	1	Ras Superfamily
5VBM_A	Ras	sGTP.1	1	Ras-like	1	Ras Superfamily
5VQ1_A	Ras	sGTP.1	1	Ras-like	1	Ras Superfamily
4EPT_A	Ras	sGTP.1	1	Ras-like	1	Ras Superfamily
4PZZ_A	Ras	sGTP.1	1	Ras-like	1	Ras Superfamily
4PZY_A	Ras	sGTP.1	1	Ras-like	1	Ras Superfamily
5VP7_A	Ras	sGTP.1	1	Ras-like	1	Ras Superfamily
4EPX_A	Ras	sGTP.1	1	Ras-like	1	Ras Superfamily
5VQ2_A	Ras	sGTP.1	1	Ras-like	1	Ras Superfamily

4EPR_A	Ras	sGTP.1	1	Ras-like	1	Ras Superfamily
4Q01_A	Ras	sGTP.1	1	Ras-like	1	Ras Superfamily
2RGB_A	Ras	sGTP.1	1	Ras-like	1	Ras Superfamily
2OCY_C	Ras	sGTP.1	1	Ras-like	1	Ras Superfamily
4Z8Y_A	Ras	sGTP.1	1	Ras-like	1	Ras Superfamily
1G17_A	Ras	sGTP.1	1	Ras-like	1	Ras Superfamily
1G16_A	Ras	sGTP.1	1	Ras-like	1	Ras Superfamily
6IYB_A	Ras	sGTP.1	1	Ras-like	1	Ras Superfamily
1YZK_A	Ras	sGTP.1	1	Ras-like	1	Ras Superfamily
2HV8_A	Ras	sGTP.1	1	Ras-like	1	Ras Superfamily
5EZ5_A	Ras	sGTP.1	1	Ras-like	1	Ras Superfamily
2D7C_A	Ras	sGTP.1	1	Ras-like	1	Ras Superfamily
4RKF_A	Ras	sGTP.1	1	Ras-like	1	Ras Superfamily
3DZ8_A	Ras	sGTP.1	1	Ras-like	1	Ras Superfamily
3RAB_A	Ras	sGTP.1	1	Ras-like	1	Ras Superfamily
2GF9_A	Ras	sGTP.1	1	Ras-like	1	Ras Superfamily
2C2H_A	Ras	sGTP.1	1	Ras-like	1	Ras Superfamily
1X3S_A	Ras	sGTP.1	1	Ras-like	1	Ras Superfamily
2ERX_A	Ras	sGTP.1	1	Ras-like	1	Ras Superfamily
2GF0_A	Ras	sGTP.1	1	Ras-like	1	Ras Superfamily
1DPF_A	Ras	sGTP.1	1	Ras-like	1	Ras Superfamily
4MIT_A	Ras	sGTP.1	1	Ras-like	1	Ras Superfamily
5FI0_B	Ras	sGTP.1	1	Ras-like	1	Ras Superfamily
1E96_A	Ras	sGTP.1	1	Ras-like	1	Ras Superfamily
1HH4_A	Ras	sGTP.1	1	Ras-like	1	Ras Superfamily
114T_D	Ras	sGTP.1	1	Ras-like	1	Ras Superfamily
2H7V_A	Ras	sGTP.1	1	Ras-like	1	Ras Superfamily
1MH1_A	Ras	sGTP.1	1	Ras-like	1	Ras Superfamily
2VRW_A	Ras	sGTP.1	1	Ras-like	1	Ras Superfamily
1G4U_R	Ras	sGTP.1	1	Ras-like	1	Ras Superfamily
3SUA_A	Ras	sGTP.1	1	Ras-like	1	Ras Superfamily
6BC1_A	Ras	sGTP.1	1	Ras-like	1	Ras Superfamily
6AGP_A	Ras	sGTP.1	1	Ras-like	1	Ras Superfamily
3RYT_C	Ras	sGTP.1	1	Ras-like	1	Ras Superfamily
4YON_B	Ras	sGTP.1	1	Ras-like	1	Ras Superfamily
2FJU_A	Ras	sGTP.1	1	Ras-like	1	Ras Superfamily
1FOE_B	Ras	sGTP.1	1	Ras-like	1	Ras Superfamily
1HE1_C	Ras	sGTP.1	1	Ras-like	1	Ras Superfamily
5N6O_A	Ras	sGTP.1	1	Ras-like	1	Ras Superfamily
3CPH_A	Ras	sGTP.1	1	Ras-like	1	Ras Superfamily
2EQB_A	Ras	sGTP.1	1	Ras-like	1	Ras Superfamily
3JZA_A	Ras	sGTP.1	1	Ras-like	1	Ras Superfamily
2RHD_A	Ras	sGTP.1	1	Ras-like	1	Ras Superfamily
3L0I_B	Ras	sGTP.1	1	Ras-like	1	Ras Superfamily
3TKL_A	Ras	sGTP.1	1	Ras-like	1	Ras Superfamily
2FOL_A	Ras	sGTP.1	1	Ras-like	1	Ras Superfamily
4IRU_B	Ras	sGTP.1	1	Ras-like	1	Ras Superfamily
3SFV_A	Ras	sGTP.1	1	Ras-like	1	Ras Superfamily
411O_A	Ras	sGTP.1	1	Ras-like	1	Ras Superfamily
4JVS_B	Ras	sGTP.1	1	Ras-like	1	Ras Superfamily
5O74_B	Ras	sGTP.1	1	Ras-like	1	Ras Superfamily
2FN4_A	Ras	sGTP.1	1	Ras-like	1	Ras Superfamily
3BFK_A	Ras	sGTP.1	1	Ras-like	1	Ras Superfamily
3A58_B	Ras	sGTP.1	1	Ras-like	1	Ras Superfamily
5LPN_A	Ras	sGTP.1	1	Ras-like	1	Ras Superfamily
4DKX_A	Ras	sGTP.1	1	Ras-like	1	Ras Superfamily
3BBP_A	Ras	sGTP.1	1	Ras-like	1	Ras Superfamily
5LEF_A	Ras	sGTP.1	1	Ras-like	1	Ras Superfamily
3CWZ_A	Ras	sGTP.1	1	Ras-like	1	Ras Superfamily
2FE4_A	Ras	sGTP.1	1	Ras-like	1	Ras Superfamily
1YZQ_A	Ras	sGTP.1	1	Ras-like	1	Ras Superfamily
6IF2_B	Ras	sGTP.1	1	Ras-like	1	Ras Superfamily
5SZH_B	Ras	sGTP.1	1	Ras-like	1	Ras Superfamily

5HZK_A	Ras	sGTP.1	1	Ras-like	1	Ras Superfamily
5C2J_B	Ras	sGTP.1	1	Ras-like	1	Ras Superfamily
6AJ4_B	Ras	sGTP.1	1	Ras-like	1	Ras Superfamily
3VHL_B	Ras	sGTP.1	1	Ras-like	1	Ras Superfamily
1AJE_A	Ras	sGTP.1	1	Ras-like	1	Ras Superfamily
1DOA_A	Ras	sGTP.1	1	Ras-like	1	Ras Superfamily
1GZS_A	Ras	sGTP.1	1	Ras-like	1	Ras Superfamily
1CEE_A	Ras	sGTP.1	1	Ras-like	1	Ras Superfamily
5CJP_A	Ras	sGTP.1	1	Ras-like	1	Ras Superfamily
5UPK_C	Ras	sGTP.1	1	Ras-like	1	Ras Superfamily
1EES_A	Ras	sGTP.1	1	Ras-like	1	Ras Superfamily
3EG5_A	Ras	sGTP.1	1	Ras-like	1	Ras Superfamily
2KB0_A	Ras	sGTP.1	1	Ras-like	1	Ras Superfamily
3QBV_A	Ras	sGTP.1	1	Ras-like	1	Ras Superfamily
4JS0_A	Ras	sGTP.1	1	Ras-like	1	Ras Superfamily
2ASE_A	Ras	sGTP.1	1	Ras-like	1	Ras Superfamily
1AM4_D	Ras	sGTP.1	1	Ras-like	1	Ras Superfamily
2L0X_A	Ras	sGTP.1	1	Ras-like	1	Ras Superfamily
4O25_A	Ras	sGTP.1	1	Ras-like	1	Ras Superfamily
4O2R_A	Ras	sGTP.1	1	Ras-like	1	Ras Superfamily
4O2L_A	Ras	sGTP.1	1	Ras-like	1	Ras Superfamily
2WBL_C	Ras	sGTP.1	1	Ras-like	1	Ras Superfamily
3GJ4_A	Ras	sGTP.1	1	Ras-like	NA	NA
5KSO_A	Ras	sGTP_DYNAMIN.1	1	Ras-like	NA	NA
5KSY_A	Ras	sGTP_DYNAMIN.1	1	Ras-like	NA	NA
4URU_R	Ras	sGTP.1	1	Ras-like	NA	NA
4DVG_A	Ras	sGTP.1	1	Ras-like	NA	NA
3REF_B	Ras	sGTP.1	1	Ras-like	NA	NA
5FBR_B	Ras	sGTP.1	1	Ras-like	NA	NA
5FBL_B	Ras	sGTP.1	1	Ras-like	NA	NA
4D0L_B	Ras	sGTP.1	1	Ras-like	NA	NA
5LDD_C	Ras	sGTP.1	1	Ras-like	NA	NA
1M7B_A	Ras	sGTP.1	1	Ras-like	NA	NA
4FMC_F	Ras	sGTP.1	1	Ras-like	NA	NA
4CZ2_A	Ras	sGTP.1	1	Ras-like	NA	NA
4CYM_A	Ras	sGTP.1	1	Ras-like	NA	NA
3T5G_A	Ras	sGTP.1	1	Ras-like	NA	NA
6BSX_A	Ras	sGTP.1	1	Ras-like	NA	NA
1XTQ_A	Ras	sGTP.1	1	Ras-like	NA	NA
2EW1_A	Ras	sGTP.1	1	Ras-like	NA	NA
3GJ0_A	Ras	sGTP.1	1	Ras-like	NA	NA
3ICQ_B	Ras	sGTP.1	1	Ras-like	NA	NA
2X19_A	Ras	sGTP.1	1	Ras-like	NA	NA
2OIL_A	Ras	sGTP.1	1	Ras-like	NA	NA
2P5S_A	Ras	sGTP.1	1	Ras-like	NA	NA
4OJK_A	Ras	sGTP.1	1	Ras-like	NA	NA
2F9L_A	Ras	sGTP.1	1	Ras-like	NA	NA
3C5C_A	Ras	sGTP.1	1	Ras-like	NA	NA
5JCP_B	RhoGAP+Ras	sGTP.1	1	Ras-like	NA	NA
4KLZ_A	Ras	sGTP.1	1	Ras-like	NA	NA
2N1B_A	Ras	sGTP.1	1	Ras-like	NA	NA
5DH9_A	Ras	sGTP.1	1	Ras-like	NA	NA
2IWR_A	Ras	sGTP.1	1	Ras-like	NA	NA
2BMJ_A	Ras	sGTP.1	1	Ras-like	NA	NA
5C1S_A	Ras+Ras	sGTP.1+sGTP.1	1	Ras-like	NA	NA
5C1T_A	Ras+Ras	sGTP.1+sGTP.1	1	Ras-like	NA	NA
3Q85_A	Ras	sGTP.1	1	Ras-like	NA	NA
3CBQ_A	Ras	sGTP.1	1	Ras-like	NA	NA
4AII_A	Ras	sGTP.1	1	Ras-like	NA	NA
2NZJ_A	Ras	sGTP.1	1	Ras-like	NA	NA
1U8Y_A	Ras	sGTP.1	1	Ras-like	NA	NA
1RYF_A	Ras+Ras	sGTP.1	1	Ras-like	NA	NA
5X9S_A	Ras	sGTP.1	1	Ras-like	NA	NA

6IAN_E	Roc	sGTP.1	1	Ras-like	NA	NA
6IA7_A	Roc	sGTP.1	1	Ras-like	NA	NA
2J1L_A	Ras	ROK-HTH.1+sGTP.1	1	Ras-like	NA	NA
2Q3H_A	Ras	sGTP.1	1	Ras-like	NA	NA
2ATV_A	Ras	ROK-HTH.1+sGTP.1	1	Ras-like	NA	NA
5A0F_A	Ras	sGTP.1	1	Ras-like	NA	NA
2FV8_A	Ras	ROK-HTH.1+sGTP.1	1	Ras-like	NA	NA
2GCO_A	Ras	sGTP.1	1	Ras-like	NA	NA
2GCN_A	Ras	sGTP.1	1	Ras-like	NA	NA
5IRC_F	Ras	sGTP.1	1	Ras-like	NA	NA
6BC0_F	Ras	sGTP.1	1	Ras-like	NA	NA
3KZ1_E	Ras	sGTP.1	1	Ras-like	NA	NA
2GJS_A	Ras	sGTP.1	1	Ras-like	NA	NA
2DPX_A	Ras	sGTP.1	1	Ras-like	NA	NA
3Q72_A	Ras	sGTP.1	1	Ras-like	NA	NA
1JAH_A	Ras	sGTP.1	1	Ras-like	NA	NA
4DSO_A	Ras	sGTP.1	1	Ras-like	NA	NA
3GFT_A	Ras	sGTP.1	1	Ras-like	NA	NA
5OCO_B	Ras	sGTP.1	1	Ras-like	NA	NA
5US4_A	Ras	sGTP.1	1	Ras-like	NA	NA
6CCH_B	Ras	sGTP.1	1	Ras-like	NA	NA
5TAR_A	Ras	sGTP.1	1	Ras-like	NA	NA
5VBE_A	Ras	sGTP.1	1	Ras-like	NA	NA
1LFD_B	Ras	sGTP.1	1	Ras-like	NA	NA
2CLD_X	Ras	sGTP.1	1	Ras-like	NA	NA
2CL0_X	Ras	sGTP.1	1	Ras-like	NA	NA
3LO5_A	Ras	sGTP.1	1	Ras-like	NA	NA
4XVQ_A	Ras	sGTP.1	1	Ras-like	NA	NA
5C2K_A	Ras+RhoGAP	sGTP.1	1	Ras-like	NA	NA
2HXS_A	Ras	sGTP.1	1	Ras-like	NA	NA
4UJ4_A	Ras	sGTP.1	1	Ras-like	NA	NA
4UJ5_A	Ras	sGTP.1	1	Ras-like	NA	NA
4UJ3_A	Ras	sGTP.1	1	Ras-like	NA	NA
1OIV_A	Ras	sGTP.1	1	Ras-like	NA	NA
1OIW_A	Ras	sGTP.1	1	Ras-like	NA	NA
5JCZ_A	Ras	sGTP.1	1	Ras-like	NA	NA
4LWZ_A	Ras	sGTP.1	1	Ras-like	NA	NA
6IXV_E	Ras	sGTP.1	1	Ras-like	NA	NA
2GZD_A	Ras	sGTP.1	1	Ras-like	NA	NA
4C4P_A	Ras	sGTP.1	1	Ras-like	NA	NA
1ZBD_A	Ras	sGTP.1	1	Ras-like	NA	NA
2IL1_A	Ras	sGTP.1	1	Ras-like	NA	NA
3Q3J_B	Ras	sGTP.1	1	Ras-like	NA	NA
4DRZ_A	Ras	sGTP.1	1	Ras-like	NA	NA
1Z0F_A	Ras	sGTP.1	1	Ras-like	NA	NA
4D0G_A	Ras	sGTP.1	1	Ras-like	NA	NA
5HZH_A	Ras+PAS_9+Ras	sGTP.1+PAS.3+sGTP.1	1	Ras-like	NA	NA
2WKQ_A	PAS_9+Ras	PAS.3+sGTP.1	1	Ras-like	NA	NA
2WKP_A	PAS_9+Ras	PAS.3+sGTP.1	1	Ras-like	NA	NA
2WKR_A	PAS_9+Ras	PAS.3+sGTP.1	1	Ras-like	NA	NA
1GWN_A	Ras	sGTP.1	1	Ras-like	NA	NA
2V55_B	Ras	sGTP.1	1	Ras-like	NA	NA
2HUP_A	Ras	sGTP.1	1	Ras-like	NA	NA
2CLS_A	Ras	sGTP.1	1	Ras-like	NA	NA
2REX_B	Ras	sGTP.1	1	Ras-like	NA	NA
2A5J_A	Ras	sGTP.1	1	Ras-like	NA	NA
4RKE_A	Ras	sGTP.1	1	Ras-like	NA	NA
4LRW_A	Ras	sGTP.1	1	Ras-like	NA	NA
4KU4_A	Ras	sGTP.1	1	Ras-like	NA	NA
3IHW_A	Ras	sGTP.1	1	Ras-like	NA	NA
5UQW_A	Ras	sGTP.1	1	Ras-like	NA	NA
6FA1_B	Ras	sGTP.1	1	Ras-like	NA	NA
6F76_B	Ras	sGTP.1	1	Ras-like	NA	NA

6GOG_A	Ras	sGTP.1	1	Ras-like	NA	NA
2C5L_A	Ras	sGTP.1	1	Ras-like	NA	NA
6FA3_A	Ras	sGTP.1	1	Ras-like	NA	NA
3KKM_A	Ras	sGTP.1	1	Ras-like	NA	NA
6FA1_A	Ras	sGTP.1	1	Ras-like	NA	NA
6GOF_A	Ras	sGTP.1	1	Ras-like	NA	NA
6GOE_A	Ras	sGTP.1	1	Ras-like	NA	NA
6GOD_A	Ras	sGTP.1	1	Ras-like	NA	NA
6Q21_A	Ras	sGTP.1	1	Ras-like	NA	NA
4L9S_A	Ras	sGTP.1	1	Ras-like	NA	NA
4K81_B	Ras	sGTP.1	1	Ras-like	NA	NA
6GQT_A	Ras	sGTP.1	1	Ras-like	NA	NA
4EFL_A	Ras	sGTP.1	1	Ras-like	NA	NA
5XCO_A	Ras	sGTP.1	1	Ras-like	NA	NA
4EFN_A	Ras	sGTP.1	1	Ras-like	NA	NA
6GQW_E	Ras	sGTP.1	1	Ras-like	NA	NA
4EFM_A	Ras	sGTP.1	1	Ras-like	NA	NA
2Q21_A	Ras	sGTP.1	1	Ras-like	NA	NA
6FA2_C	Ras	sGTP.1	1	Ras-like	NA	NA
6M9W_A	Ras	sGTP.1	1	Ras-like	NA	NA
6GJ6_A	Ras	sGTP.1	1	Ras-like	NA	NA
5WHD_A	Ras	sGTP.1	1	Ras-like	NA	NA
5WDO_A	Ras	sGTP.1	1	Ras-like	NA	NA
6AMB_A	Ras	sGTP.1	1	Ras-like	NA	NA
6EPL_R	Ras	sGTP.1	1	Ras-like	NA	NA
4QL3_A	Ras	sGTP.1	1	Ras-like	NA	NA
4L8G_A	Ras	sGTP.1	1	Ras-like	NA	NA
5WDQ_A	Ras	sGTP.1	1	Ras-like	NA	NA
6FA2_B	Ras	sGTP.1	1	Ras-like	NA	NA
6AXG_B	Ras	sGTP.1	1	Ras-like	NA	NA
6ASE_A	Ras	sGTP.1	1	Ras-like	NA	NA
5WHA_A	Ras	sGTP.1	1	Ras-like	NA	NA
4WA7_A	Ras	sGTP.1	1	Ras-like	NA	NA
4LPK_A	Ras	sGTP.1	1	Ras-like	NA	NA
4TQ9_A	Ras	sGTP.1	1	Ras-like	NA	NA
6MQN_A	Ras	sGTP.1	1	Ras-like	NA	NA
6BP1_A	Ras	sGTP.1	1	Ras-like	NA	NA
5O2S_A	Ras	sGTP.1	1	Ras-like	NA	NA
5KYK_A	Ras	sGTP.1	1	Ras-like	NA	NA
6GJ8_A	Ras	sGTP.1	1	Ras-like	NA	NA
4TQA_A	Ras	sGTP.1	1	Ras-like	NA	NA
6P0Z_A	Ras	sGTP.1	1	Ras-like	NA	NA
6ASA_A	Ras	sGTP.1	1	Ras-like	NA	NA
5E95_A	Ras	sGTP.1	1	Ras-like	NA	NA
6BOF_A	Ras	sGTP.1	1	Ras-like	NA	NA
6FA2_A	Ras	sGTP.1	1	Ras-like	NA	NA
4NYI_R	Ras	sGTP.1	1	Ras-like	NA	NA
6D55_A	Ras	sGTP.1	1	Ras-like	NA	NA
1XCM_A	Ras	sGTP.1	1	Ras-like	NA	NA
6BVJ_A	Ras	sGTP.1	1	Ras-like	NA	NA
4NYI_Q	Ras	sGTP.1	1	Ras-like	NA	NA
6MBQ_A	Ras	sGTP.1	1	Ras-like	NA	NA
6MQG_A	Ras	sGTP.1	1	Ras-like	NA	NA
2RGC_A	Ras	sGTP.1	1	Ras-like	NA	NA
1WQ1_R	Ras	sGTP.1	1	Ras-like	NA	NA
521P_A	Ras	sGTP.1	1	Ras-like	NA	NA
2RGA_A	Ras	sGTP.1	1	Ras-like	NA	NA
1IAQ_A	Ras	sGTP.1	1	Ras-like	NA	NA
1ZVQ_A	Ras	sGTP.1	1	Ras-like	NA	NA
5WPL_A	Ras	sGTP.1	1	Ras-like	NA	NA
221P_A	Ras	sGTP.1	1	Ras-like	NA	NA
421P_A	Ras	sGTP.1	1	Ras-like	NA	NA
3DDC_A	Ras	sGTP.1	1	Ras-like	NA	NA

3I3S_R	Ras	sGTP.1	1	Ras-like	NA	NA
5WLB_A	Ras	sGTP.1	1	Ras-like	NA	NA
4XVR_A	Ras	sGTP.1	1	Ras-like	NA	NA
3K9N_A	Ras	sGTP.1	1	Ras-like	NA	NA
5UFE_A	Ras	sGTP.1	1	Ras-like	NA	NA
5WDP_A	Ras	sGTP.1	1	Ras-like	NA	NA
1CLU_A	Ras	sGTP.1	1	Ras-like	NA	NA
6DZH_A	Ras	sGTP.1	1	Ras-like	NA	NA
6E6F_A	Ras	sGTP.1	1	Ras-like	NA	NA
1ZW6_A	Ras	sGTP.1	1	Ras-like	NA	NA
5UFQ_A	Ras	sGTP.1	1	Ras-like	NA	NA
2X1V_A	Ras	sGTP.1	1	Ras-like	NA	NA
721P_A	Ras	sGTP.1	1	Ras-like	NA	NA
621P_A	Ras	sGTP.1	1	Ras-like	NA	NA
1XJ0_A	Ras	sGTP.1	1	Ras-like	NA	NA
1LF0_A	Ras	sGTP.1	1	Ras-like	NA	NA
1RVD_A	Ras	sGTP.1	1	Ras-like	NA	NA
1NVV_Q	Ras	sGTP.1	1	Ras-like	NA	NA
1AGP_A	Ras	sGTP.1	1	Ras-like	NA	NA
2QUZ_A	Ras	sGTP.1	1	Ras-like	NA	NA
3V4F_A	Ras	sGTP.1	1	Ras-like	NA	NA
2Y8E_A	Ras	sGTP.1	1	Ras-like	NA	NA
2HT6_A	Ras	sGTP.1	1	Ras-like	NA	NA
2G3Y_A	Ras	ROK-HTH.1+sGTP.1	1	Ras-like	NA	NA
2CJW_B	Ras	sGTP.1	1	Ras-like	NA	NA
2CJW_A	Ras	sGTP.1	1	Ras-like	NA	NA
5ZHX_EE	Ras	sGTP.1	1	Ras-like	NA	NA
1X86_B	Ras	sGTP.1	1	Ras-like	NA	NA
5FR1_A	Ras	sGTP.1	1	Ras-like	NA	NA
4F38_A	Ras	sGTP.1	1	Ras-like	NA	NA
5FR2_A	Ras	sGTP.1	1	Ras-like	NA	NA
1OW3_B	Ras	sGTP.1	1	Ras-like	NA	NA
5M6X_B	Ras	sGTP.1	1	Ras-like	NA	NA
1LB1_B	Ras	sGTP.1	1	Ras-like	NA	NA
1CC0_A	Ras	sGTP.1	1	Ras-like	NA	NA
5SZK_B	Ras	sGTP.1	1	Ras-like	NA	NA
3SEA_A	Ras	sGTP.1	1	Ras-like	NA	NA
5OWV_C	Dynamin_N	sGTP_DYNAMIN.1	2	S_Dynamin-like	3	Dynamin1
5OWV_A	T2SSE+Dynamin_N	sGTP_DYNAMIN.1	2	S_Dynamin-like	3	Dynamin1
2J68_A	FtsK_SpoIIIE+Dynamin_N	sGTP_DYNAMIN.1+Imm49.1	2	S_Dynamin-like	3	Dynamin1
4CID_A	EHD_N+Dynamin_N+DUF5600+EF-hand_4	sGTP_DYNAMIN.1	2	S_Dynamin-like	27	Dyn-EHD
2QPT_A	EHD_N+Dynamin_N+DUF5600+EF-hand_4	sGTP_DYNAMIN.1	2	S_Dynamin-like	27	Dyn-EHD
5MTV_A	EHD_N+Dynamin_N+DUF5600+EF-hand_4	sGTP_DYNAMIN.1	2	S_Dynamin-like	27	Dyn-EHD
4AUR_A	MMR_HSR1+DLP_helical	sGTP_DYNAMIN.1	2	S_Dynamin-like	36	BDLP_36_LeoA
5GO4_A	Dynamin_N+Fzo_mitofusin	sGTP_DYNAMIN.1	2	S_Dynamin-like	NA	NA
5GOE_A	Dynamin_N+Fzo_mitofusin	sGTP_DYNAMIN.1	2	S_Dynamin-like	NA	NA
5GNR_A	Dynamin_N+Fzo_mitofusin	sGTP_DYNAMIN.1	2	S_Dynamin-like	NA	NA
5GNT_A	Dynamin_N+Fzo_mitofusin	sGTP_DYNAMIN.1	2	S_Dynamin-like	NA	NA
5YEW_A	Dynamin_N+Fzo_mitofusin	sGTP_DYNAMIN.1	2	S_Dynamin-like	NA	NA
3ZJC_A	AIG1	GTPase-AIG.1	3	AIG1-like	2	GIMAP
3V70_A	AIG1	GTPase-AIG.1	3	AIG1-like	2	GIMAP
3LXX_A	AIG1	GTPase-AIG.1	3	AIG1-like	2	GIMAP
2XTP_A	AIG1	GTPase-AIG.1	3	AIG1-like	2	GIMAP
2XTO_A	AIG1	GTPase-AIG.1	3	AIG1-like	2	GIMAP
2XTM_A	AIG1	GTPase-AIG.1	3	AIG1-like	2	GIMAP
3P1J_A	AIG1	GTPase-AIG.1	3	AIG1-like	2	GIMAP
3DEF_A	AIG1	GTPase-AIG.1	3	AIG1-like	37	Toc
3BB3_A	AIG1	GTPase-AIG.1	3	AIG1-like	37	Toc
3BB4_A	AIG1	GTPase-AIG.1	3	AIG1-like	37	Toc
2J3E_A	AIG1	GTPase-AIG.1	3	AIG1-like	37	Toc
3BB1_A	AIG1	GTPase-AIG.1	3	AIG1-like	NA	NA
1H65_A	AIG1	GTPase-AIG.1	3	AIG1-like	NA	NA
3IEV_A	MMR_HSR1+MMR_HSR1_Xtn+KH_2	sGTP.1	4	ERA-like	5	ERA

3R9W_A	MMR_HSR1+MMR_HSR1_Xtn+KH_2	sGTP.1	4	ERA-like	5	ERA
1XZP_A	TrmE_N+MnmE_helical	sGTP.1	4	ERA-like	5	ERA
5DN8_A	MMR_HSR1+MMR_HSR1+KH_dom-like	sGTP.1+sGTP.1	4	ERA-like	5	ERA
4DCS_A	MMR_HSR1+MMR_HSR1+KH_dom-like	sGTP.1+sGTP.1	4	ERA-like	5	ERA
2HJG_A	MMR_HSR1+MMR_HSR1+KH_dom-like	sGTP.1+sGTP.1	4	ERA-like	5	ERA
1MKY_A	MMR_HSR1+MMR_HSR1+KH_dom-like	sGTP.1+sGTP.1	4	ERA-like	5	ERA
4KYU_A	MMR_HSR1	sGTP.1	4	ERA-like	5	ERA
5X4B_A	MMR_HSR1	sGTP.1	4	ERA-like	5	ERA
1X18_X	MMR_HSR1+KH_2	sGTP.1	4	ERA-like	5	ERA
3GEE_A	TrmE_N+MnmE_helical	VirB11.1+sGTP.1	4	ERA-like	5	ERA
1WXQ_A	MMR_HSR1+MMR_HSR1_C+TGS	sGTP.1+TGS.2	4	ERA-like	6	Obg_CgtA
5EE0_A	MMR_HSR1+YchF-GTPase_C	sGTP.1	4	ERA-like	6	Obg_CgtA
1LNZ_A	GTP1_OBG+MMR_HSR1	sGTP.1	4	ERA-like	6	Obg_CgtA
1NI3_A	MMR_HSR1+YchF-GTPase_C	sGTP.1	4	ERA-like	6	Obg_CgtA
5M04_A	GTP1_OBG+MMR_HSR1	sGTP.1	4	ERA-like	6	Obg_CgtA
4A9A_A	MMR_HSR1+MMR_HSR1_Xtn+TGS	sGTP.1+TGS.2	4	ERA-like	6	Obg_CgtA
2DBY_A	MMR_HSR1+YchF-GTPase_C	sGTP.1	4	ERA-like	6	Obg_CgtA
6G0Z_B	MMR_HSR1	sGTP.1+GTPase-AIG.1	4	ERA-like	7	DAR_YlqF
4DHE_A	MMR_HSR1	sGTP.1	4	ERA-like	10	EngB/YxjS
3KXK_A	GTP-bdg_N+GTP-bdg_M+MMR_HSR1+MMR_HSR1_Xtn	SbcC.1+sGTP.1	4	ERA-like	11	HflX
2QTF_A	GTP-bdg_N+GTP-bdg_M+MMR_HSR1+MMR_HSR1_Xtn	SbcC.1+sGTP.1	4	ERA-like	11	HflX
3KXL_A	GTP-bdg_N+GTP-bdg_M+MMR_HSR1+MMR_HSR1_Xtn	SbcC.1+sGTP.1	4	ERA-like	11	HflX
3K53_A	FeoB_N+FeoB_Cyto	sGTP.1	4	ERA-like	15	FeoB
3A1W_A	FeoB_N	sGTP.1	4	ERA-like	15	FeoB
5FH9_A	FeoB_N+FeoB_Cyto	sGTP.1	4	ERA-like	15	FeoB
3I8S_A	FeoB_N+FeoB_Cyto	sGTP.1	4	ERA-like	15	FeoB
4R98_A	FeoB_N+FeoB_Cyto	sGTP.1	4	ERA-like	15	FeoB
3HYT_A	FeoB_N+FeoB_Cyto	sGTP.1	4	ERA-like	15	FeoB
4Q5I_A	FeoB_N+FeoB_Cyto	sGTP.1	4	ERA-like	15	FeoB
2WIA_A	FeoB_N+FeoB_Cyto	sGTP.1	4	ERA-like	15	FeoB
4Q00_A	FeoB_N+FeoB_Cyto	sGTP.1	4	ERA-like	15	FeoB
3HYR_A	FeoB_N+FeoB_Cyto	sGTP.1	4	ERA-like	15	FeoB
3IBY_A	FeoB_N	sGTP.1	4	ERA-like	15	FeoB
2WJG_A	FeoB_N	sGTP.1	4	ERA-like	15	FeoB
2WJJ_A	FeoB_N	sGTP.1	4	ERA-like	15	FeoB
2WJH_A	FeoB_N	sGTP.1	4	ERA-like	15	FeoB
2WJI_A	FeoB_N	sGTP.1	4	ERA-like	15	FeoB
3A1T_A	FeoB_N+FeoB_Cyto	sGTP.1	4	ERA-like	15	FeoB
3A1S_A	FeoB_N+FeoB_Cyto	sGTP.1	4	ERA-like	15	FeoB
3H2Y_A	MMR_HSR1	GTPase-AIG.1	4	ERA-like	24	YawG_cp
2E87_A	NOG1_N+NOG1+MMR_HSR1_Xtn	sGTP.1	4	ERA-like	31	OBG
3QQ5_A	MMR_HSR1+Spore_IV_A+HydF_dimer+HydF_tetramer	sGTP.1	4	ERA-like	45	HydF
3W5I_A	FeoB_N	sGTP.1	4	ERA-like	NA	NA
1RFL_A	MMR_HSR1	sGTP.1	4	ERA-like	NA	NA
2GJ8_A	MMR_HSR1	sGTP.1	4	ERA-like	NA	NA
2GJ9_A	MMR_HSR1	sGTP.1	4	ERA-like	NA	NA
3B1X_A	FeoB_N+FeoB_Cyto	sGTP.1	4	ERA-like	NA	NA
3TAH_A	FeoB_N+FeoB_Cyto	sGTP.1	4	ERA-like	NA	NA
3B1Z_A	FeoB_N+FeoB_Cyto	sGTP.1	4	ERA-like	NA	NA
3LX5_A	FeoB_N+FeoB_Cyto	sGTP.1	4	ERA-like	NA	NA
3B1V_A	FeoB_N+FeoB_Cyto	sGTP.1	4	ERA-like	NA	NA
3B1Y_A	FeoB_N+FeoB_Cyto	sGTP.1	4	ERA-like	NA	NA
4NON_A	FeoB_N+FeoB_Cyto	sGTP.1	4	ERA-like	NA	NA
2CXX_A	MMR_HSR1+MMR_HSR1_Xtn	sGTP.1	4	ERA-like	NA	NA
2DYK_A	MMR_HSR1+MMR_HSR1_Xtn	sGTP.1	4	ERA-like	NA	NA
3GEH_A	TrmE_N+MnmE_helical	VirB11.1+sGTP.1	4	ERA-like	NA	NA
5UCV_A	MMR_HSR1	sGTP.1	4	ERA-like	NA	NA
2QU8_A	AAA_18+NOG1	sGTP.1	4	ERA-like	NA	NA
2QTV_B	Arf	sGTP.1	5	Arf-like	4	Arf Ras Superfamily
5DI3_A	Arf	sGTP.1	5	Arf-like	4	Arf Ras Superfamily
5DE3_A	Arf	sGTP.1	5	Arf-like	4	Arf Ras Superfamily
2B6H_A	Arf	sGTP.1	5	Arf-like	4	Arf Ras Superfamily
2KSQ_A	Arf	sGTP.1	5	Arf-like	4	Arf Ras Superfamily

4Q66_F	Arf	sGTP.1	5	Arf-like	4	Arf Ras Superfamily
4GOK_B	Arf	sGTP.1	5	Arf-like	4	Arf Ras Superfamily
3AQ4_A	Arf	sGTP.1	5	Arf-like	4	Arf Ras Superfamily
2AL7_A	Arf	sGTP.1	5	Arf-like	4	Arf Ras Superfamily
1ZD9_A	Arf	sGTP.1	5	Arf-like	4	Arf Ras Superfamily
2H18_A	Arf	sGTP.1	5	Arf-like	4	Arf Ras Superfamily
4ILE_A	Arf	sGTP.1	5	Arf-like	4	Arf Ras Superfamily
4ZI2_A	Arf	sGTP.1	5	Arf-like	4	Arf Ras Superfamily
4GOJ_A	Arf	sGTP.1	5	Arf-like	4	Arf Ras Superfamily
4V0K_A	Arf	sGTP.1	5	Arf-like	4	Arf Ras Superfamily
4V0M_E	Arf	sGTP.1	5	Arf-like	4	Arf Ras Superfamily
1ZJ6_A	Arf	sGTP.1	5	Arf-like	4	Arf Ras Superfamily
2H16_A	Arf	sGTP.1	5	Arf-like	4	Arf Ras Superfamily
2H17_A	Arf	sGTP.1	5	Arf-like	4	Arf Ras Superfamily
2FMX_A	Arf	sGTP.1	5	Arf-like	4	Arf Ras Superfamily
2FA9_A	Arf	sGTP.1	5	Arf-like	4	Arf Ras Superfamily
3TJZ_A	Arf	sGTP.1	5	Arf-like	4	Arf Ras Superfamily
2K5U_A	Arf	sGTP.1	5	Arf-like	4	Arf Ras Superfamily
2X77_A	Arf	sGTP.1	5	Arf-like	4	Arf Ras Superfamily
5EE5_B	Arf	sGTP.1	5	Arf-like	4	Arf Ras Superfamily
1UPT_A	Arf	sGTP.1	5	Arf-like	4	Arf Ras Superfamily
5J5C_A	Arf	sGTP.1	5	Arf-like	4	Arf Ras Superfamily
4DCN_A	Arf	sGTP.1	5	Arf-like	4	Arf Ras Superfamily
1R4A_A	Arf	sGTP.1	5	Arf-like	4	Arf Ras Superfamily
6A8D_A	Arf	sGTP.1	5	Arf-like	4	Arf Ras Superfamily
1HUR_A	Arf	sGTP.1	5	Arf-like	4	Arf Ras Superfamily
2A5G_A	Arf	sGTP.1	5	Arf-like	4	Arf Ras Superfamily
1E0S_A	Arf	sGTP.1	5	Arf-like	4	Arf Ras Superfamily
3VHX_A	Arf	sGTP.1	5	Arf-like	4	Arf Ras Superfamily
4KAX_A	Arf	sGTP.1	5	Arf-like	4	Arf Ras Superfamily
2W83_A	Arf	sGTP.1	5	Arf-like	4	Arf Ras Superfamily
3N5C_A	Arf	sGTP.1	5	Arf-like	4	Arf Ras Superfamily
3PCR_B	Arf	sGTP.1	5	Arf-like	4	Arf Ras Superfamily
6CM9_C	Arf	sGTP.1	5	Arf-like	4	Arf Ras Superfamily
4HMY_C	Arf	sGTP.1	5	Arf-like	4	Arf Ras Superfamily
6I16_C	Arf	sGTP.1	5	Arf-like	4	Arf Ras Superfamily
1J2J_A	Arf	sGTP.1	5	Arf-like	4	Arf Ras Superfamily
6CRI_C	Arf	sGTP.1	5	Arf-like	4	Arf Ras Superfamily
4BAS_A	Arf	sGTP.1	5	Arf-like	4	Arf Ras Superfamily
2H57_A	Arf	sGTP.1	5	Arf-like	4	Arf Ras Superfamily
5DI3_B	Arf+DUF1682	sGTP.1	5	Arf-like	4	Arf Ras Superfamily
4M9Q_A	Arf	sGTP.1	5	Arf-like	4	Arf Ras Superfamily
5CK4_A	AAA_16+SRPRB	DO-GTPase2.1	5	Arf-like	28	SRP
5CK3_B	AAA_16+SRPRB	DO-GTPase2.1	5	Arf-like	28	SRP
1NRJ_B	SRPRB	NA	5	Arf-like	28	SRP
2FH5_B	SRPRB	sGTP.1	5	Arf-like	28	SRP
5YMX_A	Arf	sGTP-MglA.1	5	Arf-like	30	MglA
3T12_A	Arf	sGTP-MglA.1	5	Arf-like	30	MglA
3T1T_A	Arf	sGTP-MglA.1	5	Arf-like	30	MglA
3R7W_A	Gtr1_RagA	NA	5	Arf-like	39	Rag_Ras Superfamily
3LLU_A	Gtr1_RagA	MNS.6+sGTP_APGTPase.2	5	Arf-like	67	NA
2Q3F_A	Gtr1_RagA	sGTP_APGTPase.2	5	Arf-like	67	NA
1KSH_A	Arf	sGTP.1	5	Arf-like	NA	NA
1FZQ_A	Arf	sGTP.1	5	Arf-like	NA	NA
3DOE_A	Arf	sGTP.1	5	Arf-like	NA	NA
1KSG_A	Arf	sGTP.1	5	Arf-like	NA	NA
1KSJ_A	Arf	sGTP.1	5	Arf-like	NA	NA
2GAO_A	Arf	sGTP.1	5	Arf-like	NA	NA
6FAE_B	Arf	sGTP.1	5	Arf-like	NA	NA
3LVQ_E	ArfGap+Ank_2+Arf	ANK.1+sGTP.1	5	Arf-like	NA	NA
6BBP_A	Sec7+PH+Arf	PH.1+sGTP.1	5	Arf-like	NA	NA
6BBQ_A	Sec7+PH+Arf	PH.1+sGTP.1	5	Arf-like	NA	NA
2GED_A	SRPRB	NA	5	Arf-like	NA	NA

3O47_A	ArfGap+Arf	sGTP.1	5	Arf-like	NA	NA
4FME_C	Arf	sGTP.1	5	Arf-like	NA	NA
3BH6_A	Arf	sGTP.1	5	Arf-like	NA	NA
3BH7_A	Arf	sGTP.1	5	Arf-like	NA	NA
4Y0V_A	Arf	sGTP.1	5	Arf-like	NA	NA
4YLG_A	Arf	sGTP.1	5	Arf-like	NA	NA
2LKC_A	GTP_EFTU	sGTP.1	5	Arf-like	NA	NA
4WNR_A	Roc+COR	sGTP_APGTPase.1	7	Roc	NA	NA
3DPU_A	Roc+COR	sGTP_APGTPase.1	7	Roc	NA	NA
6HLU_A	LRR_4+LRR_4+LRR_4+LRR_4+LRR_4+LRR_4+Roc+COR	ImmLRR.1+sGTP_APGTPase.1	7	Roc	NA	NA
3D6T_B	Roc	sGTP_APGTPase.1	7	Roc	NA	NA
2ZEJ_A	Roc	sGTP_APGTPase.1	7	Roc	NA	NA
6O81_S	GTP_EFTU+GTP_EFTU_D2+eIF2_C	sGTP_DYNAMIN.1	8	Translation_Factor-like	0	Classic Trafac
3ZZ0_A	GTP_EFTU+GTP_EFTU_D2+EFG_II+EFG_IV+EFG_C	NA	8	Translation_Factor-like	0	Classic Trafac
3ZZT_A	GTP_EFTU+GTP_EFTU_D2+EFG_II+EFG_IV+EFG_C	NA	8	Translation_Factor-like	0	Classic Trafac
3ZZU_A	GTP_EFTU+GTP_EFTU_D2+EFG_II+EFG_IV+EFG_C	NA	8	Translation_Factor-like	0	Classic Trafac
6BK7_A	GTP_EFTU+GTP_EFTU_D2	sGTP-MgIA.1	8	Translation_Factor-like	0	Classic Trafac
3VQT_A	GTP_EFTU+GTP_EFTU_D2+RF3_C	NA	8	Translation_Factor-like	0	Classic Trafac
3WBI_A	GTP_EFTU+GTP_EFTU_D2+IF-2+GTP_EFTU_D4	sGTP.1	8	Translation_Factor-like	0	Classic Trafac
4QHY_A	GTP_EFTU+GTP_EFTU_D2+eIF2_C	sGTP.1	8	Translation_Factor-like	0	Classic Trafac
4RD1_A	GTP_EFTU+GTP_EFTU_D2+eIF2_C	sGTP.1	8	Translation_Factor-like	0	Classic Trafac
4NBS_A	GTP_EFTU+GTP_EFTU_D2+eIF2_C	sGTP.1	8	Translation_Factor-like	0	Classic Trafac
4RCY_A	GTP_EFTU+GTP_EFTU_D2+eIF2_C	sGTP.1	8	Translation_Factor-like	0	Classic Trafac
4QFM_A	GTP_EFTU+GTP_EFTU_D2+eIF2_C	sGTP.1	8	Translation_Factor-like	0	Classic Trafac
4M4S_A	GTP_EFTU+GTP_EFTU_D2+eIF2_C	sGTP.1	8	Translation_Factor-like	0	Classic Trafac
2AHO_A	GTP_EFTU+GTP_EFTU_D2+eIF2_C	sGTP.1	8	Translation_Factor-like	0	Classic Trafac
4M2L_A	GTP_EFTU+GTP_EFTU_D2+eIF2_C	sGTP.1	8	Translation_Factor-like	0	Classic Trafac
3SJZ_A	GTP_EFTU+GTP_EFTU_D2+eIF2_C	sGTP.1	8	Translation_Factor-like	0	Classic Trafac
4M53_A	GTP_EFTU+GTP_EFTU_D2+eIF2_C	sGTP.1	8	Translation_Factor-like	0	Classic Trafac
3PEN_A	GTP_EFTU+GTP_EFTU_D2+eIF2_C	sGTP.1	8	Translation_Factor-like	0	Classic Trafac
3J5X_Z	GTP_EFTU+GTP_EFTU_D2+EFG_II+EFG_IV+EFG_C	NA	8	Translation_Factor-like	0	Classic Trafac
3J0E_H	GTP_EFTU+GTP_EFTU_D2+EFG_II+EFG_IV+EFG_C	NA	8	Translation_Factor-like	0	Classic Trafac
4B43_A	IF2_N+GTP_EFTU	sGTP.1	8	Translation_Factor-like	0	Classic Trafac
4B47_A	IF2_N+GTP_EFTU	sGTP.1	8	Translation_Factor-like	0	Classic Trafac
4B3X_A	IF2_N+GTP_EFTU	sGTP.1	8	Translation_Factor-like	0	Classic Trafac
1U2R_A	GTP_EFTU+GTP_EFTU_D2+EFG_II+EFG_IV+EFG_C	sGTP_DYNAMIN.1	8	Translation_Factor-like	0	Classic Trafac
6N0I_A	GTP_EFTU+GTP_EFTU_D2+EFG_II+EFG_IV+EFG_C	GTPase-AIG.1	8	Translation_Factor-like	0	Classic Trafac
4FN5_A	GTP_EFTU+GTP_EFTU_D2+EFG_II+EFG_IV+EFG_C	NA	8	Translation_Factor-like	0	Classic Trafac
4ZU9_A	GTP_EFTU+GTP_EFTU_D2+SelB-wing_3	sGTP.1	8	Translation_Factor-like	0	Classic Trafac
5VH6_A	GTP_EFTU+GTP_EFTU_D2	GTPase-AIG.1	8	Translation_Factor-like	0	Classic Trafac
4NCL_A	GTP_EFTU+GTP_EFTU_D2+IF-2	sGTP.1	8	Translation_Factor-like	0	Classic Trafac
6NOT_A	GTP_EFTU+GTP_EFTU_D2+EFG_II+EFG_IV+EFG_C	sGTP.1	8	Translation_Factor-like	0	Classic Trafac
6GFU_A	GTP_EFTU+GTP_EFTU_D2+GTP_EFTU_D3	sGTP_DYNAMIN.1	8	Translation_Factor-like	0	Classic Trafac
1G7R_A	GTP_EFTU+GTP_EFTU_D2+IF-2+GTP_EFTU_D4	sGTP.1	8	Translation_Factor-like	0	Classic Trafac
4N3G_A	GTP_EFTU+GTP_EFTU_D2+IF-2+GTP_EFTU_D4	sGTP.1	8	Translation_Factor-like	0	Classic Trafac
1WDT_A	GTP_EFTU+GTP_EFTU_D2+EFG_II+EFG_IV+EFG_C	GTPase-AIG.1	8	Translation_Factor-like	0	Classic Trafac
4UPY_B	GTP_EFTU+GTP_EFTU_D2+IF-2+GTP_EFTU_D4	sGTP.1	8	Translation_Factor-like	0	Classic Trafac
3IZQ_1	HBS1_N+GTP_EFTU+GTP_EFTU_D2+GTP_EFTU_D3	sGTP.1	8	Translation_Factor-like	0	Classic Trafac
3IZY_P	GTP_EFTU+IF-2	sGTP.1	8	Translation_Factor-like	0	Classic Trafac
3J25_A	GTP_EFTU+GTP_EFTU_D2+EFG_II+EFG_IV+EFG_C	GTPase-AIG.1	8	Translation_Factor-like	0	Classic Trafac
4ZV4_A	GTP_EFTU+GTP_EFTU_D2+GTP_EFTU_D3	sGTP.1	8	Translation_Factor-like	0	Classic Trafac
1ZUN_B	GTP_EFTU+GTP_EFTU_D2	sGTP.1	8	Translation_Factor-like	0	Classic Trafac
3P26_A	GTP_EFTU+GTP_EFTU_D2+GTP_EFTU_D3	sGTP.1	8	Translation_Factor-like	0	Classic Trafac
3P27_A	GTP_EFTU+GTP_EFTU_D2+GTP_EFTU_D3	sGTP.1	8	Translation_Factor-like	0	Classic Trafac
4AC9_A	GTP_EFTU+GTP_EFTU_D2+RIBIOP_C	sGTP.1	8	Translation_Factor-like	0	Classic Trafac
2YWG_A	GTP_EFTU+GTP_EFTU_D2+RF3_C+EFG_C+LepA_C	sGTP.1	8	Translation_Factor-like	0	Classic Trafac
3TR5_A	GTP_EFTU+GTP_EFTU_D2+RF3_C	NA	8	Translation_Factor-like	0	Classic Trafac
6O7K_FF	GTP_EFTU+IF-2+GTP_EFTU_D2	sGTP.1	8	Translation_Factor-like	0	Classic Trafac
1Z01_I	GTP_EFTU+IF-2+GTP_EFTU_D2	sGTP.1	8	Translation_Factor-like	0	Classic Trafac
1R5B_A	GTP_EFTU+GTP_EFTU_D2+GTP_EFTU_D3	GTPase-AIG.1	8	Translation_Factor-like	0	Classic Trafac
5M18_A	GTP_EFTU+GTP_EFTU_D2+GTP_EFTU_D3	sGTP.1	8	Translation_Factor-like	0	Classic Trafac
5M19_A	GTP_EFTU+GTP_EFTU_D2+GTP_EFTU_D3	sGTP.1	8	Translation_Factor-like	0	Classic Trafac
5M13_A	GTP_EFTU+GTP_EFTU_D2+GTP_EFTU_D3	sGTP.1	8	Translation_Factor-like	0	Classic Trafac

5OPD_A	GTP_EFTU+GTP_EFTU_D2+GTP_EFTU_D3	sGTP.1	8	Translation_Factor-like	0	Classic Trafac
3U6B_A	GTP_EFTU+GTP_EFTU_D2+GTP_EFTU_D3	sGTP.1	8	Translation_Factor-like	0	Classic Trafac
10B2_A	GTP_EFTU+GTP_EFTU_D2+GTP_EFTU_D3	sGTP.1	8	Translation_Factor-like	0	Classic Trafac
3WBJ_A	GTP_EFTU+GTP_EFTU_D2+IF-2	sGTP.1	8	Translation_Factor-like	0	Classic Trafac
4N3S_A	GTP_EFTU+GTP_EFTU_D2+IF-2	sGTP.1	8	Translation_Factor-like	0	Classic Trafac
5FG3_A	GTP_EFTU+GTP_EFTU_D2+IF-2+GTP_EFTU_D4	sGTP.1	8	Translation_Factor-like	0	Classic Trafac
1MJ1_A	GTP_EFTU+GTP_EFTU_D2+GTP_EFTU_D3	sGTP.1	8	Translation_Factor-like	0	Classic Trafac
2Y0Y_Z	GTP_EFTU+GTP_EFTU_D2+GTP_EFTU_D3	sGTP.1	8	Translation_Factor-like	0	Classic Trafac
2C77_A	GTP_EFTU+GTP_EFTU_D2+GTP_EFTU_D3	sGTP.1	8	Translation_Factor-like	0	Classic Trafac
2Y0U_Z	GTP_EFTU+GTP_EFTU_D2+GTP_EFTU_D3	sGTP.1	8	Translation_Factor-like	0	Classic Trafac
2C78_A	GTP_EFTU+GTP_EFTU_D2+GTP_EFTU_D3	sGTP.1	8	Translation_Factor-like	0	Classic Trafac
1TTT_A	GTP_EFTU+GTP_EFTU_D2+GTP_EFTU_D3	sGTP.1	8	Translation_Factor-like	0	Classic Trafac
1EFT_A	GTP_EFTU+GTP_EFTU_D2+GTP_EFTU_D3	sGTP.1	8	Translation_Factor-like	0	Classic Trafac
1EXM_A	GTP_EFTU+GTP_EFTU_D2+GTP_EFTU_D3	sGTP.1	8	Translation_Factor-like	0	Classic Trafac
4ABR_Z	GTP_EFTU+GTP_EFTU_D2+GTP_EFTU_D3	sGTP.1	8	Translation_Factor-like	0	Classic Trafac
1AIP_A	GTP_EFTU+GTP_EFTU_D2+GTP_EFTU_D3	sGTP.1	8	Translation_Factor-like	0	Classic Trafac
4LBV_A	GTP_EFTU+GTP_EFTU_D2+GTP_EFTU_D3	sGTP.1	8	Translation_Factor-like	0	Classic Trafac
5K0Y_S	GTP_EFTU+GTP_EFTU_D2+eIF2_C	sGTP_DYNAMIN.1	8	Translation_Factor-like	0	Classic Trafac
3WY9_A	GTP_EFTU+GTP_EFTU_D2+GTP_EFTU_D3	GTPase-AIG.1	8	Translation_Factor-like	0	Classic Trafac
5TV2_A	GTP_EFTU+GTP_EFTU_D2	NA	8	Translation_Factor-like	0	Classic Trafac
3DEG_C	GTP_EFTU+GTP_EFTU_D2+EFG_II+EFG_C+LepA_C	sGTP.1	8	Translation_Factor-like	0	Classic Trafac
4CRN_P	GTP_EFTU+GTP_EFTU_D2+GTP_EFTU_D3	NA	8	Translation_Factor-like	0	Classic Trafac
5H7J_A	GTP_EFTU+GTP_EFTU_D2+EFG_II+EFG_IV+EFG_C	sGTP.1+CADHERIN.2	8	Translation_Factor-like	0	Classic Trafac
2BV3_A	GTP_EFTU+GTP_EFTU_D2+EFG_II+EFG_IV+EFG_C	sGTP.1	8	Translation_Factor-like	0	Classic Trafac
2BM1_A	GTP_EFTU+GTP_EFTU_D2+EFG_II+EFG_IV+EFG_C	sGTP.1	8	Translation_Factor-like	0	Classic Trafac
4MYU_A	GTP_EFTU+GTP_EFTU_D2+EFG_II+EFG_IV+EFG_C	sGTP.1	8	Translation_Factor-like	0	Classic Trafac
4MYT_A	GTP_EFTU+GTP_EFTU_D2+EFG_II+EFG_IV+EFG_C	sGTP.1	8	Translation_Factor-like	0	Classic Trafac
1FNM_A	GTP_EFTU+GTP_EFTU_D2+EFG_II+EFG_IV+EFG_C	sGTP.1	8	Translation_Factor-like	0	Classic Trafac
2BM0_A	GTP_EFTU+GTP_EFTU_D2+EFG_II+EFG_IV+EFG_C	sGTP.1	8	Translation_Factor-like	0	Classic Trafac
2J7K_A	GTP_EFTU+GTP_EFTU_D2+EFG_II+EFG_IV+EFG_C	sGTP.1	8	Translation_Factor-like	0	Classic Trafac
4M1K_A	GTP_EFTU+GTP_EFTU_D2+EFG_II+EFG_IV+EFG_C	sGTP.1	8	Translation_Factor-like	0	Classic Trafac
1EFG_A	GTP_EFTU+GTP_EFTU_D2+EFG_II+EFG_IV+EFG_C	sGTP.1	8	Translation_Factor-like	0	Classic Trafac
3IZP_E	GTP_EFTU+GTP_EFTU_D2+EFG_II+EFG_IV+EFG_C	sGTP.1	8	Translation_Factor-like	0	Classic Trafac
4KBT_Y	GTP_EFTU+GTP_EFTU_D2+EFG_II+EFG_IV+EFG_C	sGTP.1	8	Translation_Factor-like	0	Classic Trafac
4JUW_Y	GTP_EFTU+GTP_EFTU_D2+EFG_II+EFG_IV+EFG_C	sGTP.1	8	Translation_Factor-like	0	Classic Trafac
1ZNO_B	GTP_EFTU+GTP_EFTU+GTP_EFTU_D2+EFG_II+EFG_IV+EFG_C	sGTP.1	8	Translation_Factor-like	0	Classic Trafac
5W76_A	GTP_EFTU+GTP_EFTU_D2+GTP_EFTU_D3	sGTP.1	8	Translation_Factor-like	0	Classic Trafac
5GAN_C	EFTUD2+GTP_EFTU+GTP_EFTU_D2+EFG_II+EFG_IV+EFG_C	NA	8	Translation_Factor-like	0	Classic Trafac
3J4J_A	IF2_N+GTP_EFTU+IF-2+GTP_EFTU_D2	sGTP.1	8	Translation_Factor-like	0	Classic Trafac
4KJZ_A	IF2_N+GTP_EFTU+IF-2	sGTP.1	8	Translation_Factor-like	0	Classic Trafac
5IZK_A	GTP_EFTU+GTP_EFTU_D2	NA	8	Translation_Factor-like	0	Classic Trafac
5IZM_A	GTP_EFTU+GTP_EFTU_D2	NA	8	Translation_Factor-like	0	Classic Trafac
1XB2_A	GTP_EFTU+GTP_EFTU_D2+GTP_EFTU_D3	NA	8	Translation_Factor-like	0	Classic Trafac
1D2E_A	GTP_EFTU+GTP_EFTU_D2+GTP_EFTU_D3	sGTP.1	8	Translation_Factor-like	0	Classic Trafac
2D74_A	GTP_EFTU+GTP_EFTU_D2+eIF2_C	sGTP.1	8	Translation_Factor-like	0	Classic Trafac
1KK0_A	GTP_EFTU+GTP_EFTU_D2+eIF2_C	sGTP.1	8	Translation_Factor-like	0	Classic Trafac
1KJZ_A	GTP_EFTU+GTP_EFTU_D2+eIF2_C	sGTP.1	8	Translation_Factor-like	0	Classic Trafac
1KK3_A	GTP_EFTU+GTP_EFTU_D2+eIF2_C	sGTP.1	8	Translation_Factor-like	0	Classic Trafac
5YT0_A	GTP_EFTU+GTP_EFTU_D2	sGTP.1	8	Translation_Factor-like	0	Classic Trafac
4BYX_V	GTP_EFTU+GTP_EFTU_D2+IF-2+GTP_EFTU_D4	sGTP.1	8	Translation_Factor-like	0	Classic Trafac
4ZCM_A	GTP_EFTU+GTP_EFTU_D2+EFG_II+EFG_C	sGTP.1	8	Translation_Factor-like	0	Classic Trafac
4ZCI_A	GTP_EFTU+GTP_EFTU_D2+EFG_II+EFG_C	sGTP.1	8	Translation_Factor-like	0	Classic Trafac
1S0U_A	GTP_EFTU+GTP_EFTU_D2+eIF2_C	sGTP.1	8	Translation_Factor-like	0	Classic Trafac
1JNY_A	GTP_EFTU+GTP_EFTU_D2+GTP_EFTU_D3	sGTP_DYNAMIN.1	8	Translation_Factor-like	0	Classic Trafac
5H7K_A	GTP_EFTU+GTP_EFTU_D2	sGTP.1	8	Translation_Factor-like	0	Classic Trafac
3UOQ_W	GTP_EFTU+GTP_EFTU_D2+RF3_C	sGTP.1	8	Translation_Factor-like	0	Classic Trafac
6B8D_A	GTP_EFTU+GTP_EFTU_D2	NA	8	Translation_Factor-like	0	Classic Trafac
1EFM_A	GTP_EFTU+GTP_EFTU_D2+GTP_EFTU_D3	sGTP.1	8	Translation_Factor-like	0	Classic Trafac
2HCJ_B	GTP_EFTU+GTP_EFTU_D2+GTP_EFTU_D3	PELOTA.1	8	Translation_Factor-like	0	Classic Trafac
4W22_4	GTP_EFTU+GTP_EFTU_D2+EFG_II+EFG_IV+EFG_C	NA	8	Translation_Factor-like	0	Classic Trafac
5W7Q_A	GTP_EFTU+GTP_EFTU_D2+GTP_EFTU_D3	sGTP.1	8	Translation_Factor-like	0	Classic Trafac
4C0S_A	GTP_EFTU+GTP_EFTU_D2+GTP_EFTU_D3	sGTP_DYNAMIN.1	8	Translation_Factor-like	0	Classic Trafac
5W75_A	GTP_EFTU+GTP_EFTU_D2+GTP_EFTU_D3	sGTP.1	8	Translation_Factor-like	0	Classic Trafac

3J5Y_B	GTP_EFTU+GTP_EFTU_D2+GTP_EFTU_D3	sGTP.1	8	Translation_Factor-like	0	Classic Trafac
4D61_II	GTP_EFTU+GTP_EFTU_D2+GTP_EFTU_D3	sGTP.1	8	Translation_Factor-like	0	Classic Trafac
3WXM_A	GTP_EFTU+GTP_EFTU_D2+GTP_EFTU_D3	sGTP.1	8	Translation_Factor-like	0	Classic Trafac
3VMF_A	GTP_EFTU+GTP_EFTU_D2+GTP_EFTU_D3	sGTP.1	8	Translation_Factor-like	0	Classic Trafac
4QJT_YY	Ribosomal_L9_N+GTP_EFTU+GTP_EFTU_D2+EFG_II+EFG_C+LepA_C	sGTP.1	8	Translation_Factor-like	0	Classic Trafac
3AGP_A	EF_TS+GTP_EFTU+GTP_EFTU_D2+GTP_EFTU_D3+RNA_replicase_B	EFTS.1+sGTP.1	8	Translation_Factor-like	NA	NA
508W_A	GTP_EFTU+GTP_EFTU_D2+GTP_EFTU_D3	sGTP_DYNAMIN.1	8	Translation_Factor-like	NA	NA
4ZKD_A	GTP_EFTU	DO-GTPase2.1	8	Translation_Factor-like	NA	NA
3MMP_A	EF_TS+GTP_EFTU+GTP_EFTU_D2+GTP_EFTU_D3	EFTS.1+sGTP.1	8	Translation_Factor-like	NA	NA
4R71_A	EF_TS+GTP_EFTU+GTP_EFTU_D2+GTP_EFTU_D3	EFTS.1+sGTP.1	8	Translation_Factor-like	NA	NA
5TY0_A	GTP_EFTU+GTP_EFTU_D2	NA	8	Translation_Factor-like	NA	NA
4TMT_A	GTP_EFTU+GTP_EFTU_D2	sGTP.1	8	Translation_Factor-like	NA	NA
4TN1_B	GTP_EFTU+GTP_EFTU_D2	sGTP.1	8	Translation_Factor-like	NA	NA
4TMX_A	GTP_EFTU+GTP_EFTU_D2	sGTP.1	8	Translation_Factor-like	NA	NA
4TMV_B	GTP_EFTU+GTP_EFTU_D2	sGTP.1	8	Translation_Factor-like	NA	NA
4BYR_P	GTP_EFTU+GTP_EFTU_D2	sGTP.1	8	Translation_Factor-like	NA	NA
4IW3_B	GTP_EFTU+GTP_EFTU_D2+GTP_EFTU_D3	sGTP.1	8	Translation_Factor-like	NA	NA
3L43_A	Dynamin_N+Dynamin_M	sGTP_DYNAMIN.1	10	L_Dynamin-like	9	Dynamin
5A3F_A	Dynamin_N+Dynamin_M+PH+GED	sGTP_DYNAMIN.1+PH.1	10	L_Dynamin-like	9	Dynamin
6DLU_B	Dynamin_N+Dynamin_M+PH+GED	PH.1	10	L_Dynamin-like	9	Dynamin
3ZVR_A	Dynamin_N+Dynamin_M+PH+GED	sinutoxin.1+PH.1	10	L_Dynamin-like	9	Dynamin
3SNH_A	Dynamin_N+Dynamin_M+PH+GED	PH.1	10	L_Dynamin-like	9	Dynamin
6QL4_A	Dynamin_N+Dynamin_M	sGTP_DYNAMIN.1	10	L_Dynamin-like	9	Dynamin
6FGZ_A	Dynamin_N+Dynamin_M+GED	NA	10	L_Dynamin-like	9	Dynamin
3SZR_A	Dynamin_N+Dynamin_M+GED	NA	10	L_Dynamin-like	9	Dynamin
5GTM_A	Dynamin_N+Dynamin_M+GED	NA	10	L_Dynamin-like	9	Dynamin
6RZT_A	Dynamin_N+Dynamin_M	sGTP_DYNAMIN.1	10	L_Dynamin-like	9	Dynamin
4WHJ_A	Dynamin_N+Dynamin_M+GED	NA	10	L_Dynamin-like	9	Dynamin
3T34_A	Dynamin_N+Dynamin_M	GTPase-AIG.1	10	L_Dynamin-like	9	Dynamin
4H1U_A	Dynamin_N+Dynamin_M	sGTP_DYNAMIN.1	10	L_Dynamin-like	9	Dynamin
3W6N_A	Dynamin_N+Dynamin_M	sGTP_DYNAMIN.1	10	L_Dynamin-like	9	Dynamin
2X2E_A	Dynamin_N+Dynamin_M	NA	10	L_Dynamin-like	9	Dynamin
3ZYC_A	Dynamin_N+Dynamin_M	NA	10	L_Dynamin-like	9	Dynamin
5D3Q_A	Dynamin_N+Dynamin_M	NA	10	L_Dynamin-like	9	Dynamin
4UUD_D	Dynamin_N+Dynamin_M	NA	10	L_Dynamin-like	9	Dynamin
4UUD_A	Dynamin_N+Dynamin_M	NA	10	L_Dynamin-like	9	Dynamin
2AKA_B	Dynamin_N+Dynamin_M	NA	10	L_Dynamin-like	9	Dynamin
4BEJ_A	Dynamin_N+Dynamin_M+GED	sGTP_DYNAMIN.1	10	L_Dynamin-like	9	Dynamin
4P4U_A	Dynamin_N+Dynamin_M	NA	10	L_Dynamin-like	NA	NA
4P4T_A	Dynamin_N+Dynamin_M	NA	10	L_Dynamin-like	NA	NA
4P4S_B	Dynamin_N+Dynamin_M	NA	10	L_Dynamin-like	NA	NA
4P4S_A	Dynamin_N+Dynamin_M	NA	10	L_Dynamin-like	NA	NA
6DEF_A	Dynamin_N+Dynamin_M+GED	NA	10	L_Dynamin-like	NA	NA
6DI7_A	Dynamin_N+Dynamin_M	NA	10	L_Dynamin-like	NA	NA
1JWY_A	Myosin_N+Myosin_head+Dynamin_N+Dynamin_M	NA	10	L_Dynamin-like	NA	NA
5AR1_A	Septin	GTPase-AIG.1	11	Septin	8	Septin
4YQF_B	Septin	GTPase-AIG.1	11	Septin	8	Septin
5CYO_A	Septin	GTPase-AIG.1	11	Septin	8	Septin
5CYP_A	Septin	GTPase-AIG.1	11	Septin	8	Septin
2QNR_A	Septin	GTPase-AIG.1	11	Septin	8	Septin
2QA5_A	Septin	GTPase-AIG.1	11	Septin	8	Septin
3FTQ_A	Septin	GTPase-AIG.1	11	Septin	8	Septin
5IRR_A	Septin	GTPase-AIG.1	11	Septin	NA	NA
3SOP_A	Septin	GTPase-AIG.1	11	Septin	NA	NA
4Z54_A	Septin	GTPase-AIG.1	11	Septin	NA	NA
4Z51_A	Septin	GTPase-AIG.1	11	Septin	NA	NA
6N0B_C	Septin	GTPase-AIG.1	11	Septin	NA	NA
3T5D_A	Septin	GTPase-AIG.1	11	Septin	NA	NA
3TW4_A	Septin	GTPase-AIG.1	11	Septin	NA	NA
6OYA_A	G-alpha	sGTP-MglA.1	15	G-alpha	13	G-alpha
6OY9_A	G-alpha	NA	15	G-alpha	13	G-alpha
3V00_C	G-alpha	NA	15	G-alpha	13	G-alpha
1GOT_A	G-alpha	NA	15	G-alpha	13	G-alpha

1FQJ_A	G-alpha	NA	15	G-alpha	13	G-alpha
1TND_A	G-alpha	NA	15	G-alpha	13	G-alpha
6GDG_D	G-alpha+G-alpha	sGTP.1	15	G-alpha	13	G-alpha
5G53_C	G-alpha+G-alpha	sGTP.1	15	G-alpha	13	G-alpha
1ZCA_A	G-alpha	NA	15	G-alpha	13	G-alpha
4QJ3_A	G-alpha	NA	15	G-alpha	13	G-alpha
3AH8_A	G-alpha	NA	15	G-alpha	13	G-alpha
2BCJ_Q	G-alpha	NA	15	G-alpha	13	G-alpha
6OIJ_A	G-alpha	NA	15	G-alpha	13	G-alpha
4GNK_A	G-alpha	NA	15	G-alpha	13	G-alpha
4EKC_A	G-alpha	NA	15	G-alpha	13	G-alpha
6NBF_A	G-alpha	sGTP-MglA.1	15	G-alpha	13	G-alpha
5JS7_A	G-alpha	NA	15	G-alpha	13	G-alpha
6MHE_A	G-alpha	NA	15	G-alpha	13	G-alpha
2ODE_A	G-alpha	sGTP_APGTPase.1	15	G-alpha	13	G-alpha
4G5R_A	G-alpha	NA	15	G-alpha	13	G-alpha
4G5O_A	G-alpha	NA	15	G-alpha	13	G-alpha
2IHB_A	G-alpha	NA	15	G-alpha	13	G-alpha
1SVK_A	G-alpha	sGTP_APGTPase.1	15	G-alpha	13	G-alpha
4G5Q_A	G-alpha	NA	15	G-alpha	13	G-alpha
1Y3A_A	G-alpha	NA	15	G-alpha	13	G-alpha
3ONW_A	G-alpha	NA	15	G-alpha	13	G-alpha
3QI2_A	G-alpha	NA	15	G-alpha	13	G-alpha
2XNS_A	G-alpha	NA	15	G-alpha	13	G-alpha
1KJY_A	G-alpha	NA	15	G-alpha	13	G-alpha
3QE0_A	G-alpha	NA	15	G-alpha	13	G-alpha
2IK8_A	G-alpha	NA	15	G-alpha	13	G-alpha
2GTP_A	G-alpha	NA	15	G-alpha	13	G-alpha
1BH2_A	G-alpha	NA	15	G-alpha	13	G-alpha
2G83_A	G-alpha	NA	15	G-alpha	13	G-alpha
6D9H_A	G-alpha	NA	15	G-alpha	13	G-alpha
1ZCB_A	G-alpha	NA	15	G-alpha	13	G-alpha
3CX6_A	G-alpha	NA	15	G-alpha	13	G-alpha
3CX7_A	G-alpha	NA	15	G-alpha	13	G-alpha
4FID_A	G-alpha	viral-AAA.2	15	G-alpha	13	G-alpha
2XTZ_A	G-alpha	D5ATPase.1	15	G-alpha	13	G-alpha
3OHM_A	G-alpha	NA	15	G-alpha	13	G-alpha
5DO9_A	G-alpha	NA	15	G-alpha	13	G-alpha
1SHZ_A	G-alpha	sGTP_APGTPase.1	15	G-alpha	13	G-alpha
6QNO_A	G-alpha	sGTP_APGTPase.1	15	G-alpha	13	G-alpha
3FFA_A	G-alpha	sGTP_APGTPase.1	15	G-alpha	13	G-alpha
4N0D_A	G-alpha	sGTP_APGTPase.1	15	G-alpha	13	G-alpha
5KDL_A	G-alpha	sGTP_APGTPase.1	15	G-alpha	13	G-alpha
2ZJY_A	G-alpha	sGTP_APGTPase.1	15	G-alpha	13	G-alpha
6CRK_A	G-alpha	sGTP_APGTPase.1	15	G-alpha	13	G-alpha
4PAN_A	G-alpha	sGTP_APGTPase.1	15	G-alpha	13	G-alpha
3UMR_A	G-alpha	sGTP_APGTPase.1	15	G-alpha	13	G-alpha
3D7M_A	G-alpha	sGTP_APGTPase.1	15	G-alpha	13	G-alpha
4PAM_A	G-alpha	sGTP_APGTPase.1	15	G-alpha	13	G-alpha
5TDH_A	G-alpha	sGTP_APGTPase.1	15	G-alpha	13	G-alpha
6CMO_A	G-alpha	sGTP_APGTPase.1	15	G-alpha	13	G-alpha
6OT0_A	G-alpha	NA	15	G-alpha	13	G-alpha
3UMS_A	G-alpha	sGTP_APGTPase.1	15	G-alpha	13	G-alpha
1GP2_A	G-alpha	sGTP_APGTPase.1	15	G-alpha	13	G-alpha
1AS0_A	G-alpha	NA	15	G-alpha	13	G-alpha
1GIL_A	G-alpha	sGTP_APGTPase.1	15	G-alpha	13	G-alpha
1GG2_A	G-alpha	sGTP_APGTPase.1	15	G-alpha	13	G-alpha
6R40_B	G-alpha	sGTP-MglA.1	15	G-alpha	NA	NA
6R3Q_B	G-alpha	sGTP-MglA.1	15	G-alpha	NA	NA
1CJK_C	G-alpha	sGTP-MglA.1	15	G-alpha	NA	NA
1AZS_C	G-alpha	sGTP-MglA.1	15	G-alpha	NA	NA
6B3J_A	G-alpha	NA	15	G-alpha	NA	NA
6EG8_I	G-alpha	sGTP-MglA.1	15	G-alpha	NA	NA

1CUL_C	G-alpha	sGTP-MglA.1	15	G-alpha	NA	NA
6AU6_A	G-alpha	NA	15	G-alpha	NA	NA
6FUF_B	G-alpha+G-alpha	sGTP.1	15	G-alpha	NA	NA
6G79_A	G-alpha+G-alpha	sGTP.1	15	G-alpha	NA	NA
6OIK_A	G-alpha	NA	15	G-alpha	NA	NA
3C7K_A	G-alpha	NA	15	G-alpha	NA	NA
6CUM_A	MeaB	ParA-Soj-PloopNTPase.1+EAD2.1	16	CobW-like	23	MeaB
3TK1_A	MeaB	CR-ATPase3.1	16	CobW-like	23	MeaB
3NXS_A	MeaB	ParA-Soj-PloopNTPase.1	16	CobW-like	23	MeaB
3MD0_A	MeaB	CR-ATPase1.3	16	CobW-like	23	MeaB
2P67_A	MeaB	sGTP.1	16	CobW-like	23	MeaB
2QM8_A	MeaB	CR-ATPase1.2	16	CobW-like	23	MeaB
2QM7_A	MeaB	CR-ATPase1.2	16	CobW-like	23	MeaB
1YR6_A	ATP_bind_1	Sak4.2	16	CobW-like	34	GPN1_R
5HCN_A	ATP_bind_1	Sak4.2	16	CobW-like	34	GPN1_R
5XKT_A	cobW	NA	16	CobW-like	42	UreG
2WWW_A	MeaB	NA	16	CobW-like	NA	NA
1U0L_A	RsgA_N+RsgA_GTPase	S1.4+sGTP.1+GTPase-AIG.1	17	RsgA-like	17	RsgA
4LV5_B	IIGP	GTPase-AIG.1	18	IIGP	19	IIGP
1TPZ_A	IIGP	GTPase-AIG.1	18	IIGP	19	IIGP
1TQ4_A	IIGP	GTPase-AIG.1	18	IIGP	19	IIGP
1TQ6_A	IIGP	GTPase-AIG.1	18	IIGP	19	IIGP
1TQD_A	IIGP	GTPase-AIG.1	18	IIGP	19	IIGP
5FPH_A	IIGP	GTPase-AIG.1	18	IIGP	19	IIGP