UNIVERSIDADE FEDERAL DE MINAS GERAIS

Instituto de Ciência Biológicas

Programa de Pós-graduação em Genética

Pedro Heringer Lisboa Teixeira

ORIGEM E EVOLUÇÃO DOS *HELITRONS*

Belo Horizonte

2022

Pedro Heringer Lisboa Teixeira

Origem e Evolução dos *Helitrons*

Tese apresentada ao Programa de Pós-Graduação em Genética da Universidade Federal de Minas Gerais como requisito parcial à obtenção do título de Doutor em Genética.

Orientador: Prof. Dr. Gustavo Campos e Silva Kuhn

Belo Horizonte

2022

043 Teixeira, Pedro Heringer Lisboa.
Origem e evolução dos Helitrons [manuscrito] / Pedro Heringer Lisboa Teixeira. – 2022.
160 f. : il. ; 29,5 cm.

Orientador: Prof. Dr. Gustavo Campos e Silva Kuhn.
Tese (doutorado) – Universidade Federal de Minas Gerais, Instituto de Ciências Biológicas. Programa de Pós-Graduação em Genética.

1. Genética. 2. Elementos de DNA transponíveis. I. Kuhn, Gustavo Campos e Silva. II. Universidade Federal de Minas Gerais. Instituto de Ciências Biológicas. III. Título.

CDU: 575

Ficha catalográfica elaborada pela bibliotecária Fabiane C M Reis – CRB 6 – 2680

UNIVERSIDADE FEDERAL DE MINAS GERAIS
Instituto de Ciências Biológicas
Programa de Pós-Graduação em Genética

ATA DE DEFESA DE DISSERTAÇÃO / TESE

ATA DA DEFESA DE TESE	**153/2022** **entrada**
Pedro Heringer Lisboa Teixeira	**2º/2017** **CPF: 083.643.596-65**

Às oito horas e trinta minutos do dia **24 de fevereiro de 2022**, reuniu-se remotamente (rede mundial de computadores), a Comissão Examinadora de Tese, indicada pelo Colegiado do Programa, para julgar, em exame final, o trabalho intitulado: **"Origem e Evolução dos Helitrons"**, requisito para obtenção do grau de Doutor em **Genética.** Abrindo a sessão, o Presidente da Comissão, **Gustavo Campos e Silva Kuhn**, após dar a conhecer aos presentes o teor das Normas Regulamentares do Trabalho Final, passou a palavra ao candidato, para apresentação de seu trabalho. Seguiu-se a arguição pelos Examinadores, com a respectiva defesa do candidato. Logo após, a Comissão se reuniu, sem a presença do candidato e do público, para julgamento e expedição de resultado final. Foram atribuídas as seguintes indicações:

Prof./Pesq.	**Instituição**	**CPF**	**Indicação**
Gustavo Campos e Silva Kuhn	UFMG	260.136.648-62	Aprovado
Elgion Lucio da Silva Loreto	UFSM	324127700-34	Aprovado
Claudia Marcia Aparecida Carareto	UNESP	785924538-87	Aprovado
Leonardo Barbosa Koerich	UFMG	033.549.409-99	Aprovado
Renato Santana de Aguiar	UFMG	000.086.336-06	Aprovado

Pelas indicações, o candidato foi considerado: APROVADO.

O resultado final foi comunicado publicamente ao candidato pelo Presidente da Comissão. Nada mais havendo a tratar, o Presidente encerrou a reunião e lavrou a presente ATA, que será assinada por todos os membros participantes da Comissão Examinadora.

Belo Horizonte, 24 de fevereiro de 2022.

Gustavo Campos e Silva Kuhn - UFMG

Elgion Lucio da Silva Loreto - UFSM

Claudia Marcia Aparecida Carareto - UNESP

Leonardo Barbosa Koerich - UFMG

Renato Santana de Aguiar - UFMG

Assinatura dos membros da banca examinadora:

Documento assinado eletronicamente por **Claudia Marcia Aparecida Carareto**, **Usuário Externo**, em 24/02/2022, às 13:35, conforme horário oficial de Brasília, com fundamento no art. 5º do Decreto nº 10.543, de 13 de novembro de 2020.

Documento assinado eletronicamente por **Gustavo Campos e Silva Kuhn**, **Professor do Magistério Superior**, em 24/02/2022, às 13:53, conforme horário oficial de Brasília, com fundamento no art. 5º do Decreto nº 10.543, de 13 de novembro de 2020.

Documento assinado eletronicamente por **Renato Santana de Aguiar**, **Professor do Magistério Superior**, em 24/02/2022, às 18:15, conforme horário oficial de Brasília, com fundamento no art. 5º do Decreto nº 10.543, de 13 de novembro de 2020.

Documento assinado eletronicamente por **Leonardo Barbosa Koerich**, **Professor do Magistério Superior**, em 27/02/2022, às 20:15, conforme horário oficial de Brasília, com fundamento no art. 5º do Decreto nº 10.543, de 13 de novembro de 2020.

Documento assinado eletronicamente por **Élgion Lúcio da Silva Loreto**, **Usuário Externo**, em 04/03/2022, às 17:33, conforme horário oficial de Brasília, com fundamento no art. 5º do Decreto nº 10.543, de 13 de novembro de 2020.

A autenticidade deste documento pode ser conferida no site https://sei.ufmg.br/sei/controlador_externo.php?acao=documento_conferir&id_orgao_acesso_externo=0, informando o código verificador **1273169** e o código CRC **268CA550**.

UNIVERSIDADE FEDERAL DE MINAS GERAIS
Instituto de Ciências Biológicas
Programa de Pós-Graduação em Genética

FOLHA DE APROVAÇÃO

"Origem e Evolução dos Helitrons"

Pedro Heringer Lisboa Teixeira

Tese aprovada pela banca examinadora constituída pelos Professores:

Gustavo Campos e Silva Kuhn
UFMG

Elgion Lucio da Silva Loreto
UFSM

Claudia Marcia Aparecida Carareto
UNESP

Leonardo Barbosa Koerich
UFMG

Renato Santana de Aguiar
UFMG

Belo Horizonte, 24 de fevereiro de 2022.

sei! assinatura eletrônica

Documento assinado eletronicamente por **Claudia Marcia Aparecida Carareto**, **Usuário Externo**, em 24/02/2022, às 13:35, conforme horário oficial de Brasília, com fundamento no art. 5º do Decreto nº 10.543, de 13 de novembro de 2020.

sei! assinatura eletrônica

Documento assinado eletronicamente por **Gustavo Campos e Silva Kuhn**, **Professor do Magistério Superior**, em 24/02/2022, às 13:52, conforme horário oficial de Brasília, com fundamento no art. 5º do Decreto nº 10.543, de 13 de novembro de 2020.

sei! assinatura eletrônica

Documento assinado eletronicamente por **Renato Santana de Aguiar**, **Professor do Magistério Superior**, em 24/02/2022, às 18:15, conforme horário oficial de Brasília, com fundamento no art. 5º do Decreto nº 10.543, de 13 de novembro de 2020.

Documento assinado eletronicamente por **Leonardo Barbosa Koerich**, **Professor do Magistério Superior**, em 27/02/2022, às 20:15, conforme horário oficial de Brasília, com fundamento no art. 5º do Decreto nº 10.543, de 13 de novembro de 2020.

Documento assinado eletronicamente por **Élgion Lúcio da Silva Loreto**, **Usuário Externo**, em 04/03/2022, às 17:33, conforme horário oficial de Brasília, com fundamento no art. 5º do Decreto nº 10.543, de 13 de novembro de 2020.

A autenticidade deste documento pode ser conferida no site https://sei.ufmg.br/sei/controlador_externo.php?acao=documento_conferir&id_orgao_acesso_externo=0, informando o código verificador **1273176** e o código CRC **4FBADABC**.

Referência: Processo nº 23072.210628/2022-97 SEI nº 1273176

AGRADECIMENTOS

Gostaria de agradecer primeiramente ao Professor Gustavo C. S. Kuhn que contribuiu de forma essencial não só para a concretização deste trabalho, mas também para a minha formação acadêmica e profissional. Sua orientação desde a graduação, passando pelo mestrado, e agora no doutorado, foi fundamental para que eu pudesse me tornar o cientista que sou hoje. Obrigado por ser o melhor mentor possível e sempre confiar no meu potencial.

Agradeço aos membros atuais e os que já passaram pelo Laboratório de Citogenômica Evolutiva. Além de todas as discussões científicas interessantes, o companheirismo e a amizade de vocês ajudaram a tornar minha jornada acadêmica mais leve e agradável.

Muito obrigado à CAPES por financiar todo o período do meu mestrado e doutorado, contribuindo para que eu me mantivesse fazendo o que mais gosto na vida.

Também não poderia deixar de agradecer aos meus amigos que estão fora do meio acadêmico. Apesar de não entenderem muito bem o que eu faço, vocês sempre têm uma palavra de apoio e reconhecimento sobre o meu trabalho, e mesmo com a minha ausência em vários momentos, vocês se mantiveram companheiros.

Sou extremamente grato a todos da minha numerosa família. Aos avôs e avós que se foram, mas deixaram para mim o exemplo de vida e caráter a ser seguido, e minha querida avó que, em vida, continua a me ensinar o que de fato é importante, mesmo estando longe. Obrigado aos meus tios, tias e primos que me fazem sentir especial com tanto amor. Agradeço à minha irmã Laura, que considero um exemplo de sucesso e, mesmo estando distante, sempre foi uma das primeiras pessoas a torcer por mim.

Em especial, agradeço a minha amada esposa Isabela que tem sido uma fonte de suporte, alegria e amor constante nestes anos que se passaram. Mesmo sem ter participado diretamente, o fruto deste trabalho também é seu. Eu jamais teria chegado até aqui sem você.

Por fim, minha gratidão infinita ao meu pai e minha mãe, meus heróis e maiores inspirações. É impossível imaginar qualquer conquista na minha vida sem a participação deles. Não só este trabalho, mas a minha vida é dedicada a vocês dois.

"Nothing in biology makes sense except in the light of evolution"

Theodosius G. Dobzhansky

Resumo

Elementos de transposição (TEs) são sequências de DNA móveis e abundantes em genomas procarióticos e eucarióticos. Em eucariotos, TEs podem ser divididos em duas classes, denominadas classe I, que utilizam intermediários de RNA para se transporem, e classe II, que utilizam intermediários de DNA. Cada uma destas classes compreende diferentes subclasses, que por sua vez são divididas em superfamílias e famílias. *Helitrons* representam uma subclasse de elementos dentro da classe II que se transpõem por meio de um mecanismo único em eucariotos, sendo encontrados em todos os principais grupos taxonômicos deste domínio da vida. Estes transposons impactam genomas eucarióticos por ocuparem frações consideráveis do DNA de seus hospedeiros, além de estarem envolvidos na mobilização e duplicação de fragmentos cromossômicos adjacentes. Embora a compreensão sobre vários aspectos relacionados aos *Helitrons* tenha avançado consideravelmente nas duas décadas que sucederam a descoberta destes elementos, sua origem evolutiva e detalhes do seu mecanismo de transposição são temas que permaneceram amplamente inexplorados durante o mesmo período. Neste trabalho, investigamos a origem dos *Helitrons* através de análises evolutivas dos dois domínios principais presentes na sua transposase. Os resultados das análises de cada domínio revelam aspectos distintos, porém complementares, sobre a origem dos *Helitrons*. Em conjunto, nossos achados indicam que estes elementos descendem de plasmídeos procarióticos que, após invadirem genomas eucarióticos, passaram a utilizar a transposição como mecanismo de replicação em seus hospedeiros. Este cenário se opõe às principais hipóteses apresentadas até o momento para explicar a origem dos *Helitrons* e dos domínios da sua transposase. Além disso, com base nas evidências obtidas neste trabalho e em outros estudos, propomos que a transposase dos *Helitrons* desempenha funções catalíticas mais complexas do que havia sido sugerido anteriormente. Por fim, nossa investigação paralela sobre a evolução de uma família de *Helitrons* presente em artrópodes ilustra a capacidade notável destes transposons invadirem novos genomas hospedeiros por meio de transferências horizontais que podem ocorrer entre ordens ou mesmo classes distintas de organismos.

Palavras-chave: *Helitrons*. Elementos de transposição. Transposon. Transferência horizontal.

Abstract

Transposable elements (TEs) are mobile DNA sequences found in a large number of copies in prokaryotic and eukaryotic genomes. In eukaryotes, TEs can be divided into two classes, named class I, which use RNA intermediates to transpose, and class II, which use DNA intermediates. Each one of these classes include different subclasses, which in turn are divided into superfamilies and families. *Helitrons* represent a subclass of elements within class II that transpose by a mechanism that is unique in eukaryotes, being found in all major taxonomic groups from this domain of life. These transposons impact eukaryotic genomes by occupying considerable DNA fractions of their hosts, also being involved in the mobilization and duplication of adjacent chromosomal fragments. Although the understanding about several aspects related to *Helitrons* has advanced considerably in the two decades that followed their discovery, their evolutionary origin and details of their transposition mechanism are subjects that remained largely unexplored during the same period. In this work, we investigate the origin of *Helitrons* using evolutionary analyses of the two major domains present in their transposase. The results from the analyses of each domain reveal distinct, albeit complementary, aspects about the origin of *Helitrons*. Together, our findings indicate that these elements descend from procaryotic plasmids that, after invading eukaryotic genomes, started using transposition as the replication mechanism in their hosts. This scenario opposes the main hypotheses that have been advanced to explain the origin of *Helitrons* and the domains of their transposase. Furthermore, based on the evidence provided in this work and other studies, we propose that *Helitron* transposases execute more complex catalytic functions than it was previously suggested. Finally, our parallel investigation about the evolution of a *Helitron* family found in arthropods illustrate the marked capacity of these transposons to invade new host genomes through horizontal transfers that can occur between distinct orders or even classes of organisms.

Keywords: *Helitrons*. Transposable elements. Transposon. Horizontal transfer.

LISTA DE ILUSTRAÇÕES

Figuras do Capítulo 1

Figuras do Capítulo 2

LISTA DE ABREVIATURAS

3’-OH – Grupo hidroxila presente na extremidade 3’ do DNA

AP – *apurinic–apyrimidinic* (‘apurínica-apirimidínica’)

bp – *base pairs* (‘pares de base’)

CvBV – *Cotesia vestalis* bracovirus

dsDNA – *Double-strand DNA* (‘DNA dupla-Fita’)

Hel – Domínio helicase presente na RepHel

Hel_c35 – Família de Helitrons presente em artrópodes e descoberta no genoma de CvBV

HGT – *Horizontal Gene Transfer* (‘Transferência Horizontal de Gene’)

HT – *Horizontal Transfer* (‘Transferência Horizontal’)

HTT – *Horizontal Transposon Transfer* (‘Transferência Horizontal de Transposons’)

LTR – L*ong Terminal Repeat* (‘Repetição Terminal Longa’)

MGE – *Mobile Genetic Element* (‘Elemento Genético Móvel’)

MYA – *Million Years Ago* (‘Milhão de Anos Atrás’)

NCLDV – *Nucleocytoplasmic large DNA viruses* (‘Vírus nucleocitoplasmáticos de DNA grande’)

NMDS – *Non-metric multidimensional scaling* (‘Escalonamento Multidimensional Não-Métrico’)

ORF – *Open Reading Frame* (‘Fase de Leitura Aberta’)

PCNA – P*roliferating Cell Nuclear Antigen* (‘Antígeno Nuclear de Células em Proliferação’)

RC – *Rolling-Circle* (‘Círculo-Rolante’)

RCR – *Rolling-Circle Replication* (‘Replicação por Círculo-Rolante’)

RCRE – *RCR endonuclease domain* (‘Domínio endonuclease utilizado na RCR’)

RCT – *Rolling-Circle Transposition* (Transposição por ‘Círculo-Rolante’)

Rep – Domínio catalítico central presente na RepHel

RepHel – Transposase dos *Helitrons*

S1H – *Superfamily 1 helicase* (‘Helicase da superfamília 1’)

S3H – *Superfamily 3 helicase* (‘Helicase da superfamília 3’)

SH-aLRT – Teste da razão de verossimilhança aproximada com correção Shimodaira–Hasegawa

TE – *Transposable Element* (Elemento Transponível)

SUMÁRIO

1. INTRODUÇÃO

1.1 Elementos de transposição

Elementos de transposição (TEs), são sequências de DNA capazes de se mover nos genomas dos seus hospedeiros, e assim se replicar de maneira independente destes. TEs representam frações consideráveis do DNA de praticamente todos organismos eucariotos, sendo que a proporção ocupada por estes elementos apresenta uma forte correlação com o próprio tamanho genômico de seus hospedeiros. Além de impactar diretamente o tamanho genômico de eucariotos, TEs estão frequentemente associados a mutações, polimorfismos, rearranjos cromossômicos e, em alguns casos, são fonte de fatores moduladores da atividade gênica (revisado em Bourque et al. 2018, Wells & Feschotte 2020).

Apesar de estarem associados a inovações evolutivas benéficas para os seus hospedeiros em alguns casos isolados, os TEs representam entidades genéticas essencialmente 'egoístas' que geralmente evoluem nos genomas em que habitam de forma neutra ou afetando estes negativamente. Por esta razão, é esperado que, com o passar do tempo, linhagens de TEs sejam eliminadas dos seus genomas hospedeiros por seleção negativa e/ou deriva genética em algum momento de sua evolução. De fato, assim como outros parasitas, TEs podem utilizar diferentes estratégias para evadir tais processos que promovem sua eliminação. Entretanto, durante longos períodos evolutivos (dezenas ou centenas de milhões de anos) de transmissão vertical em seus genomas hospedeiros, tais estratégias seriam capazes de apenas adiar a extinção aparentemente inevitável destes elementos (revisado em Schaack et al. 2010).

Ao contrário da herança vertical, o processo conhecido como transferência horizontal (horizontal transfer, HT) ocorre através da transmissão de um segmento de DNA de um organismo para o genoma de outro (Wallau et al. 2018, Van Etten & Bhattacharya 2020), sendo assim uma alternativa para TEs escaparem sua extinção. Deste modo, a HT de TEs para novos genomas hospedeiros representa o principal mecanismo para explicar a persistência destes elementos no longo prazo (Schaack et al. 2010).

1.1.2 Classificação dos TEs

Quanto à sua classificação, TEs eucarióticos podem ser divididos em duas classes principais, definidas pelo tipo de intermediário de transposição gerado. Cada uma destas classes pode ser dividida em subclasses, definidas pelo mecanismo enzimático em que intermediários são gerados e inseridos, além de superfamílias e famílias, definidas pela relação filogenética dos seus membros (Bourque et al. 2018, Wells & Feschotte 2020). Elementos de classe I, também conhecidos como retrotransposons, utilizam intermediários de

RNA para se replicar. Estes intermediários são gerados por transcrição e posteriormente transcritos reversamente em DNA antes de serem integrados em um novo local do genoma hospedeiro, sendo que os elementos geradores dos intermediários permanecem intactos. Por esta razão, os elementos pertencentes à classe I também são referidos como sendo do tipo “copia-e-cola”. Já elementos de classe II, também conhecidos como transposons de DNA, utilizam intermediários de DNA para se replicar. Como a grande maioria dos grupos de TEs nesta classe geram intermediários por meio da excisão do próprio elemento doador e reinserção em uma nova localidade do genoma hospedeiro, estes elementos também são referidos como sendo do tipo “corta-e-cola”.

Entretanto, dentro da classe II, há duas subclasses de elementos que utilizam mecanismos de transposição distintos do padrão geral corta-e-cola, os Polintons (ou Mavericks) e os *Helitrons*. A primeira dessas subclasses compreende os Polintons que, apesar de não terem sido estudados em detalhe quanto ao seu mecanismo de transposição, provavelmente sintetizam intermediários de DNA diretamente a partir dos elementos doadores (Wells & Feschotte 2020). Mesmo que sejam considerados como TEs, o conjunto de evidências obtidas nos últimos anos indica de forma inequívoca que Polintons teriam se derivado de integrações virais em genomas hospedeiros e, por isso, provavelmente deveriam ser classificados como virus (Krupovic et al. 2014, Krupovic & Koonin 2015, Koonin & Krupovic 2017, Bellas & Sommaruga 2021).

1.2 *Helitrons*

A segunda subclasse de elementos da classe II que não utilizam um mecanismo de transposição do tipo corta-e-cola compreende os transposons conhecidos como *Helitrons*. Estes elementos eucarióticos foram identificados pela primeira vez em 2001 nos genomas de *Arabidopsis thaliana*, *Oriza sativa* e *Caenorhabditis elegans*, através de análises in silico (Kapitonov & Jurka 2001). Desde então, os *Helitrons* foram encontrados nos genomas de todos os principais grupos de organismos eucariotos em diferentes proporções. Por exemplo, *Helitrons* podem representar entre 0.1%-6.6% do DNA genômico em espécies de plantas e entre 0%-10% no caso de espécies animais (Kapitonov & Jurka 2007, Thomas & Pritham 2015). Estes transposons são encontrados em diferentes tamanhos que podem variar de poucas centenas de pb até poucos kb em elementos não-autônomos (que não codificam uma transposase funcional), e de poucos kb até várias dezenas de kb em elementos autônomos, dependendo do organismo hospedeiro e da família de *Helitron* em questão (e.g., Kapitonov & Jurka 2001, Pritham & Feschotte 2007, Du et al. 2009, Thomas et al. 2014, Chellapan et al. 2016).

Helitrons codificam uma transposase denominada RepHel, composta por dois domínios principais: uma endonuclease (Rep) e uma helicase (Hel) pertencente a superfamília 1 (S1H). O domínio Rep é o centro catalítico responsável pela clivagem do DNA nas extremidades do elemento doador e do sítio de inserção no cromossomo hospedeiro. Já o domínio Hel provavelmente é responsável por auxiliar na separação do DNA dupla fita (dsDNA) do elemento doador, gerando um intermediário de DNA fita simples (ssDNA). Além destes dois domínios comuns a todas transposases RepHel, *Helitrons* também possuem uma sequência palindrômica de 16-20 pb localizada ~ 11 pb antes da sua extremidade 3', capaz de formar estruturas secundárias do tipo hairpin ou stem-loop que provavelmente auxiliam no processo de transposição (Thomas & Pritham 2015) (Fig. 1).

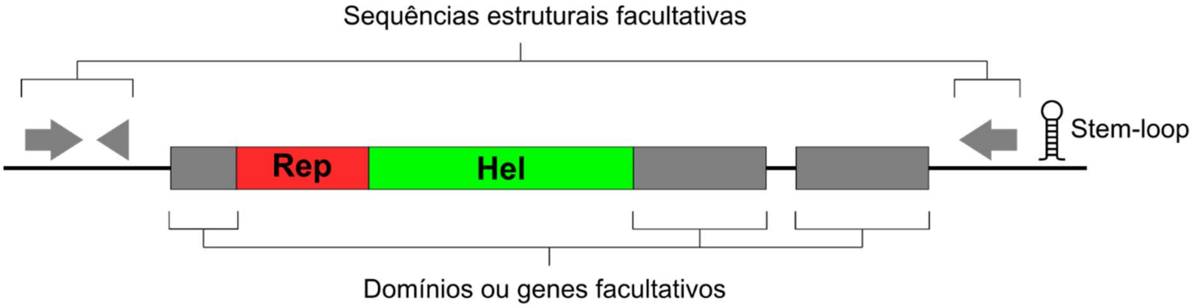

Sequências estruturais facultativas
Stem-loop
Rep
Hel
Domínios ou genes facultativos

Figura 1. Estrutura geral dos *Helitrons*. Domínios Rep e Hel estão presentes em todas transposases RepHel e estruturas do tipo stem-loop são encontradas em todos os *Helitrons*. Já os domínios, genes e estruturas restantes podem ou não estar presentes em diferentes variantes dos *Helitrons*.

Desde a descoberta dos *Helitrons*, notou-se a RepHel apresenta similaridades estruturais com transposases encontradas em elementos procarióticos (e.g., família IS91). Por isso, antes que estudos experimentais fossem conduzidos, todos os modelos sugeridos para descrever a transposição dos *Helitrons* se baseavam no mecanismo utilizado por TEs da família IS91 (Feschotte & Wessler 2001, Kapitonov & Jurka 2007, Thomas & Pritham 2015, Dias et al. 2016). Este mecanismo (Fig. 2), denominado transposição por círculo rolante (rolling-circle transposition, RCT) representa uma variação do processo conhecido como replicação por círculo rolante (rolling-circle replication, RCR), utilizado por diversos grupos de vírus e plasmídeos encontrados em organismos procariotos e eucariotos (Chandler et al. 2013, Wawrzyniak et al. 2017). Mais recentemente, análises experimentais confirmaram as principais etapas sugeridas para descrever a transposição dos *Helitrons*, além de revelar detalhes como, por exemplo, o fato de elementos circulares de dsDNA serem os intermediários viáveis de transposição (Grabundzija et al. 2016, 2018).

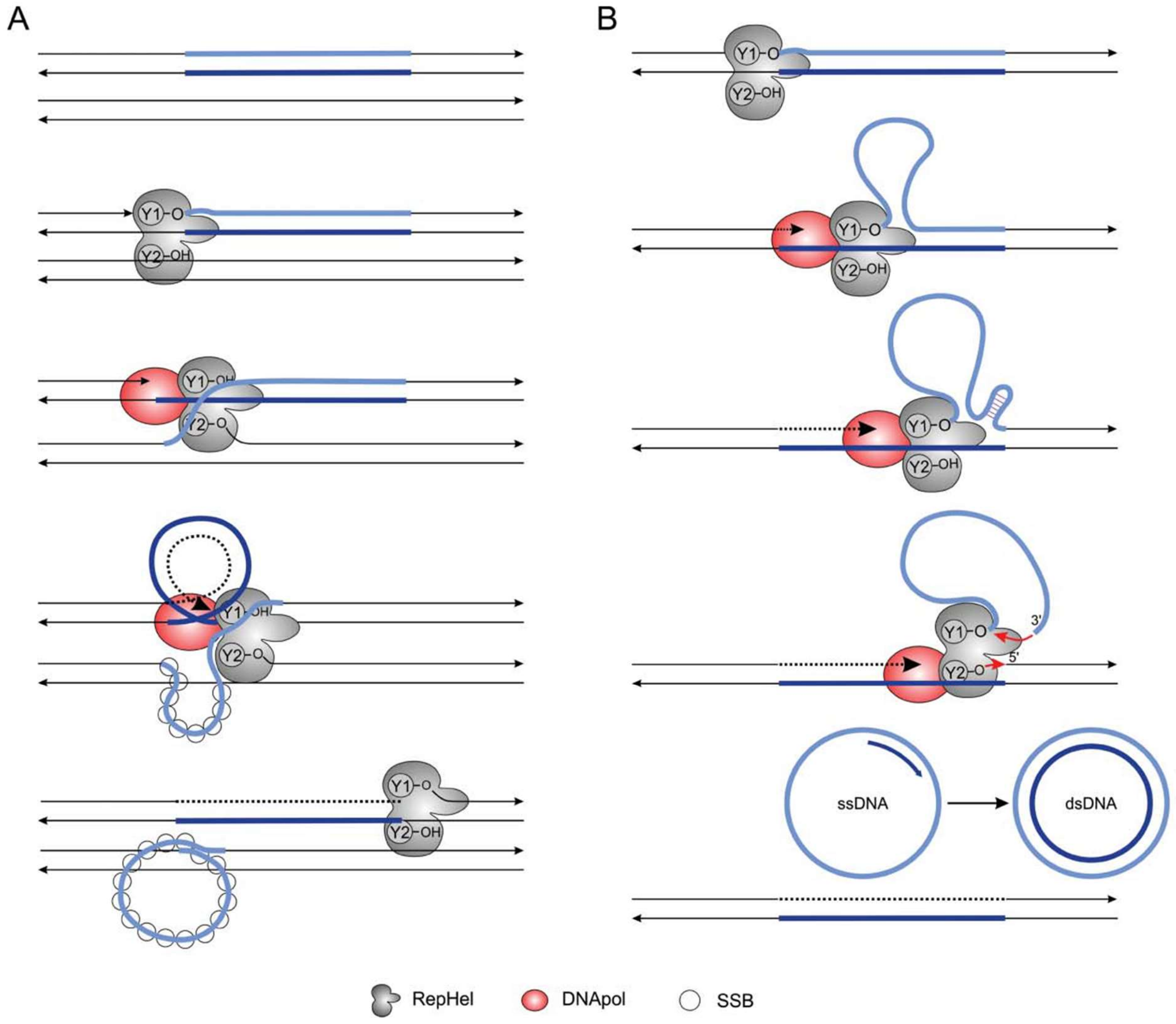


Figura 2. Mecanismos propostos para a transposição dos *Helitrons*. (A) Principal modelo sugerido para explicar a transposição dos *Helitrons* até 2016, baseado no mecanismo proposto para elementos bacterianos da família IS91. (B) Modelo alternativo sugerido pelo nosso grupo (Dias et al. 2016) para explicar inserções em tandem de *Helitrons*, baseado no fato de que elementos da família IS91 são capazes de gerar intermediários circulares. Estudos posteriores confirmaram que *Helitrons* são capazes de gerar intermediários circulares (Grabundzija et al. 2016) e que apenas intermediários circulares de dsDNA representam substratos de transposição viáveis (Grabundzija et al. 2018). Desta forma, o conjunto de dados indica que *Helitrons* geram intermediários através de um mecanismo mais próximo ao representado pelo segundo modelo (B). Em todo caso, após a geração de intermediários circulares de dsDNA, *Helitrons* provavelmente utilizam um mecanismo catalítico semelhante ao representado no primeiro modelo (A) para se integrarem no sítio receptor do hospedeiro. Figura adaptada de Dias et al. (2016).

De acordo com o que sabemos atualmente sobre os processos RCR e RCT, inclusive em *Helitrons*, a transposição destes elementos se inicia com a ligação entre a primeira tirosina catalítica do domínio Rep e a extremidade 5’ do elemento, criando um intermediário 5’-fosfotirosina e uma extremidade 3’-OH livre no sítio doador. A fita líder ligada covalentemente ao domínio Rep começa a se desassociar da sua fita complementar, provavelmente com o auxílio da atividade de translocação sentido 5’-3’ do domínio Hel. Ao mesmo tempo que a extremidade 5’ da fita líder começa a ser desassociada, uma forquilha de replicação possivelmente se forma no mesmo local, promovendo a síntese das fitas complementares tanto do intermediário em desassociação, quanto do sítio doador a partir de sua 3’-OH terminal. Desta forma, um intermediário de dsDNA é sintetizado até que a RepHel alcança o lado oposto do elemento, clivando este com sua segunda tirosina catalítica e expondo uma extremidade 3’-OH livre que ataca a primeira ligação 5’-fosfotirosina, resultando na formação de um intermediário de dsDNA circular.

Em um segundo momento, a RepHel ligada covalentemente à extremidade 5’ deste intermediário circular se associa à um segundo local do genoma hospedeiro que é clivado pela segunda tirosina presente na transposase, expondo uma extremidade 3’-OH livre do sítio receptor. Esta extremidade então ataca a ligação 5’-fosfotirosina entre a RepHel e o intermediário, ligando este ao DNA receptor. Após alcançar o lado oposto do intermediário circular, a primeira tirosina catalítica cliva este, gerando uma extremidade 3’-OH que ataca a ligação 5’-fosfotirosina entre o sítio receptor e a segunda tirosina catalítica. Tal processo resulta na inserção do *Helitron* na forma de um “loop” de DNA fita simples no sítio receptor que provavelmente é resolvido durante a replicação do genoma hospedeiro (Fig. 2).

1.2.1 Origem evolutiva dos *Helitrons*

Desde a descoberta dos *Helitrons*, foram propostas diferentes hipóteses para explicar sua origem e determinar quais seriam os elementos genéticos móveis mais próximos evolutivamente destes TEs eucarióticos. Por um lado, a semelhança estrutural e aparente semelhança funcional da RepHel com transposases de elementos procarióticos (e.g., família IS91) foi interpretada como um indício de que *Helitrons* seriam descendentes diretos ou parentes próximos destes últimos. Além disso, na época em que *Helitrons* foram descobertos em espécies de plantas e animais, vírus eucarióticos do tipo RCR haviam sido identificados apenas em espécies de plantas. Tal fato foi utilizado para sugerir a hipótese de que os *Helitrons* não só descenderiam de TEs procarióticos, mas talvez tivessem dado origem a vírus eucarióticos do tipo RCR (Kapitonov & Jurka 2001). Por outro lado, foi sugerida a hipótese alternativa de que os *Helitrons* poderiam ter se originado a partir de integrações ancestrais de vírus eucarióticos do tipo RCR (Feschotte & Wessler 2001). Esta hipótese foi baseada no fato de que, ao contrário dos transposons procarióticos do tipo RCT, os *Helitrons* codificam uma helicase e, em alguns casos, proteínas que se ligam a ssDNA (single-stranded binding proteins, SSBs), similarmente a vírus eucarióticos do tipo RCR. Além disso, integrações de vírus eucarióticos do tipo RCR já haviam sido identificadas em genomas de eucariotos, o que demonstraria a plausibilidade do cenário proposto.

Apesar de serem possíveis, ambas as hipóteses apresentam inconsistências ou requerem a ocorrência de eventos secundários para serem explicadas. No caso da primeira hipótese (origem a partir de transposons procarióticos), o principal problema se dá pelo fato de os *Helitrons* possuírem um domínio Rep seguido de uma helicase, ao contrário dos transposons procarióticos que só codificam um domínio Rep. Para explicar esta diferença foi sugerido que ancestrais dos *Helitrons* teriam adquirido seu domínio Hel por meio da captura de uma helicase proveniente de um hospedeiro eucarioto. As principais evidências que dão suporte à esta sugestão são a presença de introns no domínio Hel de alguns *Helitrons* e o fato de que o domínio Hel pertence à família de helicases Pif1 (Kapitonov & Jurka 2001, 2007, Thomas & Pritham 2015). Helicases da família Pif1 são encontradas em praticamente todos os eucariotos, sendo responsáveis por várias funções genômicas importantes como replicação e reparo do DNA, manutenção telomérica e mitocondrial, maturação de fragmentos de Okazaki, ruptura de complexos proteína-DNA, resolução de estruturas secundárias em ácidos nucleicos, dentre outras (Boule & Zakian 2006, Bochman et al. 2010, Muellner & Schmidt 2020)

A presença de uma helicase Pif1 na transposase RepHel também é inconsistente com a segunda hipótese (origem a partir de vírus eucarióticos). Apesar de *Helitrons* se assemelharem a vírus eucarióticos do tipo RCR por codificarem uma proteína com um domínio helicase, no caso dos *Helitrons* este domínio representa uma S1H, ao contrário dos vírus eucarióticos do tipo RCR em que sua helicase pertence a superfamília 3 (S3H) (Krupovic 2013, Koonin & Dolja 2014). Esta característica também é inconsistente com o cenário adicional proposto para a primeira hipótese (*Helitrons* teriam se originado de transposons procarióticos e deram origem a vírus eucarióticos). Em todo caso, até hoje nenhuma das hipóteses apresentadas acima foi investigada em detalhe, de forma que a origem dos *Helitrons*, e dos domínios presentes em sua transposase permanecem desconhecidos.

Nota-se que o conhecimento sobre os *Helitrons* tem avançado consideravelmente nas últimas duas décadas desde a sua descoberta, principalmente no que diz respeito à sua prevalência e influência nos genomas eucarióticos e, mais recentemente, ao seu mecanismo de transposição. Apesar disso, vemos que durante este mesmo período pouco, ou quase nada, foi revelado sobre a sua origem evolutiva e sua relação com outros elementos genéticos móveis.

2. OBJETIVOS

O objetivo geral do presente trabalho consistiu em investigar a origem evolutiva dos *Helitrons* utilizando análises filogenéticas moleculares das sequências de aminoácidos dos domínios presentes em sua transposase (RepHel).

Os objetivos específicos foram:

(i) Investigar as relações evolutivas entre o domínio Rep presente nos *Helitrons* e proteínas codificadas por outros elementos genéticos móveis encontrados em procariotos e eucariotos.

(ii) Testar as duas principais hipóteses acerca da origem dos *Helitrons*, sendo a primeira a de que estes teriam se originado de transposons procarióticos, e a alternativa a de que os *Helitrons* teriam se originado de vírus eucarióticos ou seriam parentes próximos destes.

(iii) Investigar as relações evolutivas entre helicases presente nos *Helitrons* e as encontradas em diferentes organismos e elementos genéticos móveis, de forma a testar a hipótese de que os *Helitrons* teriam adquirido seu domínio Hel de um gene Pif1 eucariótico.

(iv) Utilizar os dados obtidos nas análises anteriores para propor um cenário abrangente sobre a origem e evolução dos *Helitrons*.

(v) Complementarmente, decidimos reexaminar a distribuição e a história evolutiva de uma família de *Helitrons* (Hel_c35) presente em artrópodes, identificada pelo nosso grupo em um trabalho anterior, utilizando para isso análises filogenéticas moleculares das suas sequências de nucleotídeos.

3. CAPÍTULO 1

Exploring the Remote Ties between *Helitron* Transposases and Other Rolling-Circle Replication Proteins

Pedro Heringer and Gustavo C. S. Kuhn

Departamento de Genética, Ecologia e Evolução, Instituto de Ciências Biológicas, Universidade Federal de Minas Gerais, Belo Horizonte, MG, CEP 31270-901, Brazil.

International Journal of
Molecular Sciences

MDPI

Article

Exploring the Remote Ties between Helitron Transposases and Other Rolling-Circle Replication Proteins

Pedro Heringer and Gustavo C. S. Kuhn *

Departamento de Biologia Geral, Instituto de Ciências Biológicas, Universidade Federal de Minas Gerais, Belo Horizonte, MG, CEP 31270-901, Brazil; pedrohlt@ufmg.br
* Correspondence: gcskuhn@ufmg.br; Tel.: +55-(31)-3409-3062

Received: 28 August 2018; Accepted: 7 October 2018; Published: 9 October 2018

Abstract: Rolling-circle replication (RCR) elements constitute a diverse group that includes viruses, plasmids, and transposons, present in hosts from all domains of life. Eukaryotic RCR transposons, also known as Helitrons, are found in species from all eukaryotic kingdoms, sometimes representing a large portion of their genomes. Despite the impact of Helitrons on their hosts, knowledge about their relationship with other RCR elements is still elusive. Here, we compared the endonuclease domain sequence of Helitron transposases with the corresponding region from RCR proteins found in a wide variety of mobile genetic elements. To do that, we used a stepwise alignment approach followed by phylogenetic and multidimensional scaling analyses. Although it has been suggested that Helitrons might have originated from prokaryotic transposons or eukaryotic viruses, our results indicate that Helitron transposases share more similarities with proteins from prokaryotic viruses and plasmids instead. We also provide evidence for the division of RCR endonucleases into three groups (Y1, Y2, and Yx), covering the whole diversity of this protein family. Together, these results point to prokaryotic elements as the likely closest ancestors of eukaryotic RCR transposons, and further demonstrate the fluidity that characterizes the boundaries separating viruses, plasmids, and transposons.

Keywords: Helitron; rolling-circle replication; mobile genetic element; viral evolution

1. Introduction

Rolling-circle replication (RCR) proteins are essential components of many genetic elements found in all three domains of life. These proteins can be classified into three different groups according to their main function: (i) Rep proteins (vegetative replication), (ii) Mob proteins/relaxases (conjugation), and (iii) transposases (transposon mobility) [1,2]. Helitrons are the eukaryotic representatives of RCR transposable elements (TEs), found in species from all eukaryotic kingdoms in highly variable copy numbers [3,4]. Their transposition is thought to occur by a mechanism similar to the one proposed for bacterial RCR TEs, like the IS91 family of elements [4–6]. Briefly, the Helitron transposase binds to the 5′-end of the element, using one of its two catalytic tyrosines to create a 5′-phosphotyrosine intermediate and a free 3′-OH at the donor site. The leading strand covalently bound to the transposase is displaced, the lagging strand is synthesized, and the second catalytic tyrosine nicks the 3′-end, promoting the formation of a double-strand circle intermediate. The transposase then cleaves the leading strand from the circular intermediate, but this time the second tyrosine cleaves the host's genome, forming a free 3′-OH which attacks the first 5′-phosphotyrosine linkage. After the 3′-end of the circular intermediate is also joined to the recipient's free 5′-end, an integrated single-strand "loop" is formed and probably resolved during the host's genome replication. In addition, it has been recently shown that Helitron transposition shares mechanistic similarities with the replication process used by some circular viruses [7]. Despite some of the differences in their mode of propagation, the main catalytic reaction used by all RCR elements is essentially the same [1].

Helitron transposases are composed of a typical domain, the endonuclease involved in the initiation of RCR (RCRE or Rep), fused to a helicase domain (Hel) from the superfamily 1 (S1H) (Figure 1) [4,8]. This protein, also known as RepHel, belongs to the HUH (named after one of its conserved motifs with two His residues separated by a hydrophobic residue) family of endonucleases [1]. Although HUH endonucleases from eukaryotic viruses and some plasmids also have a helicase domain, they belong to the superfamily 3 (S3H), which is unrelated to the one found in Helitrons. Furthermore, prokaryote viruses only encode a RCRE domain with no helicase (Figure 1) [8,9].

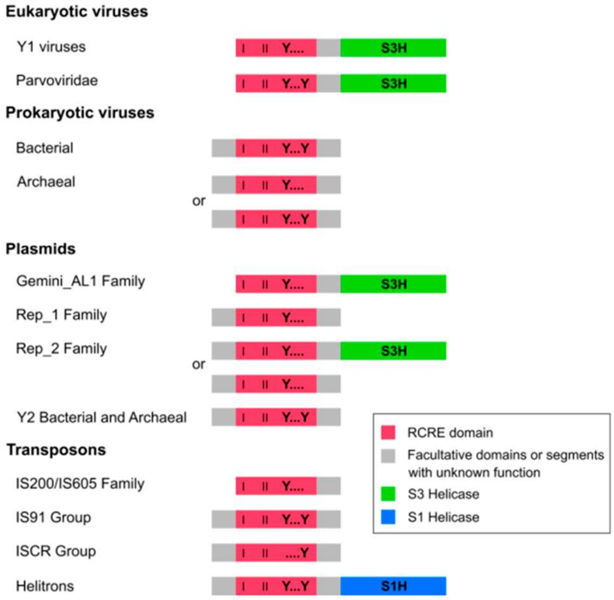


Figure 1. Modular diversity of HUH endonucleases. Schematic representation of the rolling-circle replication (RCR) proteins included in the present analysis. Rolling-circle replication endonuclease (RCRE) domains have the first two motifs (I and II), in addition to the third motif represented by one or two tyrosines (Y) in the catalytic core (dots represent variable amino acid residues). Domains are not drawn to scale, and segments after helicase domains are not represented. Based on information from Chandler et al. [1], Koonin and Dolja [8], and the Conserved Domain Database (CDD) search tool [10].

Since Helitrons were discovered [11], a few preliminary suggestions about their evolutive origins have been made. These can be generally divided in two scenarios: the first suggests that Helitrons originated from a prokaryotic ancestral RCR TE [8,11], and the second adds the possibility that Helitrons descended from an ancient eukaryotic viral integration [12]. The first scenario is mainly based on the obvious similarities in the mode of propagation of eukaryotic and prokaryotic RCR TEs, while the second scenario considers the fact that, in contrast to prokaryote RCR TEs, Helitron coding sequences include a helicase domain and sometimes a ssDNA-binding protein, similarly to some RCR proteins from eukaryotic viruses. The fact that many viral copies from geminiviruses were found to be integrated in the tobacco genome [13] was also used to support this hypothesis. In fact, since this scenario was first proposed, several studies showed copies from different eukaryotic RCR viruses in host chromosomes, revealing that viral integrations of these replicons are more common than it was previously thought (reviewed in [14]). In addition, it has been shown that several geminivirus- and parvovirus-related sequences integrated in eukaryote genomes display TE features, and have apparently shifted from a viral to a transposon-like mode of replication [15].

Despite the above considerations, some differences between the RCR proteins of Helitrons and eukaryotic viruses argue against their evolutionary relationship. Firstly, as mentioned before, helicases from these two classes of elements belong to different superfamilies. Also, with the exception of parvoviruses [16], all RCR proteins from eukaryotic ssDNA viruses contain only one tyrosine (Y1) in their catalytic core [9,17], in contrast to the RepHel from Helitrons, which has two (Y2) [4] (Figure 1). Although the number of catalytic tyrosines has been used to tentatively classify RCR proteins between two superfamilies [17], there is currently no phylogenetic support for this distinction. In view of these observations, and considering that domain rearrangements are not uncommon during protein evolution [18], the first scenario (i.e., that Helitrons originated from a prokaryotic ancestral RCR TE) seems to be more parsimonious, as the acquisition of a S1H domain would be the only major evolutionary step in a prokaryotic to eukaryotic RCR TE transition.

The relationship between Helitrons and other RCR genetic elements was initially assessed by Poulter et al. [19]. Although their results did not indicate a relationship between these TEs with specific RCR entities, they provided evidence for an ancient monophyletic origin of Helitrons, which probably occurred early on in the evolution of eukaryotes. However, the evolutionary origin of Helitrons has not been further examined, probably as a consequence of the low sequence identity of RepHel with any other group of RCR proteins [3].

In this study, we investigated the relationship of the Helitron RepHel with other RCR proteins by analyzing the RCRE amino acid sequences from a wide variety of mobile genetic elements, including TEs, plasmids, and viruses. Our results indicate that, despite being eukaryotic TEs, Helitron transposases display more sequence similarities with prokaryotic RCR proteins from bacteriophages and plasmids. In addition, we show that the HUH family of endonucleases can be divided into three major phylogenetic groups comprised of RCR proteins from highly heterogeneous mobile genetic elements.

2. Results and Discussion

2.1. Selecting and Preparing RCRE Domain Sequences

We selected a sample of 13 Helitron RepHel amino acid sequences, representing elements from distantly-related organisms across several phyla and including the main Helitron variants (Table S1). To analyze these TEs in a broad evolutionary context, at least three sequences of each family or group of RCR genetic elements from prokaryotes and eukaryotes were selected. These included single- and double-stranded viruses, plasmids, and TEs (Table S1).

Our analysis was restricted to the RCRE (or HUH) domain of the sequences (Figure 1), which has a central role in starting RCR reactions and is the only region common to all HUH endonucleases [1] (Figure 1). Modular rearrangements often occur during protein evolution [18] which is also the case for several RCR virus lineages [20]. For those reasons, and considering that flanking domains are highly variable amongst RCR elements [1], our restriction to the RCRE domain aimed to avoid spurious evolutionary inferences. Most proteins within the HUH family have three conserved motifs (I, II, and III) in the core region of the RCRE domain, despite the high sequence divergence between groups [1,2,21]. Only amino acid sequences containing all three conserved motifs in their typical arrangement (I-II-III) were selected for our analysis; this is because some HUH endonucleases display their motifs in the reverse order (e.g., III-II-I) [1,2], and these also have highly divergent amino acid sequences, which prevent reliable sequence alignments. A total of 115 amino acid sequences, representing the overall diversity of all known HUH endonucleases, were selected for the analysis (Table S1).

To reduce spurious alignments of the RCRE sequences, we conducted a stepwise alignment approach, which consisted of aligning each group of closely-related sequences separately, excluding segments flanking the RCRE domain and trimming the portions that were exclusive of individual taxa. The resulting sequences (Data S1) were aligned using PSI-Coffee, which is a method considered suitable for highly divergent protein sequences with little or no structural information available [22,23].

2.2. Major RCR Protein Phylogenetic Groups

A phylogenetic analysis was conducted and pairwise divergence values between sequences were used to generate non-metric multidimensional scaling (NMDS) ordinations. As expected for an analysis that includes highly divergent sequences, clade support values between major groups were low, although we observed an overall agreement between our results and the known topology for most of the clades (Figure 2). Our results support the monophyletic nature of all Helitron variants and the lack of any clear relationship of these TEs with other specific groups or families of mobile genetic elements, as previously suggested [19]. Nonetheless, in both the phylogenetic analysis (Figure 2) and NMDS ordinations (Figure 3) we observed an overall distinction between Y1 and Y2 RCR proteins, which we henceforth refer to as Y1 and Y2 groups. An exception is a third clade, composed of elements from both variants, which we refer to here as the Yx group because the number of tyrosines of the catalytic core of its members does not relate with the canonical Y1 and Y2 division. Although the resulting phylogeny revealed a basal segregation of Yx RCR proteins and the rest of the sequences, the Y2 group appears to be more closely related to Y1 RCR proteins, and perhaps constitutes a derivative clade of the Y1 group (Figures 2 and 3).

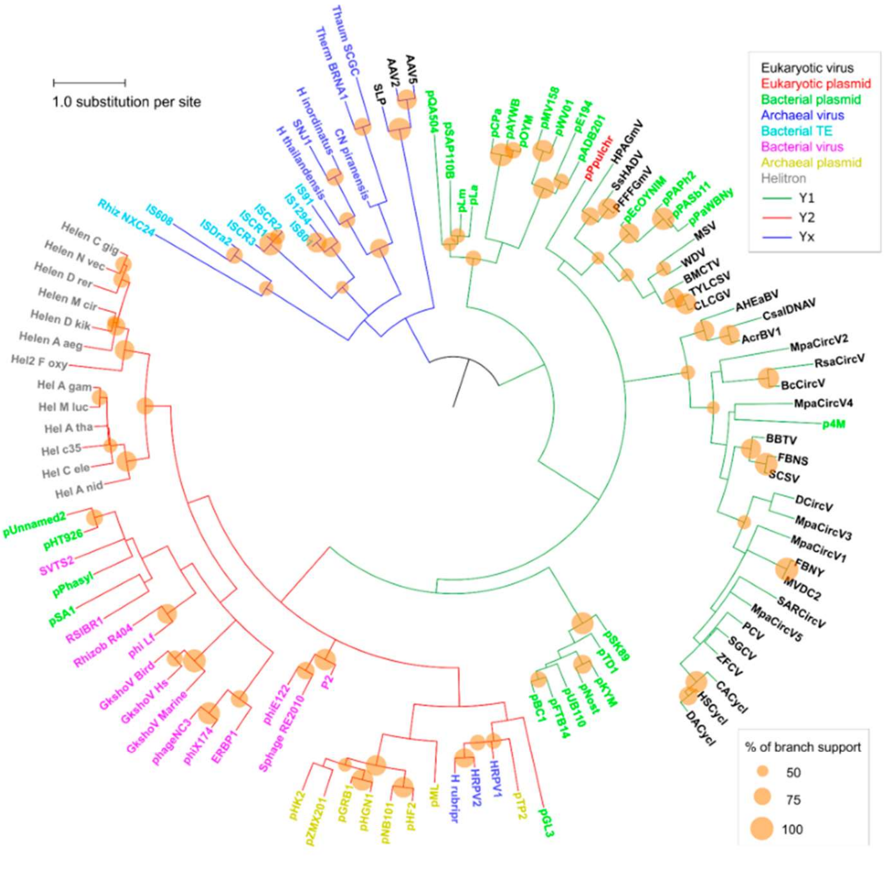


Figure 2. Phylogenetic analysis of RCRE domain sequences. Clade colors indicate each tyrosine group: Y1 (green), Y2 (red), and Yx (blue). Taxa colors represent the family of each element (box on the upper right). See Table S1 for taxa information. Phylogeny inferred by the Maximum Likelihood method (LG+G+I). The same phylogeny, with the numerical support values represented, is shown on Figure S1.

The topology observed within the Yx group is roughly in agreement with previous results [24], indicating that this clade represents a bona fide phylogenetic cluster composed of archaeal viruses and bacterial TEs. Recent analyses using different methods have also shown that parvoviruses belong to a separate clade from other eukaryotic RCR viruses [25]. However, we did not expect that parvoviral RCR proteins (AAV2, AAV5, and SLP) would group together with Yx elements (Figures 2 and 3). Although structural similarities indicate a distant relationship between parvoviral and other RCR proteins [26], the positioning of these viruses in the Yx group might also be the consequence of long branch attraction [27], so this result should be treated with caution.

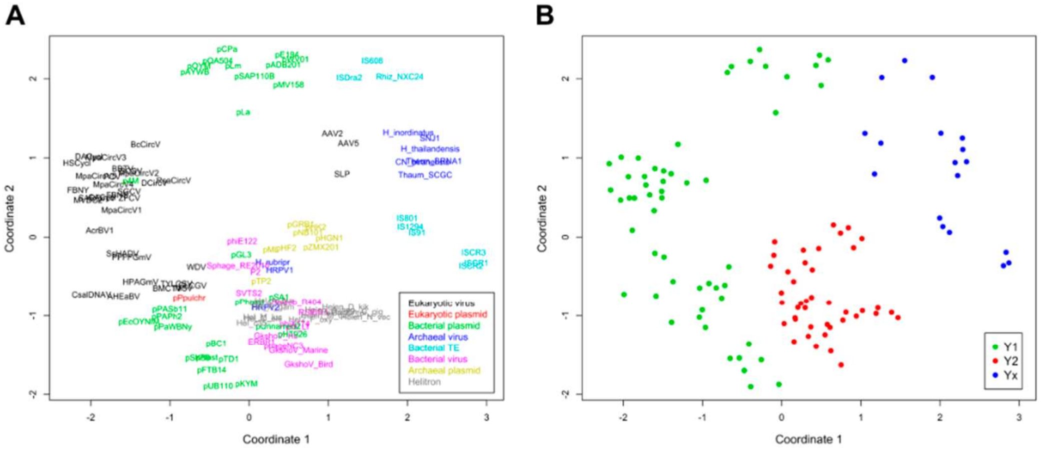


Figure 3. Non-metric multidimensional scaling (NMDS) of evolutionary divergence between RCRE domains. (**A**) Ordinations with taxa represented by their sequence abbreviations. Colors indicate the different classes of mobile genetic elements. (**B**) Same ordinations of (**A**), with colors indicating the tyrosine group of each taxa. The scaling represents euclidean distances for two dimensions (stress: 0.26382).

As revealed by the results from both analyses, the assignment to a specific catalytic tyrosine group is not contingent on the element class (Figures 2 and 3). For instance, bacterial plasmids, and eukaryotic and archaeal viruses have members in more than one group. Likewise, the element class does not always predict its topology, even within the same tyrosine group. For example, some Y1 viral families are closer to Y1 plasmids than other Y1 viruses, and the same is true in the Y2 group. This phenomenon has been observed in different studies and emphasizes the marked fluidity at the boundaries separating different classes of mobile genetic elements (reviewed in [8,9]). Thus, our results indicate that the tyrosine group division is the only informative phylogenetic feature encompassing the whole HUH endunuclease family.

2.3. Helitron Transposase is More Similar to Prokaryotic Proteins

Even though the Helitron RepHel does not appear to be phylogenetically closer to any single family of proteins, they clustered within the Y2 group which, apart from Helitrons, is exclusively composed of prokaryotic viruses and plasmids (Figures 2 and 3). On the other hand, sequences from prokaryotic TEs clustered within the Yx group, even though some of them (including the IS91 family) have two tyrosines in their catalytic core and share a similar transposition mechanism with Helitrons [4–6,28]. It is also notable that RepHel proteins appear to be only distantly related to RCR proteins from eukaryotic viruses, which almost exclusively belong to the Y1 group. These observations indicate that the core domain from Helitron transposases is more similar to proteins from prokaryotic viruses and plasmids than to prokaryotic RCR transposases or to eukaryotic viral proteins.

As we have mentioned, in addition to the RCRE domain, RepHel proteins also have a S1 helicase domain (Figure 1); more specifically, this S1 helicase belongs to the Pif1 family [4]. Although

Pif1 helicases are present in essentially all eukaryote genomes, they also have been found in some prokaryotes [29,30]. Because all known prokaryotic Y2 RCR proteins lack a helicase, this domain could have been acquired from a prokaryote host by the Helitron ancestor before it colonized the first eukaryote genome. However, considering that Pif1 helicases are ubiquitous in eukaryote genomes and found less frequently in prokaryotes, it seems more plausible that Helitrons acquired their helicase domain from a eukaryotic host. Indeed, a preliminary analysis of Pif1 sequences from Helitrons, eukaryotes, and prokaryotes indicates that the helicase domain from Helitrons is closely related to fungal proteins (Figure S2). Interestingly, the helicase domains from distinct Helitron variants formed separate clusters with different fungal proteins, suggesting that Helitrons acquired their helicase domain from at least two independent events (Figure S2).

These results support the hypothesis of an ancient origin of Helitrons during the initial radiation of eukaryotes, and suggest that neither prokaryotic TEs, nor eukaryotic viruses, are among their closest relatives. Instead, we provide evidence for a closer relationship of these eukaryotic TEs with prokaryotic viruses and plasmids with Y2 RCR proteins, even though it is not possible to determine which specific family shares the most recent common ancestor with the RepHel (Figure 4). Thus, our proposition is that Helitrons descend from a prokaryotic Y2 mobile element that integrated in the genome of an early eukaryote ancestor. Like all other known prokaryotic Y2 elements, the Helitron progenitor probably coded an RCR protein devoid of a helicase domain and was dependent of its host for correct replication/transposition. Subsequently, each of the incipient Helitron variants acquired a eukaryotic helicase by the recombination of its RCRE domain with a host helicase gene. In any case, a comprehensive understanding of the Helitron origins will probably rely on the future discovery of new groups of RCR genetic elements.

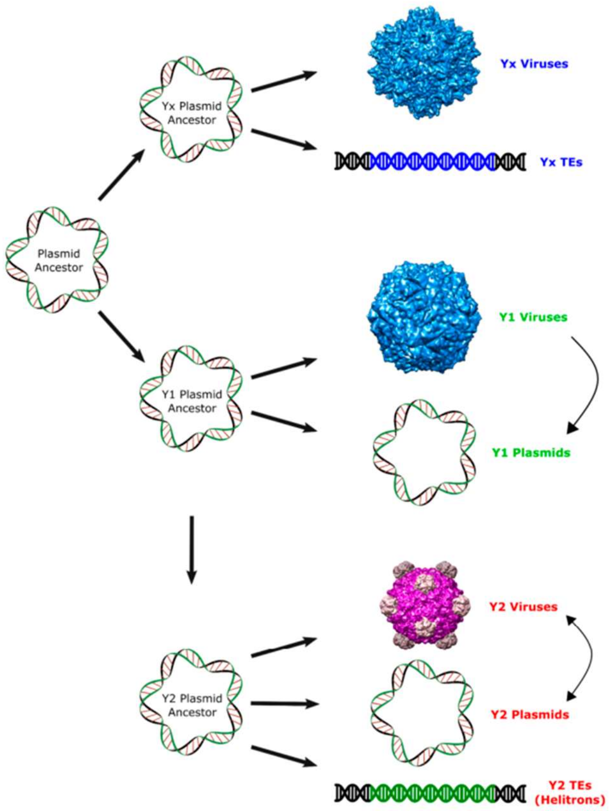


Figure 4. Proposed scenario for the origin of Helitrons and other RCR elements. Arrows represent putative pathways to explain the observed relationship among RCR elements. Virion images were obtained from VIPERdb (http://viperdb.scripps.edu) [31].

Finally, although the RCRE phylogeny does not coincide with the taxonomic division of distinct genetic elements classes (viruses, plasmids and TEs), we suggest that the HUH family of endonucleases is composed by three major radiation groups (Y1, Y2 and Yx). Interestingly, most of the HUH endonucleases can be assigned to one of these groups simply by having a tyrosine residue at a specific position in the RCRE domain, regardless of the element's class. The extreme diversity observed in each of these groups underscore the dynamic nature of mobile genetic elements which, in the long term, do not evolve under the usual taxonomic constraints acting upon their hosts.

3. Materials and Methods

3.1. Sequences Retrieval and Selection

RepHel amino acid sequences from Helitrons were retrieved from Repbase (https://www.girinst.org/repbase/) [32] and GenBank (https://www.ncbi.nlm.nih.gov/genbank/) [33], using elements from previous studies as a reference (e.g., [11,19,34]). The structure of these proteins was verified using the Conserved Domain Database (CDD) search tool (https://www.ncbi.nlm.nih.gov/Structure/cdd/wrpsb.cgi) [10]. RepHel sequences that could be clearly assigned to one of the three main Helitron variants [4] were selected: canonical Helitron (6 sequences), Helitron2 (1 sequence), and Helentron (6 sequences). Sequences representing each family or group of RCR proteins were retrieved on GenBank [33], based on several references (e.g., [9,21,24,35–37]). A total of 115 amino acid sequences were selected for the alignment (Table S1).

3.2. Sequence Alignment

Each family or group of sequences were aligned separately using the M-Coffee mode from T-Coffee (http://tcoffee.crg.cat/) [22] before being manually trimmed in order to exclude flanking portions of the RCRE domain and the segments that are exclusive of individual taxa. The trimmed sequences (Data S1) were aligned with PSI-Coffee (http://tcoffee.crg.cat/apps/tcoffee/do:psicoffee) [22] before manual correction. Alignment positions with less than 90% coverage were excluded.

3.3. NMDS and Phylogenetic Analysis

Pairwise evolutionary divergence between sequences was estimated using the Poisson correction model on MEGA7 [38]. The values were used to generate non-metric multidimensional scaling (NMDS) ordinations with the R package vegan [39], representing euclidean distances for two dimensions. NMDS and plotting of ordinations were conducted in RStudio v1.1.442 (Boston, MA, USA) [40]. The best-fit evolutionary model for the alignment (LG+G+I) was determined using MEGA7 [38] and the Smart model selection (SMS) in PhyML (http://www.atgc-montpellier.fr/phyml/) [41]. Maximum Likelihood phylogeny was inferred from 5000 replicates using MEGA7 [38], and the final phylogenetic tree edited using iTOL v4.2.3 (https://itol.embl.de/) [42].

Supplementary Materials: Supplementary materials can be found at http://www.mdpi.com/1422-0067/19/10/3079/s1.

Author Contributions: Conceptualization, P.H. and G.C.S.K.; Methodology, P.H.; Investigation, P.H.; Resources, G.C.S.K.; Data Curation, P.H.; Writing-Original Draft Preparation, P.H.; Writing-Review & Editing, G.C.S.K.; Visualization, P.H.; Funding Acquisition, G.C.S.K.

Funding: This research was funded by Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq) grant number 404620/2016-7 to G.C.S.K and doctoral fellowship from Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES) to P.H.

Acknowledgments: We thank Renan P. de Souza for the insightful discussions that helped to enrich our analysis and "Pró-Reitoria de Pesquisa da Universidade Federal de Minas Gerais" for partially covering the publication costs.

Conflicts of Interest: The authors declare no conflict of interest.

Abbreviations

RCR	Rolling-circle replication
TE	Transposable element
RCRE	Rolling-circle replication endonuclease domain
S1H	Superfamily 1 helicase
S3H	Superfamily 3 helicase
RepHel	Helitron transposase (Rep/Helicase)
ssDNA	Single-strand DNA
NMDS	Non-metric multidimensional scaling

References

1. Chandler, M.; De La Cruz, F.; Dyda, F.; Hickman, A.B.; Moncalian, G.; Ton-Hoang, B. Breaking and joining single-stranded DNA: The HUH endonuclease superfamily. *Nat. Rev. Microbiol.* **2013**, *11*, 525–538. [CrossRef] [PubMed]
2. Wawrzyniak, P.; Płucienniczak, G.; Bartosik, D. The Different Faces of Rolling-Circle Replication and Its Multifunctional Initiator Proteins. *Front. Microbiol.* **2017**, *8*, 2353. [CrossRef] [PubMed]
3. Kapitonov, V.V.; Jurka, J. Helitrons on a roll: Eukaryotic rolling-circle transposons. *Trends Genet.* **2007**, *23*, 521–529. [CrossRef] [PubMed]
4. Thomas, J.; Pritham, E.J. *Helitrons, the Eukaryotic Rolling-Circle Transposable Elements in Mobile DNA III*, 3rd ed.; ASM Press: Washington, DC, USA, 2015; pp. 893–926.
5. Dias, G.B.; Heringer, P.; Kuhn, G.C. Helitrons in *Drosophila*: Chromatin modulation and tandem insertions. *Mob. Genet. Elements* **2016**, *6*, e1154638. [CrossRef] [PubMed]
6. Grabundzija, I.; Messing, S.A.; Thomas, J.; Cosby, R.L.; Bilic, I.; Miskey, C.; Jurka, J. A Helitron transposon reconstructed from bats reveals a novel mechanism of genome shuffling in eukaryotes. *Nat. Commun.* **2016**, *7*, 10716. [CrossRef] [PubMed]
7. Grabundzija, I.; Hickman, A.B.; Dyda, F. Helraiser intermediates provide insight into the mechanism of eukaryotic replicative transposition. *Nat. Commun.* **2018**, *9*, 1278. [CrossRef] [PubMed]
8. Koonin, E.V.; Dolja, V.V. Virus world as an evolutionary network of viruses and capsidless selfish elements. *Microbiol. Mol. Biol. Rev.* **2014**, *78*, 278–303. [CrossRef] [PubMed]
9. Krupovic, M. Networks of evolutionary interactions underlying the polyphyletic origin of ssDNA viruses. *Curr. Opin. Virol.* **2013**, *3*, 578–586. [CrossRef] [PubMed]
10. Marchler-Bauer, A.; Bo, Y.; Han, L.; He, J.; Lanczycki, C.J.; Lu, S.; Chitsaz, F.; Derbyshire, M.K.; Geer, R.C.; Gonzales, N.R.; et al. CDD/SPARCLE: Functional classification of proteins via subfamily domain architectures. *Nucleic Acids Res.* **2017**, *45*, D200–D203. [CrossRef] [PubMed]
11. Kapitonov, V.V.; Jurka, J. Rolling-circle transposons in eukaryotes. *Proc. Natl. Acad. Sci. USA* **2001**, *98*, 8714–8719. [CrossRef] [PubMed]
12. Feschotte, C.; Wessler, S.R. Treasures in the attic: Rolling circle transposons discovered in eukaryotic genomes. *Proc. Natl. Acad. Sci. USA* **2001**, *98*, 8923–8924. [CrossRef] [PubMed]
13. Bejarano, E.R.; Khashoggi, A.; Witty, M.; Lichtenstein, C. Integration of multiple repeats of geminiviral DNA into the nuclear genome of tobacco during evolution. *Proc. Natl. Acad. Sci. USA* **1996**, *93*, 759–764. [CrossRef] [PubMed]
14. Krupovic, M.; Forterre, P. Single-stranded DNA viruses employ a variety of mechanisms for integration into host genomes. *Ann. N. Y. Acad. Sci.* **2015**, *1341*, 41–53. [CrossRef] [PubMed]
15. Liu, H.; Fu, Y.; Li, B.; Yu, X.; Xie, J.; Cheng, J.; Ghabrial, S.A.; Li, G.; Yi, X.; Jiang, D. Widespread horizontal gene transfer from circular single-stranded DNA viruses to eukaryotic genomes. *BMC Evol. Biol.* **2011**, *11*, 276. [CrossRef] [PubMed]
16. Hickman, A.B.; Ronning, D.R.; Kotin, R.M.; Dyda, F. Structural unity among viral origin binding proteins: Crystal structure of the nuclease domain of adeno-associated virus Rep. *Mol. Cell* **2002**, *10*, 327–337. [CrossRef]
17. Rosario, K.; Duffy, S.; Breitbart, M. A field guide to eukaryotic circular single-stranded DNA viruses: Insights gained from metagenomics. *Arch. Virol.* **2012**, *157*, 1851–1871. [CrossRef] [PubMed]

18. Björklund, Å.K.; Ekman, D.; Light, S.; Frey-Skött, J.; Elofsson, A. Domain rearrangements in protein evolution. *J. Mol. Biol.* **2005**, *353*, 911–923. [CrossRef] [PubMed]
19. Poulter, R.T.; Goodwin, T.J.; Butler, M. Vertebrate helentrons and other novel Helitrons. *Gene* **2003**, *313*, 201–212. [CrossRef]
20. Kazlauskas, D.; Varsani, A.; Krupovic, M. Pervasive Chimerism in the Replication-Associated Proteins of Uncultured Single-Stranded DNA Viruses. *Viruses* **2018**, *10*, 187. [CrossRef] [PubMed]
21. Ilyina, T.V.; Koonin, E.V. Conserved sequence motifs in the initiator proteins for rolling circle DNA replication encoded by diverse replicons from eubacteria, eucaryotes and archaebacteria. *Nucleic Acids Res.* **1992**, *20*, 3279–3285. [CrossRef] [PubMed]
22. Di Tommaso, P.; Moretti, S.; Xenarios, I.; Orobitg, M.; Montanyola, A.; Chang, J.M.; Notredame, C. T-Coffee: A web server for the multiple sequence alignment of protein and RNA sequences using structural information and homology extension. *Nucleic Acids Res.* **2011**, *39*, W13–W17. [CrossRef] [PubMed]
23. Taly, J.F.; Magis, C.; Bussotti, G.; Chang, J.M.; Di Tommaso, P.; Erb, I.; Espinosa-Carrasco, J.; Kemena, C.; Notredame, C. Using the T-Coffee package to build multiple sequence alignments of protein, RNA, DNA sequences and 3D structures. *Nat. Protoc.* **2011**, *6*, 1669–1682. [CrossRef] [PubMed]
24. Wang, Y.; Chen, B.; Cao, M.; Sima, L.; Prangishvili, D.; Chen, X.; Krupovic, M. Rolling-circle replication initiation protein of haloarchaeal sphaerolipovirus SNJ1 is homologous to bacterial transposases of the IS91 family insertion sequences. *J. Gen. Virol.* **2018**, *99*, 416–421. [CrossRef] [PubMed]
25. Aiewsakun, P.; Simmonds, P. The genomic underpinnings of eukaryotic virus taxonomy: Creating a sequence-based framework for family-level virus classification. *Microbiome* **2018**, *6*, 38. [CrossRef] [PubMed]
26. Campos-Olivas, R.; Louis, J.M.; Clérot, D.; Gronenborn, B.; Gronenborn, A.M. The structure of a replication initiator unites diverse aspects of nucleic acid metabolism. *Proc. Natl. Acad. Sci. USA* **2002**, *99*, 10310–10315. [CrossRef] [PubMed]
27. Bergsten, J. A review of long-branch attraction. *Cladistics* **2005**, *21*, 163–193. [CrossRef]
28. Garcillan-Barcia, M.P.; Bernales, I.; Mendiola, M.V.; de la Cruz, F. Single-stranded DNA intermediates in IS91 rolling-circle transposition. *Mol. Microbiol.* **2001**, *39*, 494–501. [CrossRef]
29. Bochman, M.L.; Sabouri, N.; Zakian, V.A. Unwinding the functions of the Pif1 family helicases. *DNA Repair* **2010**, *9*, 237–249. [CrossRef] [PubMed]
30. Bochman, M.L.; Judge, C.P.; Zakian, V.A. The Pif1 family in prokaryotes: What are our helicases doing in your bacteria? *Mol. Biol. Cell* **2011**, *22*, 1955–1959. [CrossRef] [PubMed]
31. Carrillo-Tripp, M.; Shepherd, C.M.; Borelli, I.A.; Venkataraman, S.; Lander, G.; Natarajan, P.; Johnson, J.E.; Brooks, C.L.; Reddy, V.S. VIPERdb2: An enhanced and web API enabled relational database for structural virology. *Nucleic Acids Res.* **2009**, *37*, D436–D442. [CrossRef] [PubMed]
32. Bao, W.; Kojima, K.K.; Kohany, O. Repbase Update, a database of repetitive elements in eukaryotic genomes. *Mob. DNA* **2015**, *6*, 11. [CrossRef] [PubMed]
33. Benson, D.A.; Cavanaugh, M.; Clark, K.; Karsch-Mizrachi, I.; Lipman, D.J.; Ostell, J.; Sayers, E.W. GenBank. *Nucleic Acids Res.* **2017**, *45*, D37–D42. [CrossRef] [PubMed]
34. Pritham, E.J.; Feschotte, C. Massive amplification of rolling-circle transposons in the lineage of the bat Myotis lucifugus. *Proc. Natl. Acad. Sci. USA* **2007**, *104*, 1895–1900. [CrossRef] [PubMed]
35. Zawar-Reza, P.; Argüello-Astorga, G.R.; Kraberger, S.; Julian, L.; Stainton, D.; Broady, P.A.; Varsani, A. Diverse small circular single-stranded DNA viruses identified in a freshwater pond on the McMurdo Ice Shelf (Antarctica). *Infect. Genet. Evol.* **2014**, *26*, 132–138. [CrossRef] [PubMed]
36. Kazlauskas, D.; Dayaram, A.; Kraberger, S.; Goldstien, S.; Varsani, A.; Krupovic, M. Evolutionary history of ssDNA bacilladnaviruses features horizontal acquisition of the capsid gene from ssRNA nodaviruses. *Virology* **2017**, *504*, 114–121. [CrossRef] [PubMed]
37. Wang, Y.; Sima, L.; Lv, J.; Huang, S.; Liu, Y.; Wang, J.; Krupovic, M.; Chen, X. Identification, characterization, and application of the replicon region of the halophilic temperate sphaerolipovirus SNJ1. *J. Bacteriol.* **2016**, *198*, 1952–1964. [CrossRef] [PubMed]
38. Kumar, S.; Stecher, G.; Tamura, K. MEGA7: Molecular evolutionary genetics analysis version 7.0 for bigger datasets. *Mol. Biol. Evol.* **2016**, *33*, 1870–1874. [CrossRef] [PubMed]
39. Dixon, P. VEGAN, a package of R functions for community ecology. *J. Veg. Sci.* **2003**, *14*, 927–930. [CrossRef]
40. RStudio Team. *RStudio: Integrated Development for R*; RStudio, Inc.: Boston, MA, USA, 2016; Available online: http://www.rstudio.com/ (accessed on 8 October 2018).

41. Lefort, V.; Longueville, J.E.; Gascuel, O. SMS: Smart model selection in PhyML. *Mol. Biol. Evol.* **2017**, *34*, 2422–2424. [CrossRef] [PubMed]

42. Letunic, I.; Bork, P. Interactive tree of life (iTOL) v3: An online tool for the display and annotation of phylogenetic and other trees. *Nucleic Acids Res.* **2016**, *44*, W242–W245. [CrossRef] [PubMed]

4. CAPÍTULO 2

Pif1 Helicases and the Evidence for a Prokaryotic Origin of *Helitrons*

Pedro Heringer and Gustavo C. S. Kuhn

Departamento de Genética, Ecologia e Evolução, Instituto de Ciências Biológicas,
Universidade Federal de Minas Gerais, Belo Horizonte, MG, CEP 31270-901, Brazil.

Pif1 Helicases and the Evidence for a Prokaryotic Origin of *Helitrons*

Pedro Heringer and Gustavo C.S. Kuhn*

Departamento de Genética, Ecologia e Evolução, Instituto de Ciências Biológicas, Universidade Federal de Minas Gerais, Belo Horizonte, Brazil

***Corresponding author:** E-mail: gcskuhn@ufmg.br.

Associate editor: Irina Arkhipova

Abstract

***Helitrons* are the only group of rolling-circle transposons that encode a transposase with a helicase domain (Hel), which belongs to the Pif1 family. Because Pif1 helicases are important components of eukaryotic genomes, it has been suggested that Hel domains probably originated after a host eukaryotic Pif1 gene was captured by a *Helitron* ancestor. However, the few analyses exploring the evolution of *Helitron* transposases (RepHel) have focused on its Rep domain, which is also present in other mobile genetic elements. Here, we used phylogenetic and nonmetric multidimensional scaling analyses to investigate the relationship between Hel domains and Pif1-like helicases from a variety of organisms. Our results reveal that Hel domains are only distantly related to genomic helicases from eukaryotes and prokaryotes, and thus are unlikely to have originated from a captured Pif1 gene. Based on this evidence, and on recent studies indicating that Rep domains are more closely related to rolling-circle plasmids and phages, we suggest that *Helitrons* are descendants of a RepHel-encoding prokaryotic plasmid element that invaded eukaryotic genomes before the radiation of its major groups. We discuss how a Pif1-like helicase domain might have favored the transposition of *Helitrons* in eukaryotes beyond simply unwinding DNA intermediates. Finally, we demonstrate that some examples in the literature describing genomic helicases from eukaryotes actually consist of Hel domains from *Helitrons*, a finding that underscores how transposons can hamper the analysis of eukaryotic genes. This investigation also revealed that two groups of land plants appear to have lost genomic Pif1 helicases independently.**

***Key words: Helitrons*, transposon, Pif1, helicase.**

Introduction

Helitrons are DNA transposable elements (TEs) found in a wide variety of species from all eukaryotic kingdoms but make up variable genomic proportions across different taxa. For instance, they constitute between 0.1% and 6.6% of the genomic DNA in plants and between 0% and 10% in animals (reviewed in Kapitonov and Jurka [2007] and Thomas and Pritham [2015]). These TEs have been shown to mobilize within a genome by a process known as rolling-circle (RC) transposition (RCT) (Grabundzija et al. 2016, 2018) which could be viewed as a variation of the RC replication (RCR) process employed by several groups of plasmids and viruses from prokaryotes and eukaryotes (reviewed in Chandler et al. [2013] and Wawrzyniak et al. [2017]). In *Helitrons*, the RCT is executed by the Rep/Helicase (RepHel) transposase, which is composed by two major domains: an endonuclease (Rep) domain and a superfamily 1 helicase (Hel) domain (Thomas and Pritham 2015) (fig. 1).

Helitrons can be classified into four structural and coding variants, namely *Helitron*, *Helentron*, *Helitron2*, and *Proto-Helentron* (Thomas and Pritham 2015). In contrast to the first three variants, which have been shown to represent distinct phylogenetic groups (Poulter et al. 2003; Thomas et al. 2014; Heringer and Kuhn 2018), *Proto-Helentron* elements seem to constitute a subtype of *Helentrons* with derived *Helitron*-like structural features (Thomas et al. 2014). Although all *Helitrons* have RepHel proteins with two major domains, distinct variants, or specific variant lineages, can encode additional domains in their transposase or/and additional genes. Likewise, specific sets of structural features, like inverted repeats, can be used to identify major lineages or variants (fig. 1).

The Hel domain present in *Helitron* transposases is a superfamily 1 helicase, more specifically from the Pif1 family (Kapitonov and Jurka 2001; Thomas and Pritham 2015). Pif1 helicases have been found in essentially all eukaryotes studied to date (Bochman et al. 2010) and are involved in several processes, like DNA replication and repair, telomere maintenance, Okazaki fragment maturation, disruption of protein–DNA complexes, resolution of nucleic acid secondary structures, mitochondrial DNA maintenance, among others (reviewed in Boule and Zakian [2006]; Bochman et al. [2010]; and Muellner and Schmidt [2020]). Although typically known as eukaryotic proteins, Pif1-like helicases can also be found in some prokaryotic species, bacteriophages, and eukaryotic viruses (Bochman et al. 2011). We henceforth refer to eukaryotic and prokaryotic proteins that perform genomic-related tasks as genomic Pif1 helicases, in order to distinguish them from Pif1-like viral helicases or Hel domains found in *Helitron* transposases.

The structural and mechanistic similarities between eukaryotic and prokaryotic RC transposons initially prompted the hypothesis that *Helitrons* could be descendants of bacterial elements (e.g., IS91 family). Furthermore, it was suggested that *Helitron* ancestors could have given rise to eukaryotic RCR viruses, as these viruses were only found in plant species at that time (Kapitonov and Jurka 2001). Conversely, because geminiviruses had been found integrated into plant chromosomes, it was also proposed that *Helitrons* could likewise be derived from an ancient genomic integration of a eukaryotic RCR virus (Feschotte and Wessler 2001). However, as revealed by recent findings, Rep domains from *Helitrons* are distantly related to proteins from prokaryotic TEs and eukaryotic viruses, and share more similarities with RCR plasmids and viruses from bacteria (Heringer and Kuhn 2018; Kazlauskas et al. 2019). In spite of these similarities, the prokaryotic plasmid and viral elements which are more closely related to *Helitrons* do not encode a helicase domain (Heringer and Kuhn 2018), what makes the origin of Hel domains a still unsolved issue. The absence of helicases on the coding sequences of prokaryotic RC TEs, together with the presence of introns in some Hel domains from plants and *Caenorhabditis elegans Helitrons*, have been considered as tentative evidences that a *Helitron* ancestor acquired its Hel domain by capturing a helicase gene from its eukaryotic host (Kapitonov and Jurka 2001, 2007; Thomas and Pritham 2015). However, we still lack information about the evolutionary origins of *Helitron* Hel domains and their relationship with other helicases, as these issues have never been investigated in detail.

The fact that Pif1 family helicases are present in virtually all eukaryotes but absent in RC mobile genetic elements (MGEs), except *Helitrons*, renders the investigation about the origin of Hel domains more difficult. Moreover, to our knowledge there are no automated methods to clearly distinguish genomic Pif1 helicases from *Helitron* Pif1-like helicases. Regarding the later issue, both genomic and *Helitron* Pif1-like sequences can be found in eukaryotic genomes and sometimes is not possible to discriminate them without a more detailed analysis. For instance, Blastp searches on eukaryotic genomes using Pif1 proteins as queries often result in multiple significant hits, even though most eukaryotic species apparently have only one or two genomic Pif1 helicases (Bochman et al. 2010). Therefore, although up to few hits are expected to represent genomic Pif1 helicases in eukaryotic species, most of them often constitute *Helitron* Pif1-like protein sequences. In addition, some eukaryotes apparently have multiple genomic Pif1 paralogs (Bochman et al. 2010, 2011; Harman and Manna 2016), which makes their distinction from *Helitron* Pif1-like helicases even more complex.

In the present study, we retrieved prokaryotic, eukaryotic and viral Pif1-like proteins in silico using a stepwise searching method to avoid classifying *Helitron* coding sequences as genomic helicases. After doing so, we were able to investigate the relationship between Hel domains and Pif1-like genes from a wide variety of organisms and MGEs. Our results reveal further valuable information about the evolution of RepHel transposases, indicating that Hel domains are only distantly related to genomic Pif1 helicases and were likely present in *Helitrons* before they invaded eukaryotic hosts. We discuss the general implications of our findings considering the known mechanistic features of RepHel transposases and Pif1 helicases, also demonstrating how the similarities between these proteins can interfere with their classification and analysis.

Results

Finding Genomic Helicases

Before conducting searches to retrieve genomic Pif1-like helicases, we first expanded our sample of *Helitrons* from different variants (*Helitrons*, *Helentrons*, and *Helitron2*) selected previously (Heringer and Kuhn 2018). Consensus sequences from the helicase domains (Hel) found in those Helitrons were used as queries to obtain Pif1-like helicases from a wide diversity of organisms (see Materials and Methods). Because *Helitrons* are found throughout a large portion of eukaryotic genomes, the distinction between genomic Pif1 and *Helitron* Pif1-like helicases (Hel domains) across individual species is highly prone to identification errors (supplementary fig. S1, Supplementary Material online). For that reason, we initially selected only organisms lacking *Helitron* Rep sequences in their genomes, so that genomic Pif1 helicases could be correctly identified before our analyses. *Helitron* Rep sequences can be used as unique identifiers for the presence of *Helitrons* as they are exclusive of these RC elements and do not have genomic counterparts in eukaryotes.

The larger or smaller representation of specific taxonomic groups in the Pif1 helicases selected initially, depended on the number of available genomes and on the presence or absence of *Helitrons* in each taxon. For instance, although our searches on Embryophyta (land plants) revealed the presence of Pif1-like proteins in most species, only the common liverwort *Marchantia polymorpha* was devoid of Rep sequences from *Helitrons*, thus being the single representative of land plants selected in the first round of searches.

Although almost all retrieved sequences from prokaryotes and eukaryotes were annotated as genomic Pif1 helicases, one of the hits from the searches on archaea was a TraA relaxase annotated as belonging to a species from the *Methanothrix* genus (*Methanothrix* sp., accession number: TFH49976.1). This hit displays a relatively low sequence coverage (62%) and identity (24%) to the query (*Helentron* Hel consensus) (supplementary data S1, Supplementary Material online). Nevertheless, as TraA relaxases constitute a group of proteins involved in conjugation of bacterial plasmids and are also known to have a helicase domain (Alt-Mörbe et al. 1996; Pérez-Mendoza et al. 2006), we decided to include additional TraA relaxase representatives in our analysis. To do that, the *Methanothrix* TraA relaxase (TFH49976.1) was used as query

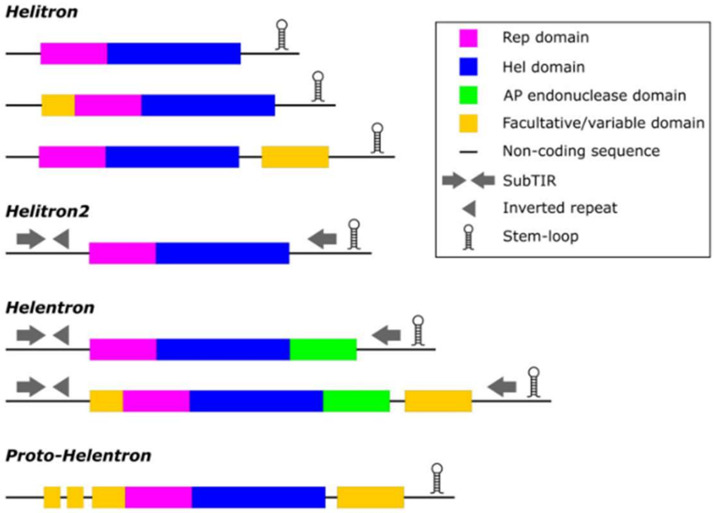


Fig. 1. *Helitron* structural and coding variants. Each variant can be identified by a set of structural (symbols) and coding sequences (colored boxes). *Helitrons*, *Helitron2*, and *Helentrons* are major phylogenetic variants, with *Proto-Helentrons* representing an internal group of *Helentrons* that have intermediate features found in *Helitrons* and *Helentrons*. Adapted from Thomas and Pritham (2015).

in Blastp searches on the nonredundant protein sequences (nr) database from GenBank (Sayers et al. 2019). Interestingly, the best hits from this search consisted of TraA sequences from the phylum Proteobacteria (supplementary table S1, Supplementary Material online), with no hits from archaeal species, indicating that TFH49976.1 could either represent a horizontally transferred gene (from a bacterium to an archaeon) or a misannotated sequence from a bacterium species (discussed in the next topic).

Using our stepwise search and selection method (schematic workflow depicted in fig. 2), we retrieved an initial sample of 76 putative genomic Pif1 helicases from a wide variety of eukaryotes, prokaryotes, and plasmids, all lacking *Helitron* sequences in their genomes. After retrieving this sample of genomic (and plasmid) helicases, we further expanded the number of proteins in our data set by selecting Pif1-like helicases in all major groups of eukaryotes, prokaryotes and viruses, without filtering taxa by the presence of *Helitron* sequences. In addition to Hel domain consensus sequences, this time we also used the *Saccharomyces cerevisiae* Pif1 (NP_013650.1) as a query in Blastp searches. The proteins identified and selected previously with the Rep-filtering procedure were used to aid in the classification of this new set of Pif1-like proteins as genomic helicases or Hel domains from Helitrons by their relationship revealed in the phylogenetic analysis. We also included eukaryotic and prokaryotic viruses in this step of Blastp searches. All taxa selected for further analyses are shown in supplementary table S1, Supplementary Material online.

Phylogenetic Analysis

We used our final sample of 310 aligned protein sequences from *Helitrons*, eukaryotic and prokaryotic organisms, plasmids and viruses, to infer their phylogenetic relationship using the Maximum Likelihood method. Our resulting phylogeny revealed seven well supported major clades (or groups), named as follows: 1) TraA, 2) *Myoviridae*, 3) nucleocytoplasmic large DNA viruses (NCLDV)/*Baculoviridae*, 4) *Helentron*/*Helitron2*, 5) *Helitron*, 6) Prokaryotic, and 7) Eukaryotic clade (fig. 3). The TraA clade included exclusively TraA relaxases and constitute a sister group of the *Myoviridae* clade, which is composed by helicases from a subset myoviruses. The NCLDV/*Baculoviridae* group included helicases from a subset of NCLDV and all retrieved baculoviruses. Together with the *Helentron*/*Helitron2* and *Helitron* clades, they represent a basal group relative to the Prokaryotic and Eukaryotic major clades, as shown in the rooted tree (supplementary fig. S2, Supplementary Material online). The Prokaryotic clade includes most bacterial, archaeal and bacteriophage sequences. In contrast, the Eukaryotic major clade, which formed a sister group with the Prokaryotic clade, included all eukaryotic sequences, plus some bacterial, archaeal, eukaryotic viruses, and bacteriophage sequences, being the most diverse group in the phylogeny.

Regarding the distribution of *Helitron* variants, we observed two distinct and well supported clades, one with *Helitron* and the other containing *Helentron* plus *Helitron2* sequences (fig. 3). However, the connection between these two clades, and between each one of them and other groups of helicases, have low branch support values, and thus are presented collapsed in the phylogeny (fig. 3; supplementary fig. S2, Supplementary Material online). Considering previous analyses involving the Rep domain (Poulter et al. 2003; Heringer and Kuhn 2018) and the fact that a monophyletic origin of all *Helitrons* seems more parsimonious, the observed paraphyletic distribution of two major *Helitron* groups in our phylogeny could represent a methodological artifact (see Discussion). Nevertheless, the

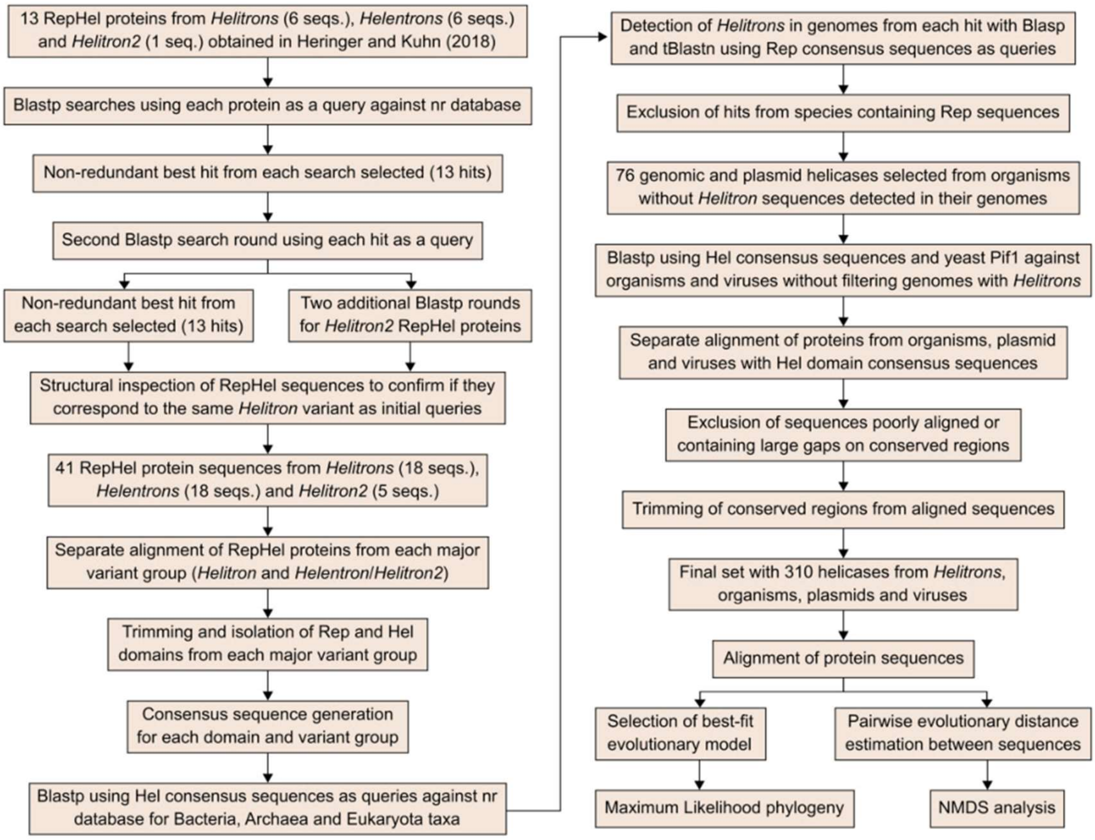

13 RepHel proteins from Helitrons (6 seqs.), Helentrons (6 seqs.) and Helitron2 (1 seq.) obtained in Heringer and Kuhn (2018)
Blastp searches using each protein as a query against nr database
Non-redundant best hit from each search selected (13 hits)
Second Blastp search round using each hit as a query
Non-redundant best hit from each search selected (13 hits)
Two additional Blastp rounds for Helitron2 RepHel proteins
Structural inspection of RepHel sequences to confirm if they correspond to the same Helitron variant as initial queries
41 RepHel protein sequences from Helitrons (18 seqs.), Helentrons (18 seqs.) and Helitron2 (5 seqs.)
Separate alignment of RepHel proteins from each major variant group (Helitron and Helentron/Helitron2)
Trimming and isolation of Rep and Hel domains from each major variant group
Consensus sequence generation for each domain and variant group
Blastp using Hel consensus sequences as queries against nr database for Bacteria, Archaea and Eukaryota taxa
Detection of Helitrons in genomes from each hit with Blasp and tBlastn using Rep consensus sequences as queries
Exclusion of hits from species containing Rep sequences
76 genomic and plasmid helicases selected from organisms without Helitron sequences detected in their genomes
Blastp using Hel consensus sequences and yeast Pif1 against organisms and viruses without filtering genomes with Helitrons
Separate alignment of proteins from organisms, plasmid and viruses with Hel domain consensus sequences
Exclusion of sequences poorly aligned or containing large gaps on conserved regions
Trimming of conserved regions from aligned sequences
Final set with 310 helicases from Helitrons, organisms, plasmids and viruses
Alignment of protein sequences
Selection of best-fit evolutionary model
Pairwise evolutionary distance estimation between sequences
Maximum Likelihood phylogeny
NMDS analysis

FIG. 2. Workflow with the methodology used in our study. See Materials and Methods for a more comprehensive description.

fact that *Helitrons* in general did not group closer to any other major clade, indicates that Hel domains are only distantly related to genomic Pif1 helicases and belong to completely independent lineages. An interesting aspect of the *Helentron/Helitron2* major clade is the presence of a Hel domain from the dinoflagellate *Symbiodinium microadriaticum* (CAE7237458.1) branching externally to the divergence of *Helitron2* and *Helentron* sequences (fig. 3; supplementary fig. S2, Supplementary Material online). This RepHel lacks the apurinic–apyrimidinic (AP) endonuclease domain typical of *Helentrons*, and the element corresponding to this transposase (CAJNJV010003184.1) is structurally more similar to a *Helitron2* variant (fig. 1). Hence, this *Helitron2*-like element appears to represent an intermediate variant that should be more closely related to the common ancestor of *Helentron* and *Helitron2* elements. To our knowledge, this is the first identification of a putative evolutionary intermediate between two *Helitron* variants. In this specific case, the putative intermediate variant was not identified before most likely because the *S. microadriaticum* sequence (CAE7237458.1) was submitted only recently (February 2021).

One of the prokaryotic sequences in the Eukaryotic major clade is a Pif1-like helicase from a Rickettsiales bacterium (MBO87943.1), positioned before the radiation including most eukaryotic Pif1 sequences (fig. 3). Most phylogenetic analyses conducted to date place the order Rickettsiales as the closest relative of mitochondria (reviewed in Roger et al. [2017]). Although this hypothesis has been challenged by some studies (Roger et al. 2017; Martijn et al. 2018), a recent analysis that used more robust methods confirmed the close relationship between Rickettsiales and the mitochondrion ancestor (Fan et al. 2020). Hence, the topology observed in our phylogeny seems to reflect the known evolutionary link between eukaryotic Pif1 proteins and their prokaryotic ancestor, which probably belonged to the symbiont that later gave rise to mitochondria (Bochman et al. 2011).

Another marked feature observed in our phylogeny is the presence of Pif1-like sequences from three eukaryotic species (*Perkinsela* sp., *Phytomonas* sp., and *Strigomonas culicis*) preceding the prokaryotic radiation within the Eukaryotic major clade (fig. 3; supplementary fig S2, Supplementary Material online). These sequences belong to kinetoplastids from the phylum Euglenozoa which, accordingly, is considered the group that diverged earliest during eukaryotic evolution (Cavalier-Smith et al. 2014). Although other kinetoplastid species are grouped separately from these three basal taxa (fig. 3), this distribution could be explained by the presence of

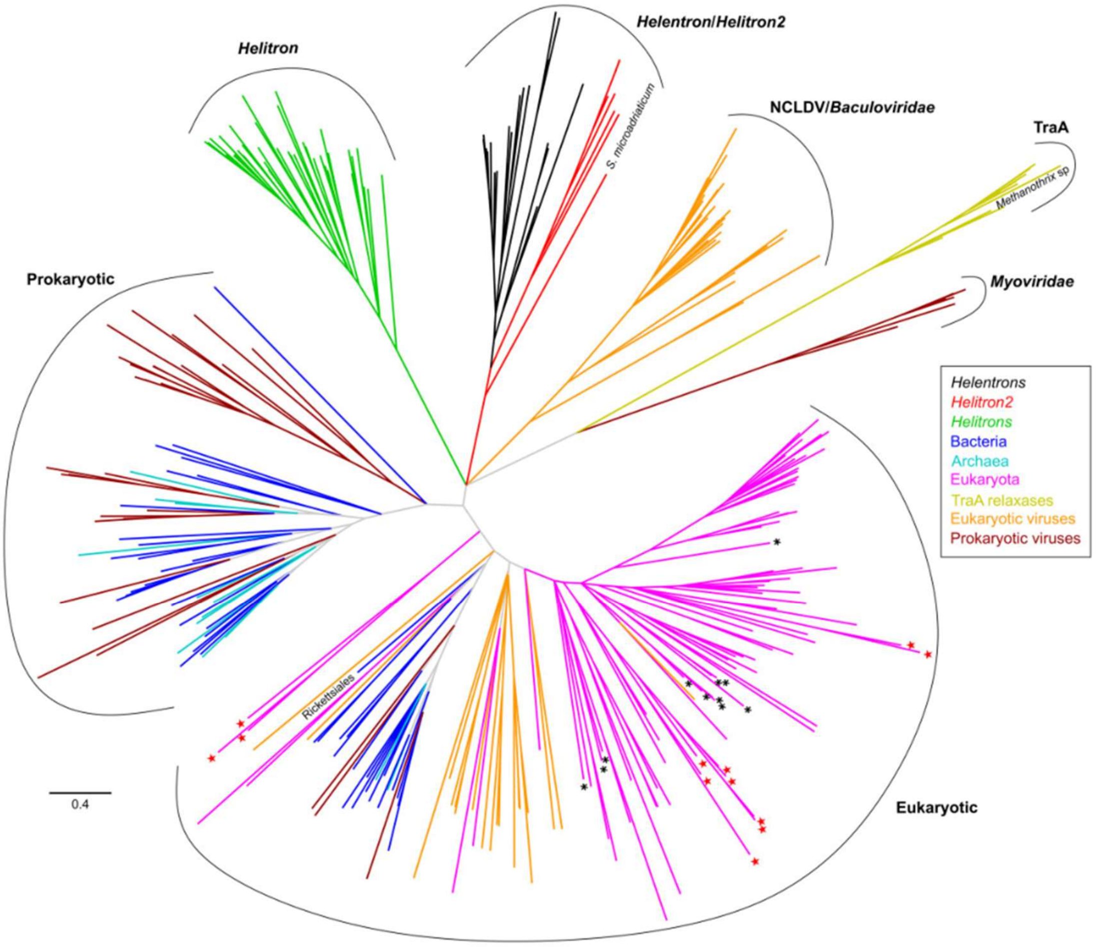


FIG. 3. Maximum-likelihood phylogeny of Pif1-like helicases. The resulting phylogeny includes Pif1-like helicases from *Helitron* variants, viruses, plasmids, and organisms, with seven major clades indicated around the tree. Specific taxa mentioned in the text are shown in branch tips. Kinetoplastids are marked with red stars and amoebae are marked with asterisks. Branches with <0.7 SH-aLRT statistical support were collapsed. The rooted tree with all taxa names and branch support values is shown in supplementary figure S2, Supplementary Material online.

multiple Pif1 paralogs in species from this class, which have been shown to encode up to eight Pif1-like genes (Liu et al. 2009; Bochman et al. 2010). If these three basal sequences represent some of the Pif1 paralogs adapted for kinetoplastid-specific functions (Bochman et al. 2010), a process of positive evolution following subfunctionalization, might have caused them to be artificially positioned externally in relation to other eukaryotic Pif1 helicases. In addition to kinetoplastids, other taxa also displayed a somewhat scattered distribution on the Eukaryotic major clade, instead of forming monophyletic clusters. For instance, amoebal Pif1 helicases were grouped in five separate clades (fig. 3). Interestingly, a scattered distribution of amoebal Pif1-like proteins was also observed in a previous study and it was explained as the result of horizontal gene transfer (HGT) and duplication events (Harman and Manna 2016). Also in the Eukaryotic major clade, eukaryotic viruses, mostly NCLDVs, were found dispersed in different clades, sometimes closer to eukaryotic and prokaryotic organisms than to other groups of viruses (fig. 3; supplementary fig. S2, Supplementary Material online). Although noteworthy, this result agrees with the growing evidence for multiple HGT events between these large viruses and a variety of organisms (reviewed in Barreat and Katzourakis [2021]).

Overall, the scattered topology observed for several taxa from the Eukaryotic major clade might have been the consequence of two main factors. First, as a result of our searching and selection method designed to retrieve Pif1-like helicases with the highest similarity to specific queries. Because we only selected the best results from each taxonomic group, and eukaryotes may have multiple Pif1 genes adapted for distinct functions, it is likely that our sampled sequences represent a mixture of paralogs and orthologs. Second, as a consequence of several HGT events between eukaryotes, prokaryotes, and viruses. Eukaryotes have been involved in HGT exchanges not only with viruses, as mentioned above, but also with multiple prokaryotic groups and sometimes with distinct eukaryotic taxa (reviewed in Husnik and McCutcheon [2018] and Van Etten and Bhattacharya [2020]). Thus, it is possible that Pif1 genes have been horizontally transferred several times during the evolution of eukaryotes.

In the Prokaryotic major clade, cases of interspersed branches from bacteria, archaea, and phages were also abundant, and indicate that several HGT events involving Pif1-like genes have occurred between these taxa (fig. 3). Although horizontally transferred sequences represent a relatively small fraction of eukaryotic genomes, in prokaryotes, HGT has long been considered a primary source of new genes and a major driver of evolution. These gene exchanges are not limited to closely related organisms, as they have been shown to cross prokaryotic domains and sometimes occur between bacteria, archaea and viruses (reviewed in Koonin [2016]). Hence, based on our phylogenetic analysis, it is reasonable to conclude that Pif1-like helicases are also members of the large set of gene families that have been horizontally transferred among prokaryotic organisms. Regardless of the particular explanations for each case, the frequent grouping of relatively distant taxa observed in the Eukaryotic and Prokaryotic major clades indicates that, in addition to ordinary vertical inheritance of genes, other events (e.g., HGTs and gene duplications) have shaped the evolution of genomic Pif1 helicases extensively.

Other interesting results were also revealed by the phylogenetic analysis. For instance, the TraA and *Myoviridae* clades formed sister groups with good branch support (fig. 3; supplementary fig. S2, Supplementary Material online). This result suggests a closer than expected relationship between replicons with completely distinct modes of propagation, underscoring the highly dynamic modularity that is typical of MGEs. Finally, as previously indicated in our Blast results, a protein annotated as belonging to the archaeon genus *Methanothrix* (TFH49976.1) grouped with TraA relaxases from Proteobacteria species, more specifically in the Desulfobacteraceae family (*Desulfobacteraceae bacterium* and *Desulfosarcina cetonica*) (fig. 3; supplementary fig. S2, Supplementary Material online). To verify whether this TraA gene derives from an HGT event or misannotation, we first used its protein sequence (TFH49976.1) as a query in separate Blastp searches against bacteria and archaea in the nr database. In this case, the query was significantly more similar to bacterial than archaeal sequences. We also used the nucleotide sequence corresponding to the protein (accession number: SPBB01000211.1) as a query in Blastn searches against bacteria and archaea in the nucleotide collection (nr/nt) and Whole Genome Shotgun (WGS) contigs databases. In this case, no hits with significant similarity were found in archaea. The query displays a significant identity (up to 75%) to bacterial genes, although limited to short stretches that cover up to 15% of the query length. Furthermore, the contig corresponding to the query only contains the TraA gene without flanking sequences that could be used to determine if this gene was integrated into an archaeal genome. Therefore, this putatively archaeal TraA gene is significantly more similar to bacterial than archaeal sequences, both at the amino acid and nucleotide level. Because this sequence is part of a metagenome assembly (BioSample: SAMN11127048), the possibility of misannotation or contamination in this case is very likely. Together, our analyses indicate that this TraA gene is likely from a bacterial plasmid misannotated as belonging to an archaeon. Regardless of those considerations, knowing the host species of this protein sequence does not change the interpretation of our results.

NMDS Analysis

The estimated evolutionary divergence between sequences were used to represent their distances in two dimensions with nonmetric multidimensional scaling (NMDS) analysis. By doing so, we intended to visualize their spatial arrangement without assuming cladistic relationships, and also verify if their distribution replicates the overall topology observed in the phylogeny.

The arrangement of Pif1-like helicases in the resulting NMDS ordination showed an overall segregation of proteins into seven major clusters (fig. 4). It also displayed a large divergence between Hel domains from the two major clades previously observed in our phylogeny (fig. 3), with *Helentron* and *Helitron2* sequences forming a single group distinctly segregated from *Helitron* variant sequences. In addition, *Helitron* Pif1-like domains from all variants did not appear to be more closely associated with any other specific major group, being roughly equidistant from genomic and viral helicases found in prokaryotes and eukaryotes (fig. 4).

Pif1 helicases from the Eukaryotic and Prokaryotic major groups formed two separate, albeit closely related clusters. Although genomic Pif1 helicases in the Eukaryotic group showed a tendency for clustering with sequences from more closely related taxa, in the Prokaryotic group, sequences from bacteria and archaea displayed a highly interspersed distribution. In both major groups viral sequences were mostly scattered among genomic Pif1 helicases (fig. 4). These distinct arrangements in the Eukaryotic and Prokaryotic major groups confirm the taxonomic incongruences and complex evolutionary history of genomic Pif1 helicases indicated by the phylogenetic analysis.

In sum, the resulting NMDS ordination recapitulates the main features observed in the phylogeny, that is, the segregation of seven major clades, the distant relationship between Hel domains from *Helitrons* and genomic helicases, and the indication of multiple HGT events involving Pif1-like helicases from eukaryotes, prokaryotes, and viruses.

Reassessing the Classification and Number of Pif1 Genes in Eukaryotes

As we have mentioned, Blastp searches on eukaryotic genomes using Pif1 helicases as queries often result in multiple significant hits. Because *Helitrons* are pervasive in most eukaryotic groups and their transposase includes a Pif1-like Hel domain, it is always possible that some of those hits constitute *Helitron* coding sequences, instead of genomic helicases. For example, during our preliminary analyses we

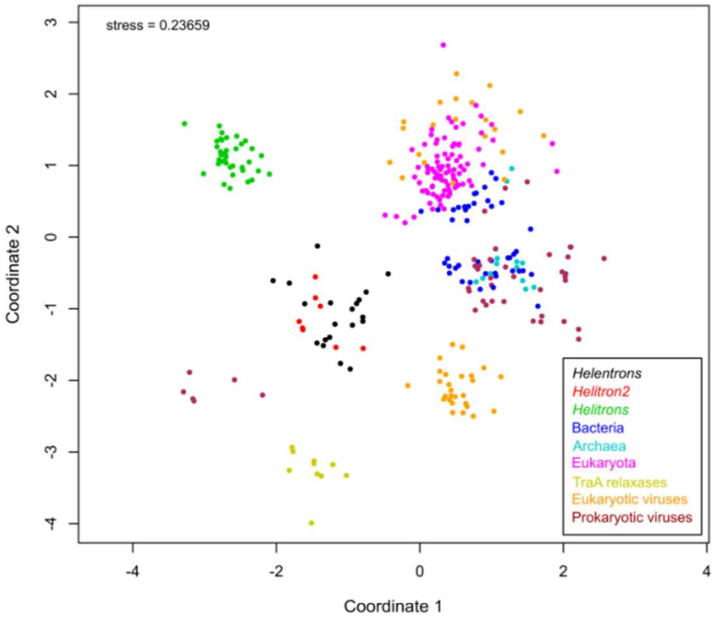


Fig. 4. NMDS plot of Pif1-like helicases. NMDS ordinations representing Euclidean distances between Pif1-like helicase sequences in two dimensions.

performed a Blastp search to identify putative genomic Pif1 helicases in the fungus *Rhizophagus clarus*, using the human Pif1 domain (6HPH_A) as a query, and found many candidate genes, together with RepHel sequences. However, a more detailed inspection revealed that some putative genomic Pif1 helicases are in fact Hel domains from *Helitron* coding sequences lacking the Rep domain in the same ORF (supplementary fig. S1, Supplementary Material online). Thus, without more careful analyses, the structural resemblance between genomic and *Helitron*-derived Pif1 domains can hinder the proper identification of sequences from this protein family. Indeed, to avoid classifying Hel domains as genomic Pif1 helicases, we excluded all species with *Helitrons* in their genomes from our initial Blast searches.

Although some eukaryotes are thought to have multiple genomic Pif1 helicases (Bochman et al. 2010, 2011; Harman and Manna 2016), most species from this domain of life apparently encode one or two Pif1 genes (Bochman et al. 2010). Considering that distantly related eukaryotes like *Schizosaccharomyces pombe* and humans only need one Pif1 helicase to carry out genomic functions, species with supposedly multiple Pif1 paralogs should be evaluated carefully. Thus, we reassessed three cases in the literature referring to genomic Pif1 genes from eukaryotes, which could have included *Helitron*-derived sequences inadvertently.

In the first example, *Arabidopsis thaliana* was described as having three genomic Pif1 helicases (CAB91581, NP_190738, and CAB63155) (Bochman et al. 2010). After examining the structure and sequence of these proteins we found that all of them are either RepHel proteins or Pif1-like sequences with significant identity to *Helitron* transposases (supplementary table S2, Supplementary Material online). Interestingly, a phylogeny of Pif1 sequences presented in the same work (Figure 1 in Bochman et al. 2010) displays a single Pif1 helicase from *Oryza sativa* (ABB47755) grouped together with the three *A. thaliana* proteins mentioned above. Because these three proteins were shown to be derived from RepHel transposases, and *Helitrons* are known to be abundant in the genomes of *A. thaliana* and *O. sativa* (Yang and Bennetzen 2009; Xiong et al. 2014), we examined this Pif1-like sequence from *O. sativa*. After inspecting its structure, we found that this *O. sativa* Pif1-like protein represents a RepHel transposase containing both of its major domains (supplementary table S2, Supplementary Material online). Hence, all these four proteins classified as genomic Pif1 helicases from *A. thaliana* and *O. sativa* constitute either RepHel transposases or Pif1-like Hel domains from *Helitrons*.

In the second example, the fungal pathogen of insects *Metarhizium robertsii* ARSEF 23 (formerly *M. anisopliae* ARSEF 23) was described as the eukaryote harboring the largest number of Pif1 genes, with 23 paralogs (Bochman et al. 2011). We conducted a Blastp search on the genome of this species using the human Pif1 domain (6HPH_A) and the *S. cerevisiae* Pif1 (NP_013650.1) as queries and found that, although *M. robertsii* appears to have up to 25 proteins with some similarity to Pif1 helicases, only 16 of them cannot be readily classified as RepHel transposases, that is, do not contain a Rep domain sequence. Of these 16 proteins, 11 either display significant similarity to RepHel transposases or belong to a cryptic RepHel ORF (truncated transposase with a Rep sequence upstream the Pif1 ORF), and one does not correspond to a Pif1 helicase (supplementary table S3, Supplementary Material online). Hence, only four helicases from *M. robertsii* could represent genomic Pif1 candidates, with the other 20 Pif1-like sequence clearly being derived from *Helitron* transposases.

In the third example, it was suggested based on in silico analyses that *A. thaliana* could have up to 11 Pif1 genes (Knoll and Puchta 2011), with this large number of paralogs being attributed to *Helitrons* capturing and multiplying genomic Pif1 sequences. However, after inspecting all *A. thaliana* Pif1-like proteins on GenBank, retrieved after a Blastp searches using the human Pif1 domain (6HPH_A) and the *S. cerevisiae* Pif1 (NP_013650.1) as queries, we found that all of them either represent RepHel proteins directly or derive from *Helitron* transposases (supplementary table S4, Supplementary Material online). Although we anticipated that some sequences would derive from *Helitrons*, the fact that all retrieved *A. thaliana* Pif1-like proteins appear to represent RepHel transposases directly or indirectly was unexpected, considering the widespread distribution of genomic Pif1 helicases in eukaryotes. To investigate whether this apparent lack of genomic Pif1 homologs is exclusive from *A. thaliana*, we conducted a Blastp search using the same method on *O. sativa*, which is estimated to have diverged from *A. thaliana* ~163 Ma (Li et al. 2019). Like what was observed in *A. thaliana*, we found many Pif1-like sequences in *O. sativa*, with all results representing RepHel transposases directly or indirectly (supplementary table S5, Supplementary Material online).

Given the distant relationship between *A. thaliana* and *O. sativa*, we tried to estimate when genomic Pif1 helicases could have been lost during the evolution of these land plant lineages. To do that, we conducted a series of Blastp searches on taxonomic ranks above *A. thaliana* and *O. sativa* using the human Pif1 domain (6HPH_A) and the yeast Pif1 (NP_013650.1) as queries. Interestingly, genomic Pif1 homologs appear to have been lost in Brassicales and commelinids, the taxonomic groups from which *A. thaliana* and *O. sativa* belong, respectively (fig. 5). The best hits within these groups corresponded to RepHel proteins (supplementary table S6, Supplementary Material online). Conversely, the best hits from searches in taxa outside Brassicales (malvids) and commelinids (Liliopsida) were Pif1 proteins with low similarity to RepHel transposases, despite some of the species with putative genomic Pif1 helicases also having *Helitron* proteins (supplementary table S6, Supplementary Material online). To further confirm the absence of genomic Pif1 homologs in the mentioned groups, we first used the best hits from searches in malvids (EOX92974.1) and Liliopsida (MQL92731.1) as queries in Blastp searches against Brassicales and commelinids, respectively. The results still indicated a lack of genomic Pif1 homologs in Brassicales and commelinids, as the best hits also corresponded to *Helitron* sequences (supplementary table S7, Supplementary Material online). Additionally, we conducted Blastn searches using the nucleotide sequences corresponding to EOX92974.1 (CM001879.1) and MQL92731.1 (NMUH01001479.1) as queries against Brassicales and commelinids, respectively. Although the search against commelinids did not retrieve hits with significant similarity to the genomic Pif1 from Liliopsida, the result from Brassicales revealed a hit in *Bretschneidera sinensis* (JACXJD010000007.1) with 74% identity to the genomic Pif1 nucleotide sequence from malvids. This hit from *B. sinensis* translates to an ORF that appears to be intact, therefore representing a Pif1 gene that has not been annotated yet, which explains its absence in Blastp results. Interestingly, *B. sinensis* (family Akaniaceae) belongs to the most basal clade from Brassicales (Edger et al. 2018), indicating that genomic Pif1 homologs were probably lost shortly after the origin of this order and before the major radiation that gave rise to most extant families of Brassicales.

Although regions flanking genomic Pif1 helicases from malvids and Liliopsida up to tens of kilobase pairs on both sides display similarity to Brassicales and commelinids sequences, this similarity covers only limited portions of their length, as indicated by Blastn searches. Because this observed similarity is not contiguous over the whole span of flanking sequences, it is not possible to define whether they correspond to homolog regions, and therefore we could not determine what caused Pif1 genes to be lost in Brassicales and commelinids. However, it is noteworthy that most of the genomic Pif1-flanking regions with significant identity to sequences from both groups correspond to TEs, particularly LTR retrotransposons, as determined by searches using the Censor tool in Repbase (Kohany et al. 2006). Although with the current data presented it is not possible to ascertain what caused genomic Pif1 helicases to be lost in Brassicales and commelinids, the presence of long TE sequences in the vicinity of those genes in the closest taxonomic groups could be related to these events. For instance, TEs flanking these Pif1 genes could have promoted ectopic recombinations between insertions, leading to the deletion of large chromosome segments in Pif1 gene loci (Kent et al. 2017). However, more extensive analyses would be necessary to pinpoint the precise boundaries of these deleted chromosomal segments and to describe the mechanisms responsible for those events. Nonetheless, our results indicate that at least two major groups of land plants appear to have lost genomic Pif1 homologs independently (fig. 5) and that usual functions performed by this gene might be carried out by different proteins in species from these taxa.

Discussion

The Evolutionary History of *Helitrons* Takes Shape

Because Pif1 helicases are known to be typically eukaryotic proteins (Bochman et al. 2010), and Hel domains found in some RepHel transposases have introns, it has been suggested that an *Helitron* ancestor likely captured a Pif1 gene from its eukaryotic host (Kapitonov and Jurka 2001, 2007; Thomas and Pritham 2015). However, our results indicate that *Helitrons* already encoded a Hel domain before invading eukaryotic genomes (fig. 6), as genomic Pif1 helicases from prokaryotes and eukaryotes formed sister groups in our analyses, with Pif1-like Hel domains being only distantly related to

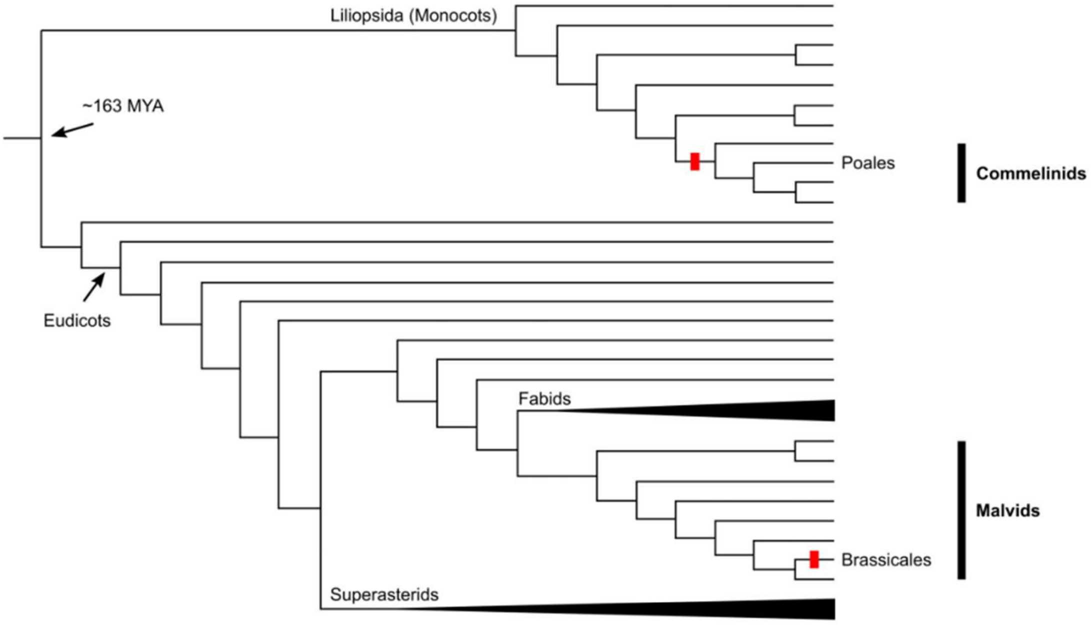


Fig. 5. Cladogram of plant groups that appear to have lost genomic Pif1 helicases. Only major clades are represented, with Poales and Brassicales indicating the orders of *O. sativa* and *A. thaliana*, respectively. Red bars mark the two branches that lack sequences with significant similarity to genomic Pif1 helicases. Phylogeny adapted from Li et al. (2019).

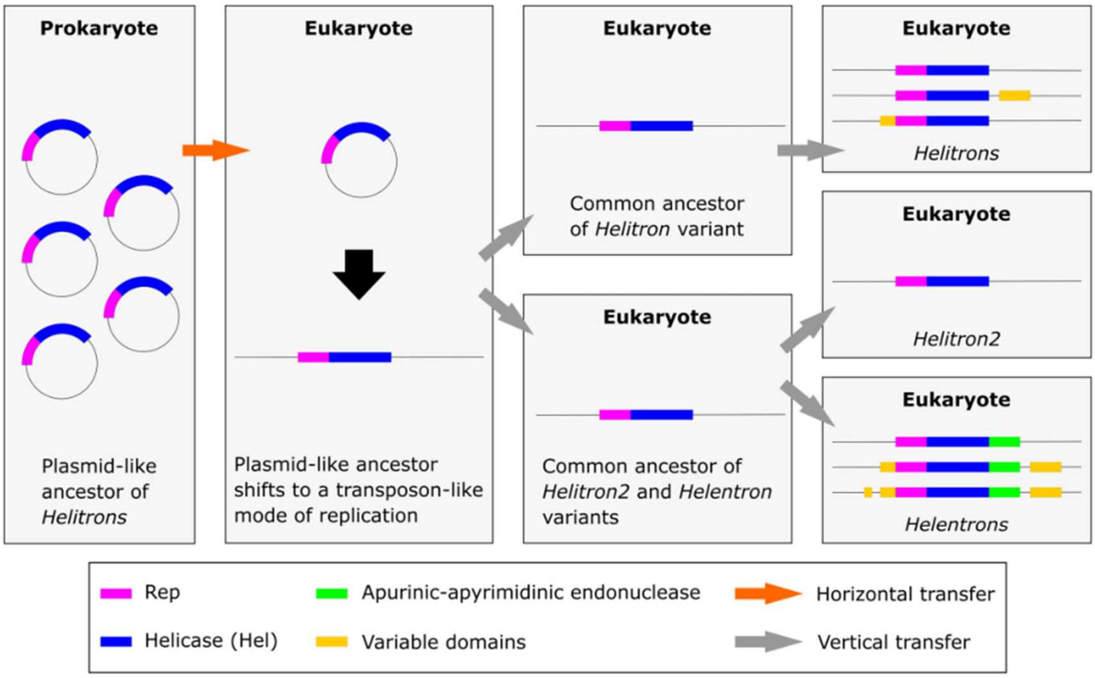

Prokaryote
Plasmid-like ancestor of *Helitrons*
Eukaryote
Plasmid-like ancestor shifts to a transposon-like mode of replication
Eukaryote
Common ancestor of *Helitron* variant
Eukaryote
Common ancestor of *Helitron2* and *Helentron* variants
Eukaryote
Helitrons
Eukaryote
Helitron2
Eukaryote
Helentrons
Rep
Helicase (Hel)
Apurinic-apyrimidinic endonuclease
Variable domains
Horizontal transfer
Vertical transfer

Fig. 6. A hypothesis for the evolution of *Helitrons*. We propose that *Helitrons* descend from prokaryotic plasmid-like elements (first box) that invaded eukaryotic cells during their early evolution. After invading eukaryotes, *Helitrons* shifted to a predominantly transposon-like mode of propagation. During their subsequent adaptation to specific hosts, *Helitrons* diverged into distinct variants (*Helitrons*, *Helentrons*, and *Helitron2*) and captured additional domains. Arrows represent major steps during the evolution of *Helitrons*.

them. Nonetheless, in addition to a RepHel with its archetypal double-domain structure, *Helentrons* also have an AP endonuclease domain in their transposase (fig. 1), which was probably captured from a non-LTR retrotransposon residing in the same eukaryotic host (Thomas and Pritham 2015). The capture of an AP endonuclease gene likely marked the evolutionary origin of *Helentrons* from *Helitron2*-like ancestors, which also gave rise to the *Helitron*2 variant. Our identification of an intermediate Hel domain from *S. microadriaticum* branching externally to *Helentron* and *Helitron2* sequences constitute the first direct evidence for a *Helitron2*-like element as the ancestor of both variants. Besides the AP endonuclease from *Helentrons*, several other domains have been incorporated to specific *Helitron* lineages during their evolution in eukaryotic genomes (Thomas and Pritham 2015) (fig. 6). However, the function of AP endonucleases and other coding sequences captured by *Helitrons* from eukaryotes have not been determined yet.

Although the evolutionary proximity of *Helentron* and *Helitron2* lineages was expected (Thomas and Pritham 2015; Heringer and Kuhn 2018), our results indicating that Hel domains from the *Helitron* variant form a distinct group from the *Helentron* and *Helitron2* variants (figs. 3 and 4) contrasts with the monophyletic distribution previously observed for *Helitron* Rep domains (Poulter et al. 2003; Heringer and Kuhn 2018). Assuming the more parsimonious scenario in which *Helitrons* constitute a monophyletic group, the resulting paraphyletic distribution of Hel domains might have been caused by faster evolutionary rates that occurred on this protein region. The same topology was not observed for Rep domains in previous studies, probably due to a higher tendency for amino acid sequence conservation in this portion of *Helitron* transposases. If Hel domains evolved under less constrained evolutionary pressures or went through a stronger positive selection imposed by their hosts, these processes could have potentially masked their monophyletic nature. Furthermore, the widespread distribution of *Helitrons* in eukaryotes (Thomas and Pritham 2015) and the overall similarity between RepHel and host phylogenies, indicate that *Helitrons* began to diverge before the emergence of most eukaryotic kingdoms (Poulter et al. 2003). As time estimates of major eukaryote radiations date back to approximately 1 billion years ago (Douzery et al. 2004; Berney and Pawlowski 2006), the first *Helitron* lineage divisions likely have a similar age. Thus, a rapid evolution of Hel domains that occurred through a very long period of time might have contributed to blur the monophyletic nature of *Helitrons* in our analyses.

An independent example supporting the hypothesis that each domain from RepHel proteins have evolved under distinct evolutionary pressures can be viewed in the phylogenies of *Helitron* Rep and Hel domains inferred by Poulter et al. (2003), which present distinct topologies. In their Rep domain phylogeny, *Helitron* sequences from the fungus *Phanerochaete chrysosporium* clustered with *Helentrons*, instead of *Helitrons*. Conversely, in the Hel domain phylogeny, all elements segregated into variant-specific clades, indicating that distinct *Helitron* variants display a more pronounced sequence divergence in this region. Furthermore, in the Hel phylogeny, *Helitron* clades were connected by relatively longer branches when compared with the Rep domain tree, similarly to the observed between our results presented here for Hel domains (supplementary fig. S2, Supplementary Material online) and on our previous study involving Rep domains (Heringer and Kuhn 2018). It is worth mentioning that, in contrast to our phylogeny, the one presented by Poulter et al. (2003) did not display a polyphyletic distribution for Hel domains. The reason for that might be related to the smaller sample size and diversity of *Helitrons* used in the latter analysis when compared with the one presented here.

Altogether, these observations suggest that each domain from RepHel transposases has evolved under distinct evolutionary rates. These differences could be derived from selective pressures that constrained the Rep amino acid sequence to a higher degree, and/or favored a more rapid evolution of the Hel domain to optimize its interaction with host components. Hence, a very early radiation of *Helitrons*, combined with relatively faster evolutionary rates that have occurred in Hel domains since they first invaded eukaryotes, probably explain the spurious paraphyletic distribution between major *Helitron* groups in our results. In this case, the observed topology could represent a result of long-branch attraction (Bergsten 2005).

In summary, our phylogenetic and NMDS analyses indicate that RepHel proteins evolved independently from genomic Pif1 helicases found in prokaryotes and eukaryotes. Thus, in spite of previous hypotheses about the origins of Hel domains, it is unlikely that a *Helitron* ancestor captured a Pif1 gene from its eukaryotic host. Instead, we suggest that, before entering eukaryotic cells, *Helitrons* already encoded RepHel proteins, branching into two major lineages after they invaded eukaryotic genomes (fig. 6). From there on, Hel domains probably evolved under relatively faster rates, which could explain their distribution into marked separate groups, in contrast to what was observed in analyses of Rep domains (Poulter et al. 2003; Heringer and Kuhn 2018).

Helitrons May Be Descendants of Plasmid-Like Elements

Although it seems clear that neither Rep nor Hel domains have originated from genomic proteins, the ancestor of *Helitrons* probably resided within a prokaryotic cell. If this ancestor already had a transposon-like mode of propagation, it is conceivable that their descendants (or their remnants) could still reside in genomes of some unknown prokaryote lineages. However, even assuming the hypothesis of a transposon ancestor as correct, it is unlikely that such elements would be found, as sequences that do not benefit cellular functioning directly (like TEs) are subject to extremely rapid turnover rates in prokaryotes (Sela et al. 2016; Wolf et al. 2016). A second possibility is that prokaryotic ancestors of *Helitrons* had a predominantly plasmid-like mode of replication before they became eukaryotic TEs. This scenario not only agrees with the current lack of *Helitron*-like sequences in prokaryotes, but with the close relationship found between Rep domains from *Helitrons* and RC bacterial plasmids (Heringer and Kuhn 2018; Kazlauskas et al. 2019) and the fact that *Helitrons* generate plasmid-like intermediates during transposition (Grabundzija et al. 2018).

It is worth mentioning that a TraA relaxase was the only protein from a MGE retrieved in our Blast searches using Hel domains as queries. Similarly to RepHel transposases, TraA and other plasmid relaxases possess Rep-like and helicase domains within the same protein (Pérez-Mendoza et al. 2006; Chandler et al. 2013). Although Rep-like domains found in relaxases display an inverted orientation of their main catalytic motifs when compared with RepHel transposases, both enzymes have an overall similar architecture, consisting of a Rep followed by a helicase domain. In addition, despite their inverted orientation, the 3D topology of these motifs in relaxases and RCR proteins is essentially the same (Chandler et al. 2013). Interestingly, the cryo-EM structure of the RepHel in complex with the *Helitron* 5′-end ssDNA was solved only recently, revealing an even higher degree of organizational similarity with relaxases, particularly with TraI (Kosek et al. 2021). As mentioned by the authors, the structural similarity between these two classes of proteins does not imply a close evolutionary relationship, which is also supported by our results and previous studies involving the Rep domain (Heringer and Kuhn 2018; Kazlauskas et al. 2019). If these structural resemblances are most likely the result of convergent evolution, they would suggest the existence of functional parallels between relaxases and RepHel transposases. Nonetheless, the fact that a group of relaxases was retrieved in our searches by sequence similarity with Hel domains from *Helitrons* could still indicate a distant evolutionary relationship between these proteins.

Based on these considerations, we propose that *Helitrons* descend from prokaryotic plasmid-like elements that shifted to a transposon mode of propagation after invading eukaryotic cells (fig. 6). Importantly, a transition from an RCR plasmid to an RC TE would likely not require major adaptations, as the replicative processes employed in both types of MGEs work by the same basic enzymatic steps, only differing in the number of DNA substrates and type of final products involved (Chandler et al. 2013; Wawrzyniak et al. 2017).

What Is the Function of Pif1 Helicases in *Helitrons*?

Experimental assays revealed that *Helitrons* have to generate dsDNA circle intermediates in order to transpose, as ssDNA circular elements transfected into human cells were not viable substrates for host genome integration (Grabundzija et al. 2018). The formation of dsDNA intermediates could be achieved by the concomitant synthesis of leading and lagging strands while the element's leading strand is being "peeled-off," or by the addition of a short lagging strand primer on the unwound leading strand before an ssDNA circle is formed. In either case, these processes would require the recruitment of replication fork and DNA repair machinery components (Grabundzija et al. 2018), both of which Pif1 helicases are part of Bochman et al. (2010) and Muellner and Schmidt (2020). For instance, Pif1 stimulates the activity of DNA polymerase δ (Pol δ) during DNA repair and replication (Pike et al. 2009; Wilson et al. 2013; Koc et al. 2016) through its interaction with the proliferating cell nuclear antigen (PCNA) (Wilson et al. 2013; Buzovetsky et al. 2017; Dahan et al. 2018). In addition, Pif1 has a role in fork convergence, resolving the stalling of these structures, which are expected to occur in the final stages of linear and circular DNA replication (Deegan et al. 2019). Another relevant feature of Pif1 helicases is their preference for binding and unwinding forked structures (dsDNA with ssDNA overhangs) (Ramanagoudr-Bhojappa et al. 2013; Li et al. 2016), which are substrates expected to be formed in the first stages of RCT, when RepHel nicks the *Helitron*'s leading strand in its 5′-end (Dias et al. 2016; Grabundzija et al. 2016, 2018).

The combination of those Pif1 attributes suggests that the Hel domain could aid in the RepHel association to forked DNA structures during the initial steps of transposition and help to recruit replication machinery components from hosts (e.g., PCNA and Pol δ). Although prokaryotic RC TEs, which are thought to transpose similarly to *Helitrons*, do not encode helicases, it is possible that a Hel domain merged to a Rep protein confers mechanistic advantages for RCT in eukaryotic cells and maybe is essential in this environment. Indeed, it has been shown that a mutation in the Walker A motif from Hel domains causes *Helitrons* to lose their transposition activity in cells (Grabundzija et al. 2016). In addition, the RepHel cryo-EM structure reveals a considerable interface between the catalytic portion of Rep and the Hel domain, suggesting that they act in conjunction to unwind dsDNA and generate sufficient ssDNA to allow strand cleavage as transposition starts (Kosek et al. 2021). Thus, it is conceivable that a Hel domain also favored the invasion and colonization of eukaryotic genomes by *Helitrons*, which would explain their pervasiveness in this domain of life that lacks other groups of RC TEs.

Additionally, the Hel domain could facilitate the final stages of transposition, when the RepHel associated with a circular intermediate binds its target site before integration. In contrast to prokaryotic RC TE insertions, which are guided by site specificity (Garcillán-Barcia et al. 2002), *Helitrons* integrate between AT, TT, or TC dinucleotides, depending on the variant, with no preference for unique sequences (Thomas and Pritham 2015). Hence, the RepHel in complex with a *Helitron* intermediate could initially bind its target site by associating with specific DNA or chromatin structures, instead of using sequence guided recognition. In this case, an initial contact would be favored by the known affinity of Pif1 helicases to DNA secondary structures typically found in recombination sites and gene promoters (Bochman et al. 2012; Byrd and Raney 2015; Muellner and Schmidt 2020). Indeed, experimental assays revealed that active *Helitrons* appear to preferentially target highly expressed gene regions (Grabundzija et al. 2016). After a structure-based association mediated also by Hel, the Rep domain would be able nick the recipient strand at a nearby AT, TT or TC dinucleotide site, before transferring an ssDNA intermediate to the host's chromosome, forming a heteroduplex and completing transposition (Kapitonov and Jurka 2007; Thomas and Pritham 2015; Dias et al. 2016).

Taken together, these features of Pif1 helicases and *Helitrons* appear to agree with a scenario in which Hel domains play a more sophisticated role during RCT, beyond simply unwinding double-stranded DNA elements. The presence of a Pif1-like Hel domain in *Helitron* transposases may have provided an advantage over the recruitment of host helicases, by concatenating the processes of DNA binding, leading strand nicking, and peeling-off, together with the formation of circular dsDNA intermediates, all conducted by the same enzyme. In addition, Hel domains could aid the association between RepHel–dsDNA intermediates and target sites on host chromosomes.

Helitrons Can Hamper the Identification of Eukaryotic Pif1 Helicases

The abundance of *Helitrons* in eukaryotic genomes, together with the general similarities between *Helitron* Pif1-like Hel domains and genomic Pif1 helicases from eukaryotes, make their distinction by in silico methods complicated. Our reevaluation of three examples in the literature describing Pif1 proteins from *A. thaliana*, *O. sativa*, and *M. robertsii* demonstrated how these problems have affected the classification and number estimation of genomic Pif1 helicases in eukaryotic species. In these cases, most, or all putative genomic Pif1 helicases described were shown to represent *Helitron*-derived sequences.

Interestingly, during our searches for genomic Pif1 candidates in *A. thaliana* and *O. sativa* we found that all Pif1-like proteins from these species either represent complete *Helitron* transposase sequences or Hel domains from broken RepHel ORFs. After investigating higher taxonomic ranks from which *A. thaliana* and *O. sativa* belong (Brassicales and commelinids, respectively), we found that both of them appear to have lost genomic Pif1 homologs independently (fig. 5). Even granting that Brassicales and commelinids may have genomic Pif1 homologs that went undetected in our searches, the fact that RepHel sequences represented the best hits to eukaryotic Pif1 helicases points to a similar evolutionary pattern in those distantly related groups. However, this issue should be further investigated to determine in more detail how the Pif1 family have evolved in land plants and if some of them have different proteins to perform the same functions of genomic Pif1 helicases.

Despite the examples described above, some eukaryotes have multiple bona fide genomic Pif1 helicases. As we have mentioned, kinetoplastids encode several Pif1 paralogs that likely participate in distinct functions related to their unique biology (Liu et al. 2009; Bochman et al. 2010). Furthermore, *Helitron* transposases are not found in kinetoplastid genomes, as indicated by our Blast searches and a previous analysis (Thomas and Pritham 2015). Hence, all Pif1 helicases found in this group might consist of genomic representatives derived from gene duplications. In addition to kinetoplastids, some amoebae also have multiple genomic Pif1 helicases, with *Acanthamoeba castellanii* encoding up to nine Pif1 genes (Harman and Manna 2016). Our Blast searches revealed that these amoebae species do not have RepHel sequences in their genomes, which confirms that these proteins indeed represent genomic Pif1 helicases. Thus, kinetoplastids and amoebae are the only eukaryotic groups so far in which there is solid evidence for species with more than two genomic Pif1 paralogs.

Altogether, it is clear that our knowledge about the distribution and number of genomic Pif1 helicases in eukaryotes is relatively limited to a small number of species. As we have shown, some of the attempts to identify genomic Pif1 proteins in eukaryotes have been hampered by the large amount of *Helitron* transposases found in this domain of life. It will be important to establish a reliable and efficient method to correctly discriminate between these two major groups of Pif1 helicases, before they are studied in large-scale analyses.

Conclusion

Although the similarity between Hel domains and genomic Pif1 helicases has been noted since the discovery of *Helitrons* 20 years ago, no study had explored their evolutionary connections. Despite previous suggestions that an *Helitron* ancestor likely acquired the Hel domain by capturing a Pif1 gene from its eukaryotic host, our results indicate that RepHel proteins already had their archetypal structure with two domains before invading eukaryotes. Furthermore, considering phylogenetic, structural, and mechanistic aspects of these elements, we propose that *Helitron* ancestors probably had a plasmid-like mode of replication in prokaryotic hosts, before invading eukaryotes and shifting into a transposon. Based on the known features of Pif1 helicases and RepHel proteins, we also hypothesize that Hel domains likely perform a more complex function during transposition, beyond simply unwinding *Helitron* double-stranded DNA.

In addition, our reassessment of the literature describing eukaryotic Pif1 helicases revealed that many of these examples actually represent complete or partial RepHel transposases from *Helitrons*, which are commonly abundant in eukaryotic genomes. This finding highlights the need for a careful inspection before classifying Pif1-like proteins as genomic helicases in eukaryotes, particularly in species that appear to harbor multiple Pif1-like genes. We also found that two distantly related groups of land plants appear to lack genomic Pif1 homologs, despite having multiple Pif1-like Hel domain sequences derived from *Helitrons*. This observation should be studied in more detail, as Pif1 helicases have been considered essential in many genomic processes that are conserved in all eukaryotes studied to date.

Materials and Methods

Selection of RepHel Sequences

We used RepHel protein sequences obtained in our previous study (Heringer and Kuhn 2018), belonging to the three main *Helitron* variants (*Helitron*, *Helentron*, or *Helitron2*) (Thomas and Pritham 2015), as initial queries in a series of Blastp searches on the nonredundant protein sequences (nr) database from GenBank (Sayers et al. 2019). With this strategy, we were able to retrieve a sample with a larger variety of RepHel representatives, thus enabling the generation of more accurate consensus sequences of each domain (Rep and Hel). Each one of the initial 13 *Helitron* protein sequences was used as a query to select an additional RepHel, which in turn, was used as a query to select another sequence in a second Blastp search round. In each of these searches the best hit, sorted by Max Score, was selected, excluding sequences found in genomes of the same genus in a previous round. For the *Helitron2* variant we applied four rounds of consecutive searches to increase the number of sequences, as it had a single representative in our previous analysis (Heringer and Kuhn 2018). To determine whether the additional RepHel sequences belonged to the same variant as the initial queries, we visually inspected their structure with the Conserved Domain Database (CDD) search tool (Lu et al. 2020), following the classification provided by Thomas and Pritham (2015). This classification considers differences in amino acids within conserved regions from the Rep domain and the presence or absence of specific domains in the RepHel protein. A total of 41 RepHel protein sequences were selected for further analyses: 18 from *Helitrons*, 18 from *Helentrons*, and 5 from *Helitron2* elements. Sequences from *Helitron* and *Helentron*/*Helitron2* variants were aligned separately using the auto mode from the MAFFT online service (Katoh et al. 2019). *Helentron* and *Helitron2* sequences were aligned as a single group because these variants are known to be closely related (Thomas and Pritham 2015; Heringer and Kuhn 2018). Rep and Hel domains from each protein were isolated and trimmed, keeping only well-defined conserved regions among aligned sequences. These conserved regions were used to generate consensus sequences of each domain from *Helitron* and *Helentron*/*Helitron2* variants, considering the most common amino acid in each site (supplementary data S1, Supplementary Material online), using the Advanced Consensus Maker tool from the HIV Database (https://www.hiv.lanl.gov/content/sequence/CONSENSUS/AdvCon.html; last accessed November 16, 2021).

Stepwise Search and Selection of Helicase Protein Sequences

The Hel domain consensus sequences of *Helitron* and *Helentron*/*Helitron2* variants (supplementary data S3, Supplementary Material online) were used as queries in Blastp searches against the nr database from GenBank (Sayers et al. 2019), which includes all available annotated proteins for a given taxa. A sample of protein sequences representing a wide variety of organisms were retrieved from distinct taxonomic levels, depending on their number of resulting hits in preliminary Blastp searches. For example, in eukaryotes, searches were conducted from the kingdom down to the class level, as this domain displayed a large number of significant results distributed heterogeneously across thousands of genomes. Conversely, in bacteria we conducted searches at the phylum level, and in archaea the whole sample was retrieved at the domain level itself. The best hits (sorted by Max Score) from Blastp searches using consensus sequences of both *Helitron* and *Helentron*/*Helitron2* variants were selected. Each species containing best hits had one or two protein sequence representatives selected, depending on whether searches using different variant consensuses retrieved the same or different best hits, respectively. To verify if *Helitrons* were present in the genomes of species containing selected hits, we carried out a second round of searches in these taxa, this time using Rep consensus sequences as queries. Blastp searches were conducted against the nr database and tBlastn searches were conducted against the WGS contigs database. Because the aim of our study was to investigate the relationship between Hel domains from *Helitrons* and genomic Pif1 helicases, taxa containing hits corresponding to Rep sequences in any of the two searches (Blastp or tBlastn) were excluded at this stage. By doing so, we expected to have avoided the inclusion of helicases derived from *Helitrons* during the retrieval of putative genomic helicases, which could result in false phylogenetic inferences. Using these criteria, we were able to select 76 Pif1-like sequences from a wide variety of organisms lacking Rep sequences in their genomes. To expand our sample, we used Hel domain consensus sequences and the *S. cerevisiae* Pif1 (NP_013650.1) as queries in Blastp searches against the same groups of organisms from the previous analysis, this time without filtering taxa with Rep sequences in their genomes and including eukaryotic and prokaryotic viruses. Because Pif1-like proteins selected in the initial searches could be more readily identified as either genomic or *Helitron*-derived helicases, they were used to aid in the classification of sequences retrieved without the Rep-filtering procedure by their relationship revealed later in the phylogenetic analysis.

Alignment and Isolation of Helicase Domains

Helicase sequences from each major taxon group (Eukaryota, Bacteria, Archaea, plasmids, eukaryotic, and prokaryotic viruses) were aligned separately with the Hel domain consensus sequences from *Helitrons* and *Helentrons*/*Helitron2* using the auto mode from the MAFFT online service (Katoh et al. 2019) in order to identify a common region among them. Sequences that aligned poorly or displayed large gaps on conserved regions were excluded using the MAFFT data set refinement tool also available in the MAFFT online service (Katoh et al. 2019). Segments extending upstream and downstream the central conserved regions were visualized using MEGAX (Kumar et al. 2018) and trimmed to avoid spurious alignments between nonrelated portions of proteins. This procedure is important considering that a large majority of prokaryotic and eukaryotic proteins contain multiple domains that have evolved through modular rearrangements (Bornberg-Bauer et al. 2005; Wang and Caetano-Anollés 2009). Even among genomic Pif1-like domains from eukaryotes, there are low levels of sequence and size similarity in their N- and C-terminal regions extending beyond a conserved core (Boule and Zakian 2006). Thus, when conducting a phylogenetic analysis of highly divergent protein sequences, it is preferable to only consider limited domain regions as evolutionary units, because flanking segments can evolve through distinct selective constraints. A total of 310 helicases from *Helitrons* (65 sequences), eukaryotic (89 sequences) and prokaryotic organisms (56 sequences), plasmids (10 sequences), eukaryotic viruses (48 sequences), and prokaryotic viruses (42 sequences) were selected for the next step of our analyses (supplementary table S1, Supplementary Material online). Trimmed helicase domains from all taxa, including *Helitrons*, were aligned using the E-INS-i method combined with mafft-homologs in the MAFFT online service (Katoh et al. 2019). The final alignment containing all sequences used in the following analyses are available in supplementary data S2, Supplementary Material online.

Phylogenetic and NMDS Analyses

The best-fit evolutionary model for the alignment (LG + G + I) was selected using the smart model selection in PhyML (Lefort et al. 2017). The maximum likelihood phylogeny of aligned amino acid sequences was inferred with the SPR method of tree topology search, six random plus one parsimony starting trees and six substitution rate categories across sites modeled with estimated gamma-shaped distribution parameter and proportion of invariant sites. Branch supports were estimated using the approximate likelihood ratio test (aLRT) with the nonparametric Shimodaira–Hasegawa correction (SH-aLRT). All these procedures were conducted on PhyML 3.1 (Guindon et al. 2010). Branches with <0.7 SH-aLRT statistical support were collapsed using TreeGraph 2 (Stöver and Müller 2010) and the final tree visualized using FigTree v.1.4.2 (http://tree.bio.ed.ac.uk/software/figtree/; last accessed November 16, 2021). For the NMDS analysis, pairwise evolutionary distances between aligned sequences were estimated with the JTT matrix-based model and the rate variation among sites modeled with a gamma distribution on MEGAX (Kumar et al. 2018). NMDS ordinations with Euclidean distances of the sequences represented in two dimensions were generated using the R package vegan v2.5-6 (Dixon 2003). The NMDS analysis and plotting were executed in RStudio v1.3.959 (RStudio Team 2020) with R v4.0.0 (R Core Team 2020). All the methodology described heretofore is represented as a schematic workflow in figure 2.

Search and Classification of Pif1-Like Proteins in Eukaryotic Species

To reexamine selected examples from the literature describing genomic Pif1 helicases, which could in fact constitute RepHel-derived sequences, we inspected the structure of those proteins using the CDD search tool (Lu et al. 2020). To reassess the description of species containing multiple genomic Pif1 helicases we conducted Blastp searches in the protein sequences from the corresponding taxa available in the nr database from GenBank (Sayers et al. 2019) using the human Pif1 domain (6HPH_A) and *S. cerevisiae* Pif1 protein (NP_013650.1) as queries. In order to verify if the resulting sequences corresponded to RepHel transposases, all hits had their structural features inspected with the CDD search tool (Lu et al. 2020). Hits that did not included a conserved Rep domain identified by the CDD search tool were used as queries in a second round of Blastp searches against the nr database from GenBank to check if they might constitute Hel domains from broken *Helitron* transposases (Hel domains highly similar to RepHel proteins) or cryptic RepHel proteins (truncated transposase with a Rep sequence upstream the Pif1 ORF). If the best hits (sorted by Max Score) from this second round of searches corresponded to RepHel proteins, queries were considered as derived from *Helitrons*. In contrast, if the resulting best hits did not correspond to RepHel sequences, queries were classified as putative genomic Pif1 helicases.

Supplementary Material

Supplementary data are available at *Molecular Biology and Evolution* online.

Acknowledgments

We would like to thank Dr Guilherme B. Dias for the helpful comments on an earlier version of the manuscript and three anonymous reviewers whose comments and suggestions helped to improve several aspects of the paper. We are also grateful to Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq) (fellowship 308386/2018-3 to G.C.S.K.) and Coordenação de Aperfeiçoamento de Pessoal de Nível Superior—Brasil (CAPES) (doctoral fellowship to P.H.).

Data Availability

The data underlying this article are available in the article and in its supplementary material.

References

Alt-Mörbe J, Stryker JL, Fuqua C, Li PL, Farrand SK, Winans SC. 1996. The conjugal transfer system of *Agrobacterium tumefaciens* octopine-type Ti plasmids is closely related to the transfer system of an IncP plasmid and distantly related to Ti plasmid vir genes. *J Bacteriol*. 178(14):4248–4257.

Barreat JGN, Katzourakis A. 2021. Paleovirology of the DNA viruses of eukaryotes. *Trends Microbiol.* https://doi.org/10.1016/j.tim.2021.07.004.

Bergsten J. 2005. A review of long-branch attraction. *Cladistics* 21(2):163–193.

Berney C, Pawlowski J. 2006. A molecular time-scale for eukaryote evolution recalibrated with the continuous microfossil record. *Proc R Soc B*. 273(1596):1867–1872.

Bochman ML, Sabouri N, Zakian VA. 2010. Unwinding the functions of the Pif1 family helicases. *DNA Repair* 9(3):237–249.

Bochman ML, Judge CP, Zakian VA. 2011. The Pif1 family in prokaryotes: what are our helicases doing in your bacteria? *Mol Biol Cell*. 22(12):1955–1959.

Bochman ML, Paeschke K, Zakian VA. 2012. DNA secondary structures: stability and function of G-quadruplex structures. *Nat Rev Genet*. 13(11):770–780.

Bornberg-Bauer E, Beaussart F, Kummerfeld SK, Teichmann SA, Weiner J. 2005. The evolution of domain arrangements in proteins and interaction networks. *Cell Mol Life Sci*. 62(4):435–445.

Boule JB, Zakian VA. 2006. Roles of Pif1-like helicases in the maintenance of genomic stability. *Nucleic Acids Res*. 34(15):4147–4153.

Buzovetsky O, Kwon Y, Pham NT, Kim C, Ira G, Sung P, Xiong Y. 2017. Role of the Pif1-PCNA complex in Pol δ-dependent strand displacement DNA synthesis and break-induced replication. *Cell Rep*. 21(7):1707–1714.

Byrd AK, Raney KD. 2015. A parallel quadruplex DNA is bound tightly but unfolded slowly by Pif1 helicase. *J Biol Chem*. 290(10):6482–6494.

Cavalier-Smith T, Chao EE, Snell EA, Berney C, Fiore-Donno AM, Lewis R. 2014. Multigene eukaryote phylogeny reveals the likely protozoan ancestors of opisthokonts (animals, fungi, choanozoans) and Amoebozoa. *Mol Phylogenet Evol*. 81:71–85.

Chandler M, De La Cruz F, Dyda F, Hickman AB, Moncalian G, Ton-Hoang B. 2013. Breaking and joining single-stranded DNA: the HUH endonuclease superfamily. *Nat Rev Microbiol*. 11(8):525–538.

Dahan D, Tsirkas I, Dovrat D, Sparks MA, Singh SP, Galletto R, Aharoni A. 2018. Pif1 is essential for efficient replisome progression through lagging strand G-quadruplex DNA secondary structures. *Nucleic Acids Res*. 46(22):11847–11857.

Deegan TD, Baxter J, Bazán MÁO, Yeeles JT, Labib KP. 2019. Pif1-family helicases support fork convergence during DNA replication termination in eukaryotes. *Mol Cell*. 74(2):231–244.

Dias GB, Heringer P, Kuhn GCS. 2016. Helitrons in *Drosophila*: chromatin modulation and tandem insertions. *Mob Genet Elements* 6(2):e1154638

Dixon P. 2003. VEGAN, a package of R functions for community ecology. *J Veg Sci*. 14(6):927–930.

Douzery EJ, Snell EA, Bapteste E, Delsuc F, Philippe H. 2004. The timing of eukaryotic evolution: does a relaxed molecular clock reconcile proteins and fossils? *Proc Natl Acad Sci U S A*. 101(43):15386–15391.

Edger PP, Hall JC, Harkess A, Tang M, Coombs J, Mohammadin S, Schranz ME, Xiong Z, Leebens-Mack J, Meyers BC, et al. 2018. Brassicales phylogeny inferred from 72 plastid genes: a reanalysis of the phylogenetic localization of two paleopolyploid events and origin of novel chemical defenses. *Am J Bot*. 105(3):463–469.

Fan L, Wu D, Goremykin V, Xiao J, Xu Y, Garg S, Zhang C, Martin WF, Zhu R. 2020. Phylogenetic analyses with systematic taxon sampling show that mitochondria branch within Alphaproteobacteria. *Nat Ecol Evol*. 4(9):1213–1219.

Feschotte C, Wessler SR. 2001. Treasures in the attic: rolling circle transposons discovered in eukaryotic genomes. *Proc Natl Acad Sci U S A*. 98(16):8923–8924.

Garcillán-Barcia MP, Bernales I, Mendiola MV, De La Cruz F. 2002. IS91 rolling-circle transposition. In: Craig N, Craigie R, Gellert M, Lambowitz A, editors. Mobile DNA II. Washington, DC: ASM Press. p. 891–904.

Grabundzija I, Messing SA, Thomas J, Cosby RL, Bilic I, Miskey C, Gogol-Döring A, Kapitonov V, Diem T, Dalda A, et al. 2016. A Helitron transposon reconstructed from bats reveals a novel mechanism of genome shuffling in eukaryotes. *Nat Commun*. 7:10716.

Grabundzija I, Hickman AB, Dyda F. 2018. Helraiser intermediates provide insight into the mechanism of eukaryotic replicative transposition. *Nat Commun*. 9(1):1278.

Guindon S, Dufayard JF, Lefort V, Anisimova M, Hordijk W, Gascuel O. 2010. New algorithms and methods to estimate maximum-likelihood phylogenies: assessing the performance of PhyML 3.0. *Syst Biol*. 59(3):307–321.

Harman A, Manna S. 2016. Identification of Pif1 helicases with novel accessory domains in various amoebae. *Mol Phylogenet Evol*. 103:64–74.

Heringer P, Kuhn GCS. 2018. Exploring the remote ties between Helitron transposases and other rolling-circle replication proteins. *Int J Mol Sci*. 19(10):3079.

Husnik F, McCutcheon JP. 2018. Functional horizontal gene transfer from bacteria to eukaryotes. *Nat Rev Microbiol*. 16(2):67–79.

Kapitonov VV, Jurka J. 2001. Rolling-circle transposons in eukaryotes. *Proc Natl Acad Sci U S A*. 98(15):8714–8719.

Kapitonov VV, Jurka J. 2007. Helitrons on a roll: eukaryotic rolling-circle transposons. *Trends Genet*. 23(10):521–529.

Katoh K, Rozewicki J, Yamada KD. 2019. MAFFT online service: multiple sequence alignment, interactive sequence choice and visualization. *Brief Bioinform*. 20(4):1160–1166.

Kazlauskas D, Varsani A, Koonin EV, Krupovic M. 2019. Multiple origins of prokaryotic and eukaryotic single-stranded DNA viruses from bacterial and archaeal plasmids. *Nat Commun*. 10(1):3425.

Kent TV, Uzunović J, Wright SI. 2017. Coevolution between transposable elements and recombination. *Phil Trans R Soc B*. 372(1736):20160458.

Knoll A, Puchta H. 2011. The role of DNA helicases and their interaction partners in genome stability and meiotic recombination in plants. *J Exp Bot*. 62(5):1565–1579.

Koc KN, Singh SP, Stodola JL, Burgers PM, Galletto R. 2016. Pif1 removes a Rap1-dependent barrier to the strand displacement activity of DNA polymerase δ. *Nucleic Acids Res*. 44(8):3811–3819.

Kohany O, Gentles AJ, Hankus L, Jurka J. 2006. Annotation, submission and screening of repetitive elements in Repbase: repbaseSubmitter and Censor. *BMC Bioinf*. 7:474.

Koonin EV. 2016. Horizontal gene transfer: essentiality and evolvability in prokaryotes, and roles in evolutionary transitions. *F1000Res*. 5:1805.

Kosek D, Grabundzija I, Lei H, Bilic I, Wang H, Jin Y, Peaslee GF, Hickman AB, Dyda F. 2021. The large bat Helitron DNA transposase forms a compact monomeric assembly that buries and protects its covalently bound 5′-transposon end. *Mol Cell*. 81(20):4271–4286.

Kumar S, Stecher G, Li M, Knyaz C, Tamura K. 2018. MEGA X: molecular evolutionary genetics analysis across computing platforms. *Mol Biol Evol*. 35(6):1547–1549.

Lefort V, Longueville JE, Gascuel O. 2017. SMS: smart model selection in PhyML. *Mol Biol Evol*. 34(9):2422–2424. Available from: http://www.atgc-montpellier.fr/phyml/. Accessed November 16, 2021.

Li J-H, Lin W-X, Zhang B, Nong D-G, Ju H-P, Ma J-B, Xu C-H, Ye F-F, Xi XG, Li M, et al. 2016. Pif1 is a force-regulated helicase. *Nucleic Acids Res*. 44(9):4330–4339.

Li HT, Yi TS, Gao LM, Ma PF, Zhang T, Yang JB, Gitzendanner MA, Fritsch PW, Cai J, Luo Y, et al. 2019. Origin of angiosperms and the puzzle of the Jurassic gap. *Nat Plants* 5(5):461–470.

Liu B, Wang J, Yaffe N, Lindsay ME, Zhao Z, Zick A, Shlomai J, Englund PT. 2009. Trypanosomes have six mitochondrial DNA helicases with one controlling kinetoplast maxicircle replication. *Mol Cell* 35(4):490–501.

Lu S, Wang J, Chitsaz F, Derbyshire MK, Geer RC, Gonzales NR, Gwadz M, Hurwitz DI, Marchler GH, Song JS, et al. 2020. CDD/SPARCLE: the conserved domain database in 2020. *Nucleic Acids Res*. 48(D1):D265–D268. Available from: https://www.ncbi.nlm.nih.gov/Structure/cdd/. Accessed November 16, 2021.

Martijn J, Vosseberg J, Guy L, Offre P, Ettema TJ. 2018. Deep mitochondrial origin outside the sampled alphaproteobacteria. *Nature* 557(7703):101–105.

Muellner J, Schmidt KH. 2020. Yeast genome maintenance by the multifunctional PIF1 DNA helicase family. *Genes* 11(2):224.

Pérez-Mendoza D, Lucas M, Munoz S, Herrera-Cervera JA, Olivares J, de la Cruz F, Sanjuán J. 2006. The relaxase of the *Rhizobium etli* symbiotic plasmid shows nic site cis-acting preference. *J Bacteriol*. 188(21):7488–7499.

Pike JE, Burgers PM, Campbell JL, Bambara RA. 2009. Pif1 helicase lengthens some Okazaki fragment flaps necessitating Dna2 nuclease/helicase action in the two-nuclease processing pathway. *J Biol Chem*. 284(37):25170–25180.

Poulter RT, Goodwin TJ, Butler MI. 2003. Vertebrate helentrons and other novel Helitrons. *Gene*. 313:201–212.

R Core Team. 2020. R: A language and environment for statistical computing. Vienna (Austria): R Foundation for Statistical Computing. Available from: https://www.R-project.org/. Accessed November 16, 2021.

Ramanagoudr-Bhojappa R, Chib S, Byrd AK, Aarattuthodiyil S, Pandey M, Patel SS, Raney KD. 2013. Yeast Pif1 helicase exhibits a one-base-pair stepping mechanism for unwinding duplex DNA. *J Biol Chem*. 288(22):16185–16195.

Roger AJ, Muñoz-Gómez SA, Kamikawa R. 2017. The origin and diversification of mitochondria. *Curr Biol*. 27(21):R1177–R1192.

RStudio Team. 2020. RStudio: integrated development for R. RStudio. Boston: PBC. Available from: http://www.rstudio.com/. Accessed November 16, 2021.

Sayers EW, Cavanaugh M, Clark K, Ostell J, Pruitt KD, Karsch-Mizrachi I. 2019. GenBank. *Nucleic Acids Res*. 47(D1):D94–D99. Available from: https://www.ncbi.nlm.nih.gov/genbank/. Accessed November 16, 2021.

Sela I, Wolf YI, Koonin EV. 2016. Theory of prokaryotic genome evolution. *Proc Natl Acad Sci U S A*. 113(41):11399–11407.

Stöver BC, Müller KF. 2010. TreeGraph 2: combining and visualizing evidence from different phylogenetic analyses. *BMC Bioinf*. 11:7.

Thomas J, Pritham EJ. 2015. Helitrons, the eukaryotic rolling-circle transposable elements. *Microbiol Spectr*. 3(4):MDNA3-0049-2014.

Thomas J, Vadnagara K, Pritham EJ. 2014. DINE-1, the highest copy number repeats in *Drosophila melanogaster* are non-autonomous endonuclease-encoding rolling-circle transposable elements (Helentrons). *Mob Dna*. 5:18.

Van Etten J, Bhattacharya D. 2020. Horizontal gene transfer in eukaryotes: not if, but how much? *Trends Genet*. 36(12):915–925.

Wang M, Caetano-Anollés G. 2009. The evolutionary mechanics of domain organization in proteomes and the rise of modularity in the protein world. *Structure* 17(1):66–78.

Wawrzyniak P, Płucienniczak G, Bartosik D. 2017. The different faces of rolling-circle replication and its multifunctional initiator proteins. *Front Microbiol*. 8:2353.

Wilson MA, Kwon Y, Xu Y, Chung WH, Chi P, Niu H, Mayle R, Chen X, Malkova A, Sung P, et al. 2013. Pif1 helicase and Polδ promote recombination-coupled DNA synthesis via bubble migration. *Nature* 502(7471):393–396.

Wolf YI, Makarova KS, Lobkovsky AE, Koonin EV. 2016. Two fundamentally different classes of microbial genes. *Nat Microbiol*. 2:16208.

Xiong W, He L, Lai J, Dooner HK, Du C. 2014. HelitronScanner uncovers a large overlooked cache of Helitron transposons in many plant genomes. *Proc Natl Acad Sci U S A*. 111(28):10263–10268.

Yang L, Bennetzen JL. 2009. Structure-based discovery and description of plant and animal Helitrons. *Proc Natl Acad Sci U S A*. 106(31):12832–12837.

5. CAPÍTULO 3

Multiple horizontal transfers of a *Helitron* transposon associated with a Bracovirus

Pedro Heringer and Gustavo C. S. Kuhn

Departamento de Genética, Ecologia e Evolução, Instituto de Ciências Biológicas, Universidade Federal de Minas Gerais, Belo Horizonte, MG, CEP 31270-901, Brazil.

Multiple horizontal transfers of a *Helitron* transposon associated with a Bracovirus

Pedro Heringer and Gustavo C. S. Kuhn*

Departamento de Genética, Ecologia e Evolução, Instituto de Ciências Biológicas,
Universidade Federal de Minas Gerais, Belo Horizonte, MG, CEP 31270-901, Brazil.

* Corresponding author: Gustavo C. S. Kuhn, gcskuhn@ufmg.br

Abstract

In a previous study we found that a *Helitron* transposon became integrated as a segment in the genome of a symbiotic *Cotesia vestalis* bracovirus (CvBV) from the parasitoid wasp *C. vestalis*. We presented evidence that this *Helitron*, named Hel_c35, initially invaded the *C. vestalis* genome through a horizontal transfer (HT) event from a dipteran species and was later transferred horizontally from *C. vestalis* to a lepidopteran species. We have also anticipated that, as more species would have their genomes sequenced, more HT events involving Hel_c35 might be detected. Here, we investigated the evolution of Hel_c35 in arthropods using a more updated data set to reassess our previous findings. Most species (95%) in the present analysis had their genomes sequenced only after our initial study was published, thus representing new descriptions of taxa harboring Hel_c35. Our results expand considerably the number of putative HTs involving Hel_c35 and suggest that several recent HTs took place in Europe, probably from *C. vestalis* to other insects. We argue that many of these HT events were likely favored by the behavior of this wasp and the stability conferred to Hel_c35 DNA circles by CvBV particles.

Introduction

Horizontal Transfer (HT) events are defined as the exchange of DNA segments between organisms without the involvement of vertical inheritance (Wallau et al. 2018, Van Etten and Bhattacharya 2020). Although HTs are major drivers of evolutionary change in prokaryotes, they are considerably less frequent in eukaryotes, especially in multicellular organisms (Husnik and McCutcheon 2018, Van Etten and Bhattacharya 2020).

In contrast to most genomic components, transposons are DNA segments capable of moving from a locus to another and, as a consequence, they can be found in multiple copies on most eukaryotic genomes, thus being one of the genetic entities most likely to be involved in successful HTs among eukaryotes. Indeed, as the number of eukaryotic sequenced genomes has increased considerably in the last few decades, the number of described examples of horizontal transposon transfers (HTTs) between eukaryotes has also increased, as well as the availability of new bioinformatic methods to detect those events (Schaack et al 2010, Wallau et al. 2018).

We have previously described a *Helitron* transposon from the parasitoid wasp *Cotesia vestalis*, which was found to represent one of the circular segments of the symbiotic virus *C. vestalis* bracovirus (CvBV) (Heringer et al. 2017). This *Helitron* was named Hel_c35, as it was first characterized from the CvBV segment 35 (HQ009558.1). The Hel_c35 has 5,294 bp and appears to be autonomous, containing a 4,538 bp gene encoding its transposase (AEE09607.1) consisting of 1,384 amino acids. In the same work, we showed that, not only this CvBV *Helitron* originated after a HTT event (from a *Drosophila* species to *C. vestalis*), but also that this transposon was later transferred horizontally from *C. vestalis* to the domestic silk moth (*Bombyx mori*). Those HTTs were probably facilitated by the close interactions between *C. vestalis* and its potential hosts, which are mediated by CvBV and a fundamental part of this wasp's life cycle. However, as we anticipated in our study, any HT analysis is subject to a different interpretation in the future as more species with sequenced genomes become available (Heringer et al. 2017).

Here, we reassessed our earlier propositions using an updated data set that includes genomes sequenced more recently, providing both a larger and more diverse sample of species. Our results reveal that Hel_c35 elements can be found in a considerably wider range of arthropod species from different orders than it was previously suggested. Likewise, our analysis indicates that presence of Hel_c35 sequences in a large number of species are most likely the result of HT events. In particular, the investigation of sequences more similar to Hel_c35 elements from *C. vestalis* suggests that several recent putative HTs took place in Europe and were probably facilitated by the parasitoid behavior of this wasp, together with the association between Hel_c35 and CvBV.

Results and Discussion

We Blastn searched sequences similar to Hel_c35 (> 80% identity covering > 70% of the query) in all arthropod genomes available on GenBank (Sayers et al. 2019) using the complete CvBV *Helitron* sequence as a query. A total of 285 sequences from 117 species were retrieved for further analyses (Table S1). Although the vast majority of taxa consisted of Lepidoptera species, several different insect orders and two spider species were found to harbor Hel_c35 sequences.

After aligning all the retrieved Hel_c35 sequences, we conducted a phylogenetic analysis using the Maximum Likelihood method. The resulting phylogeny shows that Hel_c35 sequences from specific taxa (insect order or Lepidoptera superfamily) are mostly scattered across different branches, instead of representing the overall topology expected from their evolutionary relationships (Fig. 1). In addition, although several lepidopterans from the same superfamily grouped together, many of those clades contain species from distinct families. At the same time, taxa from the same family were found in separate clades, even though they were grouped with species from the same superfamily (Fig. S1, Table S1).

Despite the diversity and incongruent topology observed in the resulting phylogeny, its Hel_c35 sequences have > 80% sequence identity, what would place its earliest origin at ~ 33 million years ago (MYA), assuming that this transposon evolves neutrally. This diverge time is at least 15 times more recent than the one estimated for the split between arachnids and insects (> 500 MYA) (Kumar et al. 2017) and several times more recent than the estimated time of divergence between most insect orders (Misof et al. 2014). The patchy distribution of taxa, together with the marked deviation between observed and expected divergence times among sequences, strongly indicate that Hel_c35 has been involved in multiple HTT events during its evolution.

Given the large number of sequences included in our phylogeny, we decided to focus our analysis in the main clade containing the CvBV Hel_c35 sequence (zoomed in clade on Fig. 1). This well supported clade (SH-aLRT branch support = 0.95) (see also Fig. S1 for support values) contains species from seven insect orders, along with a variety of lepidopteran species from 6 different superfamilies. Similarly to the phylogeny as a whole, most of this clade topology does not reflect the evolutionary relationships between species. Moreover, the estimated evolutionary distances between many sequences in this clade (Table S2) also strongly deviate from their expected divergence times. For example, the cat flea *Ctenocephalides felis* and *Drosophila ficusphila* were the two species with the largest number of pairwise differences per site (0.0751) between their Hel_c35 sequences. Using a conservative assumption of one generation per year for all species, this clade would have originated ~ 12.5 MYA, which strongly contrasts with the estimated divergence time between most taxa included in this clade. For instance, *C. felis* and *D. ficusphila* are estimated to have diverged > 200 MYA, and all Lepidoptera species are estimated to have diverged from *Gryllus bimaculatus* (Orthoptera) > 300 MYA (Kumar et al. 2017). In both examples, if Hel_c35 has been exclusively evolving neutrally and being inherited vertically, no sequence homology would be expected in Hel_c35 copies between groups. This contrasts strikingly with the observed sequence nucleotide identity > 92% between all sequences in this clade.

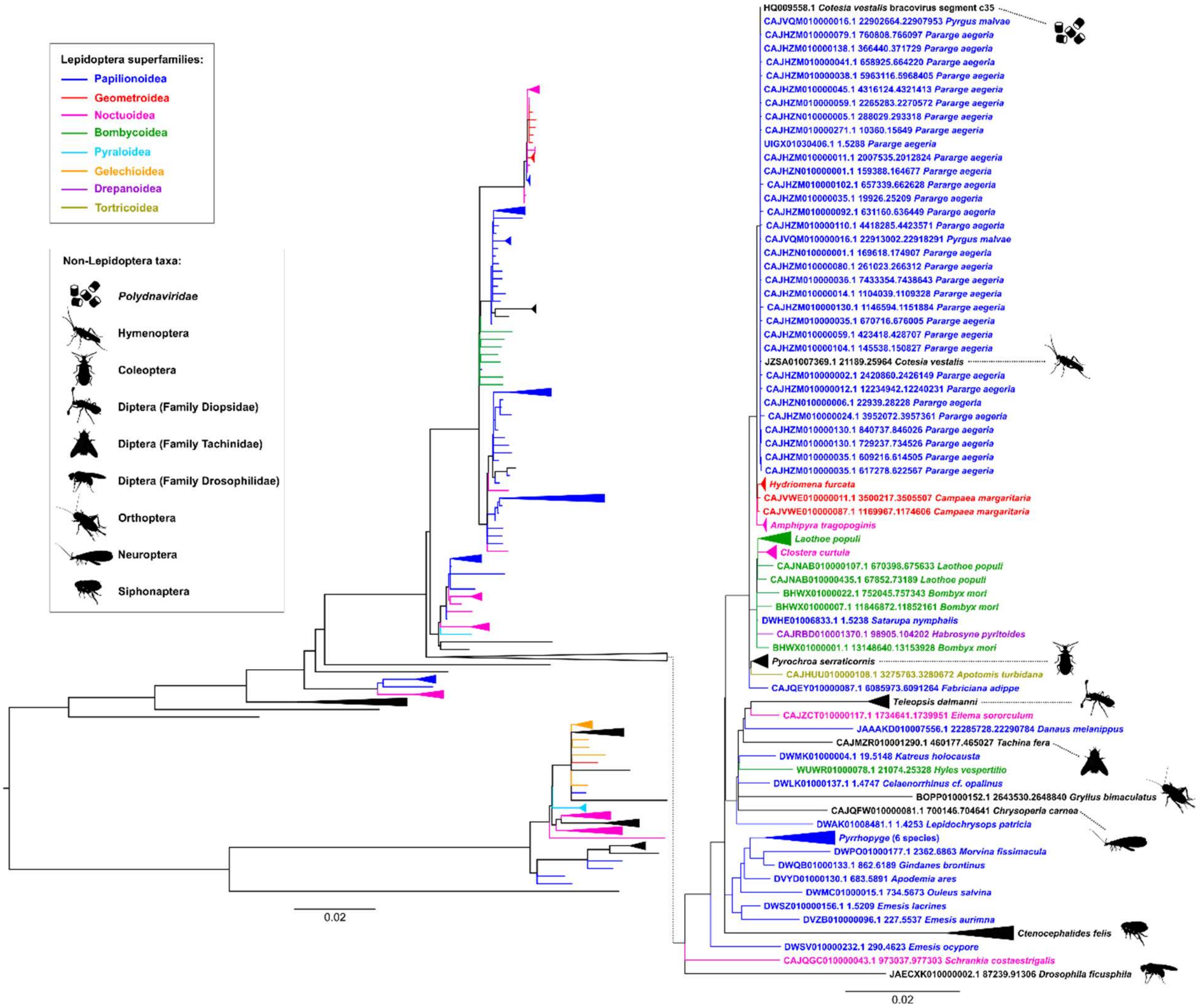

Lepidoptera superfamilies:
Papilionoidea
Geometroidea
Noctuoidea
Bombycoidea
Pyraloidea
Gelechioidea
Drepanoidea
Tortricoidea
Non-Lepidoptera taxa:
Polydnaviridae
Hymenoptera
Coleoptera
Diptera (Family Diopsidae)
Diptera (Family Tachinidae)
Diptera (Family Drosophilidae)
Orthoptera
Neuroptera
Siphonaptera
0.02
HQ009558.1 Cotesia vestalis bracovirus segment c35
CAJVQM010000016.1 22902664.22907953 Pyrgus malvae
CAJHZM010000079.1 760808.766097 Pararge aegeria
CAJHZM010000138.1 366440.371729 Pararge aegeria
CAJHZM010000041.1 658925.664220 Pararge aegeria
CAJHZM010000038.1 5963116.5968405 Pararge aegeria
CAJHZM010000045.1 4316124.4321413 Pararge aegeria
CAJHZM010000059.1 2265283.2270572 Pararge aegeria
CAJHZN010000005.1 288029.293318 Pararge aegeria
CAJHZM010000271.1 10360.15649 Pararge aegeria
UIGX01030406.1 1.5288 Pararge aegeria
CAJHZM010000011.1 2007535.2012824 Pararge aegeria
CAJHZN010000001.1 159388.164677 Pararge aegeria
CAJHZM010000102.1 657339.662628 Pararge aegeria
CAJHZM010000035.1 19926.25209 Pararge aegeria
CAJHZM010000092.1 631160.636449 Pararge aegeria
CAJHZM010000110.1 4418285.4423571 Pararge aegeria
CAJVQM010000016.1 22913002.22918291 Pyrgus malvae
CAJHZN010000001.1 169618.174907 Pararge aegeria
CAJHZM010000080.1 261023.266312 Pararge aegeria
CAJHZM010000036.1 7433354.7438643 Pararge aegeria
CAJHZM010000014.1 1104039.1109328 Pararge aegeria
CAJHZM010000130.1 1146594.1151884 Pararge aegeria
CAJHZM010000035.1 670716.676005 Pararge aegeria
CAJHZM010000059.1 423418.428707 Pararge aegeria
CAJHZM010000104.1 145538.150827 Pararge aegeria
JZSA01007369.1 21189.25964 Cotesia vestalis
CAJHZM010000002.1 2420860.2426149 Pararge aegeria
CAJHZM010000012.1 12234942.12240231 Pararge aegeria
CAJHZN010000006.1 22939.28228 Pararge aegeria
CAJHZM010000024.1 3952072.3957361 Pararge aegeria
CAJHZM010000130.1 840737.846026 Pararge aegeria
CAJHZM010000130.1 729237.734526 Pararge aegeria
CAJHZM010000035.1 609216.614505 Pararge aegeria
CAJHZM010000035.1 617278.622567 Pararge aegeria
Hydriomena furcata
CAJVWE010000011.1 3500217.3505507 Campaea margaritaria
CAJVWE010000087.1 1169967.1174606 Campaea margaritaria
Amphipyra tragopoginis
Laothoe populi
Clostera curtula
CAJNAB010000107.1 670398.675633 Laothoe populi
CAJNAB010000435.1 67852.73189 Laothoe populi
BHWX01000022.1 752045.757343 Bombyx mori
BHWX01000007.1 11846872.11852161 Bombyx mori
DWHE01006833.1 1.5238 Satarupa nymphalis
CAJRBD010001370.1 98905.104202 Habrosyne pyritoides
BHWX01000001.1 13148640.13153928 Bombyx mori
Pyrochroa serraticornis
CAJHUU010000108.1 3275763.3280672 Apotomis turbidana
CAJQEY010000087.1 6085973.6091264 Fabriciana adippe
Teleopsis dalmanni
CAJZCT010000117.1 1734641.1739951 Eilema sororculum
JAAAKD010007556.1 22285728.22290784 Danaus melanippus
CAJMZR010001290.1 460177.465027 Tachina fera
DWMK01000004.1 19.5148 Katreus holocausta
WUWR01000078.1 21074.25328 Hyles vespertilio
DWLK01000137.1 1.4747 Celaenorrhinus cf. opalinus
BOPP01000152.1 2643530.2648840 Gryllus bimaculatus
CAJQFW010000081.1 700146.704641 Chrysoperla carnea
DWAK01008481.1 1.4253 Lepidochrysops patricia
Pyrrhopyge (6 species)
DWPO01000177.1 2362.6863 Morvina fissimacula
DWQB01000133.1 862.6189 Gindanes brontinus
DVYD01000130.1 683.5891 Apodemia ares
DWMC01000015.1 734.5673 Ouleus salvina
DWSZ010000156.1 1.5209 Emesis lacrines
DVZB010000096.1 227.5537 Emesis aurimna
Ctenocephalides felis
DWSV010000232.1 290.4623 Emesis ocypore
CAJQGC010000043.1 973037.977303 Schrankia costaestrigalis
JAECXK010000002.1 87239.91306 Drosophila ficusphila
0.02

Figure 1. Phylogeny of Hel_c35 sequences. Maximum Likelihood phylogeny including all 285 Hel_c35 sequences retrieved from arthropod genomes is represented on the left. A clade containing sequences closely related to the CvBV Hel_c35 is featured on the right. Lepidoptera species from different superfamilies are represented by different colors. Non-lepidopteran arthropods are represented in black. Branches with < 0.7 SH-aLRT statistical support were collapsed. The same phylogeny with branch supports and all taxa names is shown on Fig. S1.

A deviation from the expected pairwise nucleotide differences per site between species is even more pronounced in the clade comprising taxa with sequences more closely related to the CvBV Hel_c35 (zoomed in clade on Fig. 2). All sequences in this proximal clade have > 99% identity between each other, even though they include species from 3 insect orders that diverged up to > 300 MYA (e.g., Hymenoptera and Lepidoptera) and 6 Lepidoptera superfamilies that diverged up to > 100 MYA (e.g., Bombycoidea and Tortricoidea) (Kumar et al. 2017). Considering the largest value of pairwise nucleotide differences per site among taxa in this clade, which is found between *Apotomis turbidana* and *Habrosyne pyritoides*

(0.009830), its earliest origin would be ~ 1.64 MYA, in contrast to the estimated divergence time for some species included, which are higher by up to two orders of magnitude.

Some of the most conspicuous examples of recent HTTs are shown on the clade containing the Hel_c35 sequences from *C. vestalis* (including CvBV), *Pararge aegeria* and *Pyrgus malvae* (Fig. 1). The phylogenetic relationships between these three species are represented as a polytomy containing sequences with > 99.95% identity, what puts its earliest date of origin at 0.068 MYA (68 thousand years ago). Considering that *P. aegeria* and *P. malvae* diverged > 70 MYA and these two Lepidoptera species have diverged from *C. vestalis* > 300 MYA, these values are at least three orders of magnitude higher than the maximum estimated divergence time for Hel_c35 sequences in this clade.

Even though the phylogenetic topology and level of identity between Hel_c35 sequences strongly suggest the occurrence of multiple HTTs, these events also require some degree of geographic overlap between species to be inferred (Loreto et al. 2008). To verify if the geographical distribution of the analyzed species provides further evidence for HTT events, we represented our phylogeny by color coding the taxa according to the geographical locations where the species were sampled. Sample locations were assigned into one of seven regions defined by their biogeographic realm, bioregions and/or expected migration barriers. Several topological incongruencies consisting of distantly related taxa grouping together on Figure 1 represent species sampled in the same region (Fig. 2 and Fig. S2), indicating that, in some cases, the geographical distribution of species appears to better explain the phylogenetic relationships of their Hel_c35 sequences.

Interestingly, 11 from the 14 species belonging to the CvBV immediate clade correspond to samples from Europe (zoomed in clade on Fig. 2). From those 11 species, nine were derived from the island of Great Britain (Table S1). The remaining two species were collected in Romania (*P. malvae* and *Fabriciana adippe*) but can also be found in Great Britain (Butterfly Conservation 2022a, 2022b). Although the *C. vestalis* samples used in our analysis derive from East Asia (China and South Korea), this wasp species can also be found in several European countries (Furlong et al. 2013), including Great Britain (Broad et al. 2016). Hence, 12 out of 14 species in this clade containing the CvBV Hel_c35 sequence overlap geographically in Great Britain, indicating this island as the most probable region where those HTT events occurred.

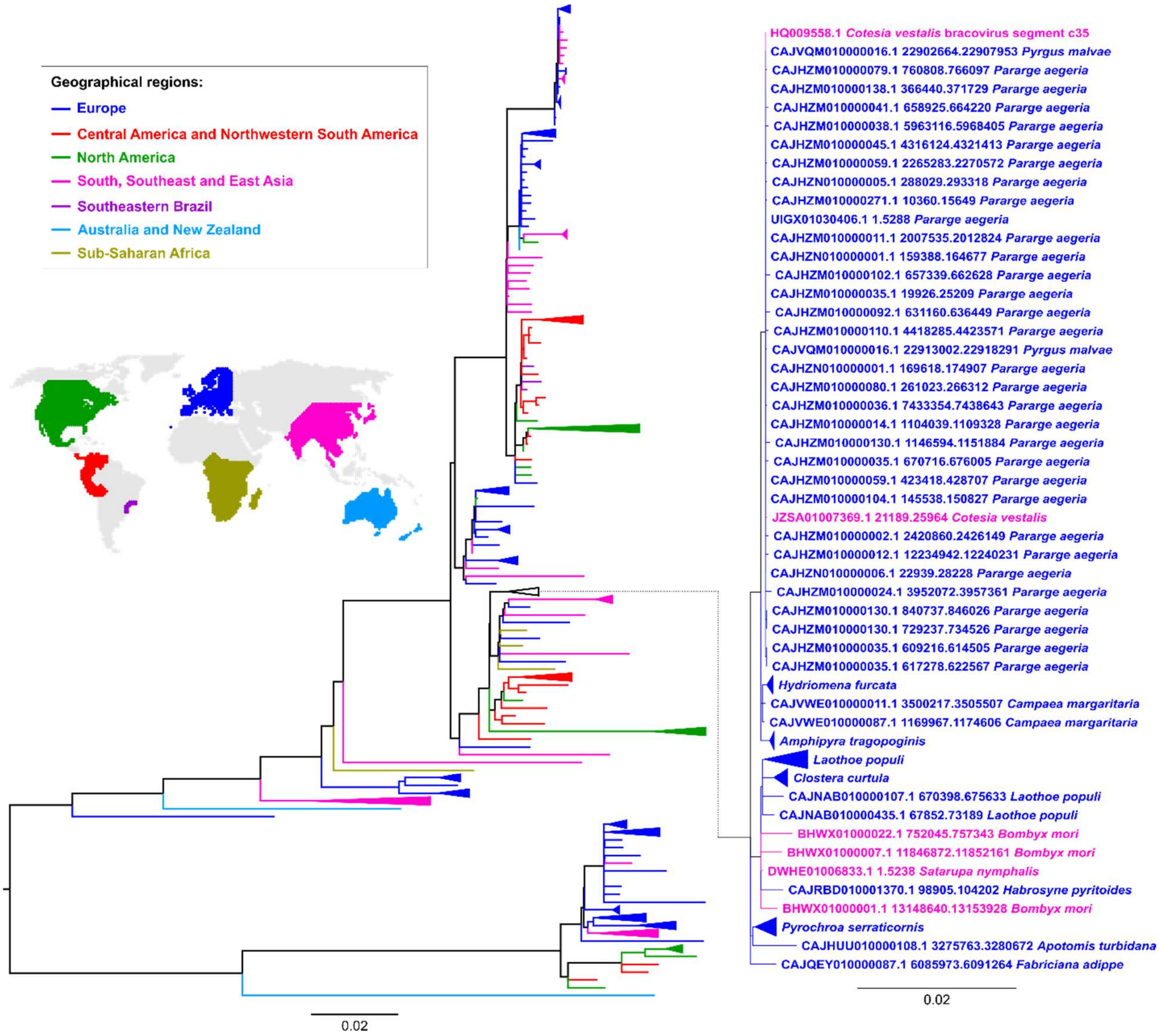

Geographical regions:
Europe
Central America and Northwestern South America
North America
South, Southeast and East Asia
Southeastern Brazil
Australia and New Zealand
Sub-Saharan Africa
HQ009558.1 Cotesia vestalis bracovirus segment c35
CAJVQM010000016.1 22902664.22907953 Pyrgus malvae
CAJHZM010000079.1 760808.766097 Pararge aegeria
CAJHZM010000138.1 366440.371729 Pararge aegeria
CAJHZM010000041.1 658925.664220 Pararge aegeria
CAJHZM010000038.1 5963116.5968405 Pararge aegeria
CAJHZM010000045.1 4316124.4321413 Pararge aegeria
CAJHZM010000059.1 2265283.2270572 Pararge aegeria
CAJHZN010000005.1 288029.293318 Pararge aegeria
CAJHZM010000271.1 10360.15649 Pararge aegeria
UIGX01030406.1 1.5288 Pararge aegeria
CAJHZM010000011.1 2007535.2012824 Pararge aegeria
CAJHZN010000001.1 159388.164677 Pararge aegeria
CAJHZM010000102.1 657339.662628 Pararge aegeria
CAJHZM010000035.1 19926.25209 Pararge aegeria
CAJHZM010000092.1 631160.636449 Pararge aegeria
CAJHZM010000110.1 4418285.4423571 Pararge aegeria
CAJVQM010000016.1 22913002.22918291 Pyrgus malvae
CAJHZN010000001.1 169618.174907 Pararge aegeria
CAJHZM010000080.1 261023.266312 Pararge aegeria
CAJHZM010000036.1 7433354.7438643 Pararge aegeria
CAJHZM010000014.1 1104039.1109328 Pararge aegeria
CAJHZM010000130.1 1146594.1151884 Pararge aegeria
CAJHZM010000035.1 670716.676005 Pararge aegeria
CAJHZM010000059.1 423418.428707 Pararge aegeria
CAJHZM010000104.1 145538.150827 Pararge aegeria
JZSA01007369.1 21189.25964 Cotesia vestalis
CAJHZM010000002.1 2420860.2426149 Pararge aegeria
CAJHZM010000012.1 12234942.12240231 Pararge aegeria
CAJHZN010000006.1 22939.28228 Pararge aegeria
CAJHZM010000024.1 3952072.3957361 Pararge aegeria
CAJHZM010000130.1 840737.846026 Pararge aegeria
CAJHZM010000130.1 729237.734526 Pararge aegeria
CAJHZM010000035.1 609216.614505 Pararge aegeria
CAJHZM010000035.1 617278.622567 Pararge aegeria
Hydriomena furcata
CAJVWE010000011.1 3500217.3505507 Campaea margaritaria
CAJVWE010000087.1 1169967.1174606 Campaea margaritaria
Amphipyra tragopoginis
Laothoe populi
Clostera curtula
CAJNAB010000107.1 670398.675633 Laothoe populi
CAJNAB010000435.1 67852.73189 Laothoe populi
BHWX01000022.1 752045.757343 Bombyx mori
BHWX01000007.1 11846872.11852161 Bombyx mori
DWHE01006833.1 1.5238 Satarupa nymphalis
CAJRBD010001370.1 98905.104202 Habrosyne pyritoides
BHWX01000001.1 13148640.13153928 Bombyx mori
Pyrochroa serraticornis
CAJHUU010000108.1 3275763.3280672 Apotomis turbidana
CAJQEY010000087.1 6085973.6091264 Fabriciana adippe
0.02
0.02

Figure 2. Geographical distribution of arthropod species containing Hel_c35. The same phylogeny of Hel_c35 sequences from Fig. 1 is represented, but with colors corresponding to the geographical location where the species were sampled (Table S1). A clade with species containing sequences closely related to CvBV Hel_c35 is featured expanded on the right. The same phylogeny with branch supports and all taxa names is shown on Fig. S2.

Although it is difficult to infer the direction of HTTs, the diversity of Lepidoptera superfamilies at the base of most clades suggests that species in this order are the earliest donors of horizontally transferred Hel_c35 sequences. However, considering the large number of potential HTTs in the presented phylogeny, it is also possible that Lepidoptera species could have received Hel_c35 sequences by secondary HTT events. For example, a HTT from a lepidopteran to a dipteran, which later transferred this transposon to another Lepidoptera species. The diversity and broad distribution of dipterans in the phylogeny (Fig. 1 and Fig. S1) indicate that species from this order were also basal donors of Hel_c35 elements. Nonetheless, because of mechanical and physiological constraints, direct HTs between insects should be considered rare events. In those cases, it is reasonable to expect the involvement of species like *C. vestalis* as likely HT vectors or intermediates, due to their life history which is thought to facilitate those events (Schaack et al 2010, Wallau et al. 2018). That is particularly relevant for the putative HTTs in the clades more closely related to the CvBV Hel_c35 (Fig. 1). This *Helitron* appears to be autonomous and its copies are likely protected by a viral capsid and envelope when injected every time *C. vestalis* lay eggs in its potential hosts (Heringer et al. 2017). Hence, we suggest a preferred direction for those specific HTT events, which is from the parasitoid to other species.

Considering the topology revealed by our phylogenetic analysis, the geographical distribution of the species and their natural history, we suggest the following hypothesis to explain the putative HTTs involving sequences more closely related to the *C. vestalis* Hel_c35 element. The originally Palearctic/eastern Asian distribution of *C. vestalis* (Hiroyoshi et al. 2017) and several other lepidopterans and dipterans harboring closely related Hel_c35 sequences (Heringer et al. 2017) indicates that *C. vestalis* acquired Hel_c35 by HT from an insect species within those orders, less than 12.5 MYA. In our previous work (Heringer et al. 2017) we suggested a drosophilid as the most probable donor of the *C. vestalis* Hel_c35, given the evidence available at the time. Although our results showing eastern Asian drosophilids near the base of the CvBV Hel_c35 clade provide some support for that hypothesis, we cannot reject that lepidopterans from the same geographical region could also have been potential donors. In any case, after this HTT event, a Hel_c35 sequence became one of CvBV segments, which in turn facilitated other HTTs from *C. vestalis* to multiple species from several insect orders (Fig. 3).

Lepidoptera species are overrepresented in our phylogeny, what could indicate a genome sequencing bias favoring this order. On the other hand, this could likewise be a consequence of lepidopterans being more frequently attacked by parasitoid wasps. This feature might be particularly relevant to explain the putative HTTs indicated in the immediate clade containing the CvBV Hel_c35 sequence (Fig. 1). Despite being considered a specialist parasitoid of the diamondback moth (*Plutella xylostella*), *C. vestalis* is known to attack lepidopterans from at least ten different families within eight superfamilies (Hiroyoshi et al. 2017). In view of the high diversity of lepidopteran larvae that can be targeted by *C. vestalis*, it is reasonable to expect that unspecific attacks to larvae from other insect orders could also occur in some conditions, even if rarely. In fact, the diversity of insect orders found in the main clade containing *C. vestalis*/CvBV in itself might be considered as evidence for the occurrence of those unspecific attacks. As we previously suggested (Heringer et al. 2017), the detection of HTs involving parasitoid wasps and species outside the known range of hosts targeted by those wasps could be used to indicate potential cryptic interactions to be confirmed in future ecological and behavioral studies.

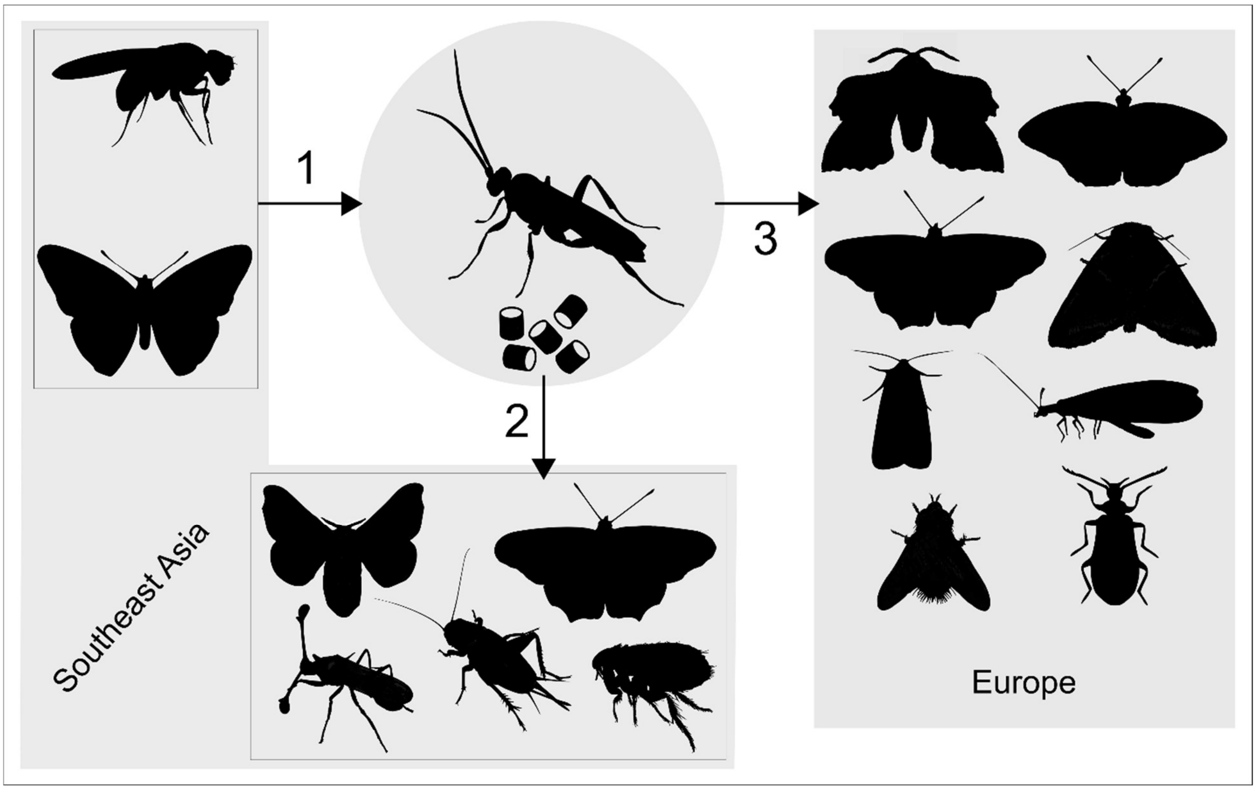


Figure 3. Hypothesis for HTTs involving Hel_c35 sequences closely related to the one found in CvBV. Arrows represent the probable direction of HTTs and numbers indicate the order which most HTTs events in each geographical region occurred. The earliest event from a Diptera or Lepidoptera species to *C. vestalis* and CvBV (1) was followed by HTTs from CvBV to multiple insects from several orders, initially to species found in Southeast Asia (2) and more recently to species from Europe (3). Although most HTTs in 2 appear to have occurred earlier than those in 3, some European species are interspersed with, or more basal in relation to some Southeast Asian species, indicating that this chronological division is not clear cut.

Overall, the results presented here differ from our previous findings (Heringer et al. 2017) in some important aspects. Firstly, the single best hits from 24 species were retrieved in our earlier work, as opposed to the current analysis, in which 285 sequences from 117 species were included, even though we used a more stringent selection criteria in the latter. For instance, here we considered the same minimum query coverage (> 70%) and identity (> 80%) as previously, but using the whole CvBV Hel_c35 (5,294 bp) as a reference, as opposed to a region of ~ 838 bp only containing the Rep coding sequence. The sampled species in our former analysis belonged to five insect orders, with one spider species, in contrast to the current sample that comprises 117 species from eight insect orders and two spider families. Therefore, not only the resulting data set presented here is larger, but is also more diverse. It is also worthwhile mentioning that using this more stringent sequence selection criteria, only five out of 24 species from the previous analysis were included in the current study. Only one of the new species included in the present data set (*Heliconius wallacei*) had its genome sequence already available before the previous study was conducted (November 2015), although we cannot explain the reason for this absence. The remaining 111 new species all had their genome sequences made available only after our previous work (Heringer et al 2017) was submitted (September 2017) and represent 95% of the current data set (Table S1).

The larger number of species in the present analysis revealed a more complex scenario regarding the evolutionary history of Hel_c35 sequences more closely related to the one found in *C. vestalis*. We previously suggested that East/Southeast Asia was probably the geographical region in which the most recent HTTs of Hel_c35 involving *C. vestalis* had occurred (Heringer et al. 2017). Although the evidence provided here still is consistent with a scenario in which the *C. vestalis* Hel_c35 originated from a HTT that probably occurred in East/Southeast Asia < 12.5 MYA, our current results also indicate that this *Helitron* was probably horizontally transferred more recently to multiple insect species in Europe in the last few million years. In spite of those significant differences, our results presented here confirm the previous hypothesis that, as new genome sequencing projects would become available, new HT events would probably be detected, resulting in new interpretations about the evolution of Hel_c35.

Given the large amount of putative HTTs involving *C. vestalis* as a donor of Hel_c35 sequences to other species, and the evidence for CvBV being an important promoter of these events, we consider that future sequencing *C. vestalis* and/or CvBV genomes from different lineages and geographical locations will be essential to confirm our proposed scenario. For instance, we expect that if Hel_c35 copies turn out to be absent in genomes from European lineages of *C. vestalis*, our main hypotheses regarding the direction and geographical location of the most recent HTTs would be refuted, at least partially. Likewise, an absence of Hel_c35 in CvBV genomes from outside East Asia would contradict our suggestion that CvBV has been a major HTT vector of Hel_c35 copies.

Materials and Methods

We Blastn searched all arthropod genomes available (as in October 2021) on the Whole Genome Shotgun (WGS) contigs database from GenBank (Sayers et al. 2019) using the Hel_c35 sequence from CvBV (HQ009558.1) as a query. In order to include only highly similar elements in our analysis, we downloaded all Blast aligned sequences from hits with > 80% sequence identity covering > 70% of the query. Those downloaded hits are sometimes composed by multiple separate matches which, together, cover > 70% of the query, instead of continuous sequences with the minimum query cover size. Hence, to include only sequences covering > 70% (3,705 bp) of the query, we adapted a Biopython (Cock et al. 2009) script for that purpose and also to edit FASTA sequence descriptions in order to contain only the hit accession number, the sequence match range and the species name (Data S1). The resulting 285 sequences (Data S2) were aligned using the E-INS-i method in the MAFFT online service (Katoh et al. 2019). For the phylogenetic analysis, the best-fit evolutionary model (GTR+G+I) was selected using the Smart Model Selection (SMS) in PhyML (Lefort et al. 2017). The maximum likelihood phylogeny of sequences was inferred using the best topology from NNI and SPR methods, six random plus one parsimony starting trees and 10 substitution rate categories across sites, modelled with estimated gamma-shaped distribution parameter and a proportion of invariant sites. Branch supports were estimated using the approximate likelihood ratio test (aLRT) with the nonparametric Shimodaira–Hasegawa correction (SH-aLRT). The phylogenetic analysis procedures described above were conducted on PhyML 3.1 (Guindon et al. 2010). All branches with < 0.7 SH-aLRT statistical support were collapsed using TreeGraph 2 (Stöver and Müller 2010), with the final tree edited and visualized using FigTree v.1.4.2 (http://tree.bio.ed.ac.uk/software/figtree/; last accessed December 15, 2022). The species taxonomy and sample collection locations were obtained from their corresponding accession on GenBank (Sayers et al. 2019), and additional information about the geographical distribution of organisms included in our analysis was obtained from various Web sources. The average nucleotide differences per site between groups in the main clade containing CvBV Hel_c35 (Table S2) was calculated using MEGA X (Kumar et al. 2018), and their divergence time estimated using the equation:

$$T = \frac{K}{2r}$$

in which T is the number of generations, K is the number of substitutions per site, and r is the rate of nucleotide substitution. We considered that r is equal to the mutation rate (μ), as expected for neutral mutations (Graur and Li 2000), and a value of μ equal to 3.0 x 10^{-9} for insect species (Liu et al. 2017). To obtain a conservative estimation for the maximum time of divergence between sequences we considered one generation per year for all insect species. Hence, in our equation, the value found for T is equal to the diverge time between species given in number of years.

References

Broad, G. R., Shaw, M. R., Godfray, H. C. J. (2016). Checklist of British and Irish Hymenoptera - Braconidae. Biodiversity Data Journal, 4:e8151.

Butterfly Conservation. (2022a). Grizzled Skipper, Pyrgus malvae. Available at: https://butterflyconservation.org/butterflies/grizzled-skipper (last accessed January 22, 2022).

Butterfly Conservation. (2022b). High Brown Fritillary, Fabriciana adippe. Available at: https://butterfly-conservation.org/butterflies/high-brown-fritillary (last accessed January 22, 2022).

Cock, P. J., Antao, T., Chang, J. T., Chapman, B. A., Cox, C. J., Dalke, A., ... & De Hoon, M. J. (2009). Biopython: freely available Python tools for computational molecular biology and bioinformatics. Bioinformatics, 25(11), 1422-1423.

Furlong, M. J., Wright, D. J., & Dosdall, L. M. (2013). Diamondback moth ecology and management: problems, progress, and prospects. Annual review of entomology, 58, 517-541.

Graur, D., and W.-H. Li. 2000. Fundamentals of Molecular Evolution, Ed. 2. Sinauer Associates, Sunderland, MA.

Guindon S, Dufayard JF, Lefort V, Anisimova M, Hordijk W, Gascuel O. 2010. New algorithms and methods to estimate maximum-likelihood phylogenies: assessing the performance of PhyML 3.0. Syst Biol. 59(3):307-321.

Heringer, P., Dias, G. B., & Kuhn, G. C. (2017). A horizontally transferred autonomous Helitron became a full polydnavirus segment in Cotesia vestalis. G3: Genes, Genomes, Genetics, 7(12), 3925-3935.

Hiroyoshi, S., Harvey, J. A., Nakamatsu, Y., Nemoto, H., Mitsuhashi, J., Mitsunaga, T., & Tanaka, T. (2017). Potential host range of the larval endoparasitoid Cotesia vestalis (= plutellae) (Hymenoptera: Braconidae). International Journal of Insect Science, 9. https://doi.org/10.1177/1179543317715623

Husnik, F., & McCutcheon, J. P. (2018). Functional horizontal gene transfer from bacteria to eukaryotes. Nature Reviews Microbiology, 16(2), 67-79.

Katoh, K., Rozewicki, J., & Yamada, K. D. (2019). MAFFT online service: multiple sequence alignment, interactive sequence choice and visualization. Briefings in bioinformatics, 20(4), 1160-1166.

Kumar, S., Stecher, G., Suleski, M., & Hedges, S. B. (2017). TimeTree: a resource for timelines, timetrees, and divergence times. Molecular biology and evolution, 34(7), 1812-1819.

Kumar, S., Stecher, G., Li, M., Knyaz, C., & Tamura, K. (2018). MEGA X: molecular evolutionary genetics analysis across computing platforms. Molecular biology and evolution, 35(6):1547-1549.

Lefort V, Longueville JE, Gascuel O. 2017. SMS: smart model selection in PhyML. Mol Biol Evol. 34(9):2422-2424. (Available from: http://www.atgc-montpellier.fr/phyml/).

Liu, H., Jia, Y., Sun, X., Tian, D., Hurst, L. D., & Yang, S. (2017). Direct determination of the mutation rate in the bumblebee reveals evidence for weak recombination-associated mutation and an approximate rate constancy in insects. Molecular biology and evolution, 34(1), 119-130.

Loreto, E. L. S., Carareto, C. M. A., & Capy, P. (2008). Revisiting horizontal transfer of transposable elements in Drosophila. Heredity, 100, 545-554.

Misof, B., Liu, S., Meusemann, K., Peters, R. S., Donath, A., Mayer, C., ... & Zhou, X. (2014). Phylogenomics resolves the timing and pattern of insect evolution. Science, 346(6210), 763-767.

Sayers EW, Cavanaugh M, Clark K, Ostell J, Pruitt KD, Karsch-Mizrachi I. 2019. GenBank. Nucleic Acids Res. 47(D1):D94-D99. Available from: https://www.ncbi.nlm.nih.gov/genbank/. Accessed October 30, 2021.

Schaack, S., Gilbert, C., & Feschotte, C. (2010). Promiscuous DNA: horizontal transfer of transposable elements and why it matters for eukaryotic evolution. Trends in ecology & evolution, 25(9), 537-546.

Stöver BC, Müller KF. 2010. TreeGraph 2: combining and visualizing evidence from different phylogenetic analyses. BMC Bioinf. 11:7.

Van Etten, J., Bhattacharya, D. (2020). Horizontal gene transfer in eukaryotes: not if, but how much?. Trends in Genetics. 36(12):915–925.

Wallau, G. L., Vieira, C., & Loreto, É. L. S. (2018). Genetic exchange in eukaryotes through horizontal transfer: connected by the mobilome. Mobile DNA, 9:6.

6. DISCUSSÃO GERAL

Os resultados apresentados no Capítulo 1 sugerem que, a despeito de os *Helitrons* serem transposons exclusivamente eucarióticos, estes elementos pertencem a uma linhagem filogenética de replicons tipicamente procarióticos. Apesar de os *Helitrons* terem uma origem procariótica, modo de replicação por transposição e possuírem uma proteína com atividade enzimática semelhante às encontradas em transposons procarióticos que utilizam RCT, nossos resultados indicam que *Helitrons* não são parentes próximos destes últimos. Por outro lado, a hipótese de que *Helitrons* seriam descendentes ou mesmo teriam dado origem a vírus eucarióticos do tipo RCR também não é sustentada pelos resultados das nossas análises utilizando o domínio Rep.

Ao contrário, nossos dados indicam que *Helitrons* são parentes são mais proximamente relacionados a plasmídeos e vírus procarióticos, formando com estes um grupo filogenético composto por elementos circulares que se replicam por RCR e possuem duas tirosinas catalíticas no seu domínio Rep. Após sua publicação (Heringer & Kuhn 2018), estes resultados foram corroborados por um estudo independente que analisou as relações evolutivas entre proteínas Rep de elementos procarióticos e eucarióticos (Kazlauskas et al. 2019).

Já os resultados no Capítulo 2 argumentam contra a hipótese de que os *Helitrons* teriam adquirido seu domínio Hel após a captura de uma helicase Pif1 eucariótica. Apesar de helicases Pif1 serem tipicamente codificadas por genomas de eucariotos, esta família de proteínas também é encontrada em diversos genomas de arqueias e bactérias, além de vírus eucarióticos e procarióticos. A distribuição filogenética das proteínas analisadas demonstra que o domínio Hel evoluiu independentemente de da linhagem que deu origem a helicases Pif1 eucarióticas, indicando que *Helitrons* já possuíam uma transposase contendo seus dois domínios antes de invadirem seus primeiros hospedeiros eucariotos.

Sugerimos que *Helitrons* representam um grupo de plasmídeos procarióticos que, após invadirem organismos eucariotos, passaram a se replicar por transposição nos genomas de seus hospedeiros (Fig. 6 do Cap. 2). Esta hipótese se baseia no conjunto de dados revelados no presente trabalho e em outros estudos, sendo estas evidências apresentadas a seguir. Primeiramente, apesar de terem se tornado transposons, *Helitrons* geram intermediários de dsDNA circulares para se mover no genoma (Grabundzija et al. 2018). Além disso, estes elementos possuem em sua transposase um domínio Rep mais proximamente relacionado com proteínas de vírus circulares e plasmídeos (Cap. 1, Kazlauskas et al. 2019).

Por outro lado, o domínio helicase presente em relaxases TraA de plasmídios parece ser filogeneticamente relacionado à família Pif1, que inclui o domínio Hel (Cap. 2). Apesar de remota, esta relação sugere que *Helitrons* poderiam representar parentes distantes de plasmídeos atuais. Mesmo considerando que a similaridade entre a transposase RepHel e a relaxase TraA provavelmente resulta de convergência evolutiva, tal fato ainda indicaria a existência de paralelos entre os processos enzimáticos conduzidos por estas duas proteínas distintas. Recentemente, a estrutura da RepHel associada à extremidade 5’ ssDNA do *Helitron* foi resolvida por crio-microscopia eletrônica, revelando que esta transposase apresenta uma estrutura tridimensional notavelmente similar a encontrada na relaxase TraI (Kosek et al. 2021). Assim como a semelhança na sequência de aminoácidos observada para o caso da relaxase TraA, a similaridade estrutural entre TraI e RepHel muito provavelmente resulta de convergência evolutiva pelo fato de ambas as proteínas desempenharem reações catalíticas análogas.

Apesar de possuírem características de plasmídeos, o conjunto de resultados apresentados nos dois primeiros capítulos indicam que cada um dos dois domínios principais da transposase RepHel se assemelha mais a proteínas encontradas em elementos genéticos móveis de grupos distintos. De um lado, o domínio Rep claramente pertence a um grupo de proteínas responsáveis pela replicação de plasmídeos e vírus procarióticos do tipo RCR (Cap. 1); do outro, o domínio Hel representa um dos clados mais basais de helicases Pif1 (Cap. 2). De fato, a divergência dos dois grandes grupos do domínio Hel (*Helitrons* e *Helentrons*/*Helitron2*) parece ser tão antiga quanto as principais radiações basais de proteínas semelhantes a helicases Pif1. A profundidade desta divergência evolutiva entre domínios Hel de *Helitrons* e *Helentrons*/*Helitron2* é tão acentuada que domínios Hel sequer formam grupos monofiléticos na nossa análise (Fig. 3 e Fig. S2 do Cap. 2).

Por fim, no Capítulo 3 exemplificamos a capacidade que os *Helitrons* possuem de se propagar horizontalmente cruzando a barreira das espécies, muitas vezes entre organismos de ordens ou mesmo classes diferentes. Em um trabalho anterior (Heringer et al. 2017) havíamos identificado um *Helitron*, denominado Hel_c35, que se tornou um dos segmentos do vírus simbionte *Cotesia vestalis* bracovirus (CvBV) associado à vespa parasitoide *C. vestalis*. Neste último estudo, também havíamos demonstrado que elementos Hel_c35 se encontravam distribuídos de forma desigual em genomas de diversas espécies de insetos em diferentes ordens, além de uma espécie de aracnídeo. Tal distribuição desigual e irregular já indicava que este *Helitron* estaria envolvido em vários eventos de HT. De fato, nossos resultados sugeriam que o próprio elemento Hel_c35 presente em CvBV teria se originado após a HT de um díptero para *C. vestalis*, seguida pela inserção deste *Helitron* no genoma proviral de CvBV. Além disso, nossas análises apontavam para um segundo evento de HT, de *C. vestalis* para a espécie de mariposa *Bombyx mori*.

Os resultados apresentados no Capítulo 3 descrevem a evolução de elementos Hel_c35 e sua distribuição em genomas de artrópodes utilizando uma amostra consideravelmente maior de espécies e análises mais robustas. Além de atualizar nossos achados anteriores (Heringer et al. 2017) ao revelar uma quantidade e diversidade consideravelmente maior de espécies com elementos Hel_c35, nossos resultados sugerem que esta família de *Helitrons* possivelmente está envolvida em dezenas de eventos de HT. Várias destas HTs estão associadas a espécies que contém sequências mais similares ao elemento Hel_c35 encontrado em *C. vestalis*, provavelmente foram transferidas horizontalmente desta vespa parasitoide para outras espécies de insetos e facilitada pela presença do *Helitron* Hel_c35 em partículas virais de CvBV.

7. CONCLUSÕES

Desde que os *Helitrons* foram descritos pela primeira vez em 2001, estes elementos têm se revelado cada vez mais como componentes genômicos importantes e versáteis em diversos grupos de organismos eucariotos. Algumas das características mais bem estabelecidas sobre os *Helitrons* nas últimas duas décadas dizem respeito a sua capacidade de ocupar frações consideráveis dos seus genomas hospedeiros, capturar, mobilizar e duplicar fragmentos cromossômicos. Apesar disso, informações sobre a sua origem e mecanismo de transposição permaneceram obscuras até recentemente. O objetivo central deste trabalho foi o de elucidar a origem e relações evolutivas destes elementos através do estudo de sua estrutura codificante, composta por dois domínios principais. Nossos resultados indicam que *Helitrons* representam transposons descendentes de plasmídeos procarióticos que invadiram o genoma dos seus primeiros hospedeiros eucarióticos em um período próximo à origem deste domínio da vida. Apesar do domínio catalítico central da sua transposase RepHel se assemelhar mais a proteínas encontradas em um grupo de plasmídeos e vírus bacterianos, *Helitrons* diferem destes últimos por codificarem um domínio helicase em sua transposase.

Em conjunto com dados revelados em outros estudos, nossos resultados sugerem que este domínio helicase não representa uma aquisição evolutiva posterior à invasão dos *Helitrons* em genomas eucarióticos. Ao contrário, a estrutura composta por dois domínios principais na transposase RepHel parece anteceder a origem dos *Helitrons* em eucariotos e ser indispensável para a transposição destes elementos. De fato, a similaridade estrutural entre a transposase RepHel e relaxases encontradas em plasmídeos indica que o domínio Hel desempenha uma função complexa que vai além da simples atividade típica de uma helicase. Neste cenário, os domínios Rep e Hel desempenhariam funções enzimáticas essenciais, complementares e necessariamente concatenadas nas principais etapas do processo de transposição dos *Helitrons*.

Para além dos aspectos fundamentais sobre a origem e mecanismo de transposição destes elementos, nosso estudo de uma família de *Helitrons* encontrada em artrópodes ilustra como a evolução destes transposons em genomas hospedeiros pode ser altamente complexa. Tal complexidade se dá pela capacidade dos *Helitrons* de invadir novas espécies por transferência horizontal, sendo que análises filogenéticas de suas sequências comumente resultam em topologias incongruentes com as relações evolutivas de suas espécies hospedeiras. A evolução desta família de *Helitrons* analisada no nosso último capítulo é particularmente notável não só por incluir múltiplos eventos de transferência horizontal entre diferentes ordens de artrópodes, mas também pela associação entre um elemento desta família com o vírus simbiótico de uma vespa parasitóide.

Os aspectos revelados sobre os *Helitrons* neste trabalho, e em outros estudos recentes, sobre a sua origem, evolução, mecanismo de transposição e estrutura da sua transposase, abrem caminho para futuras investigações mais profundas sobre cada um destes temas. No campo das análises in silico, o aumento no número de espécies com genomas sequenciados poderá contribuir com cenários mais completos sobre a origem dos *Helitrons*, seja revelando variantes estruturalmente mais semelhantes à sua forma ancestral ou replicons evolutivamente mais próximos dos *Helitrons*. Já análises in vitro poderão confirmar se as similaridades estruturais entre transposases RepHel e relaxases de fato se traduzem em semelhanças funcionais. Por fim, a compreensão mais detalhada da estrutura e processos enzimáticos conduzidos por esta transposase única em genomas eucariotos cria novas possibilidades na investigação de ferramentas de engenharia genética.

8. REFERÊNCIAS

Bellas, C. M., & Sommaruga, R. (2021). Polinton-like viruses are abundant in aquatic ecosystems. Microbiome, 9:13.

Bochman, M. L., Sabouri, N., & Zakian, V. A. (2010). Unwinding the functions of the Pif1 family helicases. DNA repair, 9(3), 237-249.

Boule, J. B., & Zakian, V. A. (2006). Roles of Pif1-like helicases in the maintenance of genomic stability. Nucleic acids research, 34(15), 4147-4153.

Bourque, G., Burns, K. H., Gehring, M., Gorbunova, V., Seluanov, A., Hammell, M., ... & Feschotte, C. (2018). Ten things you should know about transposable elements. Genome biology, 19:199.

Chandler, M., De La Cruz, F., Dyda, F., Hickman, A. B., Moncalian, G., & Ton-Hoang, B. (2013). Breaking and joining single-stranded DNA: the HUH endonuclease superfamily. Nature Reviews Microbiology, 11(8), 525-538.

Chellapan, B. V., van Dam, P., Rep, M., Cornelissen, B. J., & Fokkens, L. (2016). Non-canonical Helitrons in Fusarium oxysporum. Mobile DNA, 7:27.

Dias, G. B., Heringer, P., & Kuhn, G. C. (2016). Helitrons in Drosophila: Chromatin modulation and tandem insertions. Mobile genetic elements, 6(2), e1154638.

Du, C., Fefelova, N., Caronna, J., He, L., & Dooner, H. K. (2009). The polychromatic Helitron landscape of the maize genome. Proceedings of the National Academy of Sciences, 106(47), 19916-19921.

Feschotte, C., & Wessler, S. R. (2001). Treasures in the attic: rolling circle transposons discovered in eukaryotic genomes. Proceedings of the National Academy of Sciences, 98(16), 8923-8924.

Grabundzija, I., Messing, S. A., Thomas, J., Cosby, R. L., Bilic, I., Miskey, C., ... & Ivics, Z. (2016). A Helitron transposon reconstructed from bats reveals a novel mechanism of genome shuffling in eukaryotes. Nature communications, 7:10716.

Grabundzija, I., Hickman, A. B., & Dyda, F. (2018). Helraiser intermediates provide insight into the mechanism of eukaryotic replicative transposition. Nature communications, 9:1278.

Heringer, P., Dias, G. B., & Kuhn, G. C. (2017). A horizontally transferred autonomous Helitron became a full polydnavirus segment in Cotesia vestalis. G3: Genes, Genomes, Genetics, 7(12), 3925-3935.

Heringer, P., & Kuhn, G. (2018). Exploring the remote ties between helitron transposases and other rolling-circle replication proteins. International journal of molecular sciences, 19(10), 3079.

Kapitonov, V. V., & Jurka, J. (2001). Rolling-circle transposons in eukaryotes. Proceedings of the National Academy of Sciences, 98(15), 8714-8719.

Kapitonov, V. V., & Jurka, J. (2007). Helitrons on a roll: eukaryotic rolling-circle transposons. TRENDS in Genetics, 23(10), 521-529.

Kazlauskas, D., Varsani, A., Koonin, E. V., & Krupovic, M. (2019). Multiple origins of prokaryotic and eukaryotic single-stranded DNA viruses from bacterial and archaeal plasmids. Nature communications, 10:3425

Koonin, E. V., & Dolja, V. V. (2014). Virus world as an evolutionary network of viruses and capsidless selfish elements. Microbiology and Molecular Biology Reviews, 78(2), 278-303.

Koonin, E. V., & Krupovic, M. (2017). Polintons, virophages and transpovirons: a tangled web linking viruses, transposons and immunity. Current opinion in virology, 25, 7-15.

Kosek, D., Grabundzija, I., Lei, H., Bilic, I., Wang, H., Jin, Y., ... & Dyda, F. (2021). The large bat Helitron DNA transposase forms a compact monomeric assembly that buries and protects its covalently bound 5′-transposon end. Molecular Cell, 81(20), 4271-4286.

Krupovic, M. (2013). Networks of evolutionary interactions underlying the polyphyletic origin of ssDNA viruses. Current opinion in virology, 3(5), 578-586.

Krupovic, M., Bamford, D. H., & Koonin, E. V. (2014). Conservation of major and minor jelly-roll capsid proteins in Polinton (Maverick) transposons suggests that they are bona fide viruses. Biology direct, 9:6.

Krupovic, M., & Koonin, E. V. (2015). Polintons: a hotbed of eukaryotic virus, transposon and plasmid evolution. Nature Reviews Microbiology, 13, 105-115.

Muellner, J., & Schmidt, K. H. (2020). Yeast genome maintenance by the multifunctional PIF1 DNA helicase family. Genes, 11(2), 224.

Pritham, E. J., & Feschotte, C. (2007). Massive amplification of rolling-circle transposons in the lineage of the bat Myotis lucifugus. Proceedings of the National Academy of Sciences, 104(6), 1895-1900.

Schaack, S., Gilbert, C., & Feschotte, C. (2010). Promiscuous DNA: horizontal transfer of transposable elements and why it matters for eukaryotic evolution. Trends in ecology & evolution, 25(9), 537-546.

Thomas, J., & Pritham, E. J. (2015). Helitrons, the eukaryotic rolling-circle transposable elements. Microbiology spectrum, 3(4):MDNA3-0049-2014.

Thomas, J., Vadnagara, K., & Pritham, E. J. (2014). DINE-1, the highest copy number repeats in Drosophila melanogaster are non-autonomous endonuclease-encoding rolling-circle transposable elements (Helentrons). Mobile DNA, 5:18.

Van Etten, J., Bhattacharya, D. (2020). Horizontal gene transfer in eukaryotes: not if, but how much?. Trends in Genetics. 36(12):915–925.

Wallau, G. L., Vieira, C., & Loreto, É. L. S. (2018). Genetic exchange in eukaryotes through horizontal transfer: connected by the mobilome. Mobile DNA, 9:6.

Wawrzyniak, P., Płucienniczak, G., & Bartosik, D. (2017). The different faces of rolling-circle replication and its multifunctional initiator proteins. Frontiers in microbiology, 8, 2353.

Wells, J. N., & Feschotte, C. (2020). A field guide to eukaryotic transposable elements. Annual review of genetics, 54, 539-561.

9. ANEXOS

9.1 Material suplementar do Capítulo 1

Supplementary Material

Supplementary Table S1. Taxa information

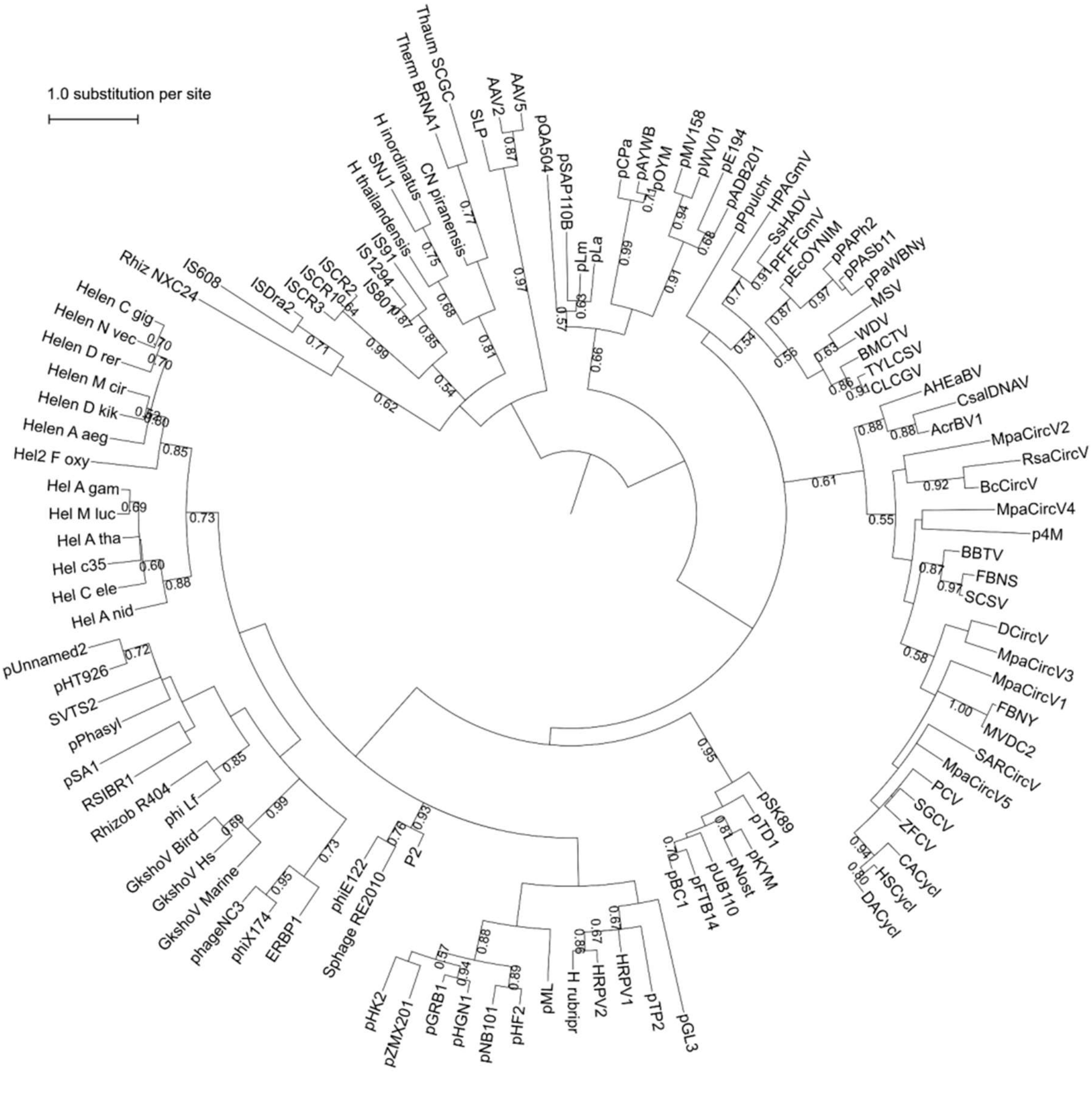
Group	Sequence ID	Taxon name	Family/Group[a]	# of tyr[b]	Accession
Eukaryotic viruses					
	MSV	Maize streak virus	Geminiviridae	1	AAF97764.1
	WDV	Wheat dwarf virus	Geminiviridae	1	CAA57625.1
	BMCTV	Beet mild curly top virus	Geminiviridae	1	AAC54875.1
	TYLCSV	Tomato yellow leaf curl Sardinia virus	Geminiviridae	1	CAA43466.1
	CLCGV	Cotton leaf curl Gezira virus	Geminiviridae	1	AAF97439.1
	SsHADV	Sclerotinia sclerotiorum hypovirulence associated DNA virus 1	Genomoviridae	1	YP_003104796.1
	PFFFGmV	Pacific flying fox faeces associated gemycircularvirus 12	Genomoviridae	1	AMH87729.1
	HPAGmV	Human plasma-associated gemycircularvirus	Genomoviridae	1	YP_009181996.1
	BBTV	Banana bunchy top virus	Nanoviridae	1	NP_604483.1
	FBNS	Faba bean necrotic stunt virus	Nanoviridae	1	YP_003104737.1
	SCSV	Subterranean clover stunt virus	Nanoviridae	1	Q9ICP7.1
	FBNY	Faba bean necrotic yellows C11 alphasatellite	Nanovirus-associated alphasatellite	1	NP_619565.1
	MVDC2	Milk vetch dwarf C2 alphasatellite	Nanovirus-associated alphasatellite	1	NP_619760.1
	PCV	Porcine circovirus 1	Circoviridae	1	NP_065678.1
	SGCV	Silurus glanis circovirus	Circoviridae	1	YP_009091696.1
	ZFCV	Zebra finch circovirus	Circoviridae	1	YP_009134739.1
	HSCycl	Cyclovirus PK5510 (*H. sapiens*)	Circoviridae	1	ADD62457.1
	DACycl	Dragonfly associated cyclovirus 1	Circoviridae	1	YP_009021893.1
	CACycl	Chicken associated cyclovirus 1 (NGchicken8)	Circoviridae	1	ADU77011.1
	DCircV	Diporeia sp. associated circular vírus	Unclassified[c]	1	AGG39813.1
	SARCircV	Circovirus-like genome SAR-A	Unclassified[c]	1	ACQ78172.2
	MpaCircV1	McMurdo Ice Shelf pond-associated circular DNA virus 1	Unclassified[c]	1	AIF71501.1
	MpaCircV2	McMurdo Ice Shelf pond-associated circular DNA virus 2	Unclassified[c]	1	AIF71504.1
	MpaCircV3	McMurdo Ice Shelf pond-associated circular DNA virus 3	Unclassified[c]	1	AIF71507.1
	MpaCircV4	McMurdo Ice Shelf pond-associated circular DNA virus 4	Unclassified[c]	1	AIF71509.1
	MpaCircV5	McMurdo Ice Shelf pond-associated circular DNA virus 5	Unclassified[c]	1	AIF71512.1
	RsaCircV	Rodent stool-associated circular genome virus	Unclassified[c]	1	AEM05803.1
	BcCircV	Bat circovirus ZS/China/2011	Unclassified[c]	1	AEL87784.1
	CsalDNAV	Chaetoceros salsugineum DNA virus	Bacilladnaviridae[d]	1	YP_473359.1
	AcrBV1	Amphibola crenata associated bacilladnavirus 1	Bacilladnaviridae[d]	1	YP_009345107.1
	AHEaBV	Avon-Heathcote estuary associated bacilladnavirus	Bacilladnaviridae[d]	1	YP_009345097.1
	AAV2	Adeno-associated virus 2	Parvoviridae	2	YP_680422.1
	AAV5	Adeno-associated virus 5	Parvoviridae	2	YP_068408.1
	SLP	Slow loris parvovirus 1	Parvoviridae	2	YP_009111339.1
Bacterial viruses					
	phiX174	Enterobacteria phage phiX174	Microviridae	2	NP_040703.1
	phageNC3	Enterobacteria phage NC3	Microviridae	2	AAZ49040.1
	ERBP1	Eel River basin pequenovirus	Microviridae	2	YP_009126954.1
	P2	Escherichia virus P2	Myoviridae	2	NP_046795.1
	Sphage_RE2010	Salmonella phage RE-2010	Myoviridae	2	YP_007003504.1
	phiE122	Burkholderia virus phiE122	Myoviridae	2	YP_001111165.1
	phi_Lf	Xanthomonas phage Lf	Inoviridae	2	AAC54630.1
	SVTS2	Spiroplasma phage SVTS2	Inoviridae	2	AAF18311.2
	Rhizob_R404	Rhizobacter sp. Root404 (Inovirus Gp2 family protein)	Inoviridae	2	WP_056466193.1
	RSIBR1	Ralstonia virus RSIBR1	Inoviridae	2	ATW64834.1
	GkshoV_Hs	Gokushovirus WZ-2015a (H.sapiens)	Microviridae	2	ALS03579.1
	GkshoV_Bird	Gokushovirus WZ-2015a (Bird)	Microviridae	2	ALS03530.1
	GkshoV_Marine	Marine gokushovirus	Microviridae	2	YP_008798246.1

Archaeal viruses					
	HRPV1	Halorubrum pleomorphic virus 1	Pleolipoviridae	2	YP_002791886.1
	HRPV2	Halorubrum pleomorphic virus 2	Pleolipoviridae	2	YP_005454258.1
	H_rubripr	Haloarcula rubripromontorii	Haloarculaceae[e]	2	KOX95265.1
	SNJ1	Natrinema virus SNJ1	Sphaerolipoviridae	1	NC_003158.1[f]
	H_inordinatus	Halopelagius inordinatus	Haloferacaceae[e]	1	WP_092894117.1
	H_thailandensis	Halococcus thailandensis JCM 13552	Halococcaceae[e]	1	EMA56448.1
	CN_piranensis	Candidatus Nitrosopumilus piranensis	Nitrosopumilaceae[e]	1	AJM92193.1
	Therm_BRNA1	Thermoplasmatales archaeon BRNA1	unclassified Thermoplasmatales[e]	1	WP_015491922.1
	Thaum_SCGC	Marine Group I thaumarchaeote SCGC AAA799-P11	unclassified Thaumarchaeota[e]	1	WP_048071526.1
Prokaryotic TEs					
	IS91	Insertion sequence IS91 (*Escherichia coli*)	IS91 Group	2	S23782
	IS801	Insertion sequence IS801 (*Pseudomonas savastanoi*)	IS91 Group	2	P24607.1
	IS1294	Insertion sequence IS1294 (*Escherichia coli*)	IS91 Group	2	CAA07835.1
	ISCR1	Insertion sequence ISCR1 (*Citrobacter freundii*)	ISCR Group	1	AFL38296.1
	ISCR2	Insertion sequence ISCR2 (*Klebsiella pneumoniae*)	ISCR Group	1	SBN37579.1
	ISCR3	Insertion sequence ISCR3 (*Pseudomonas aeruginosa*)	ISCR Group	1	ATE47644.1
	IS608	Insertion sequence IS608 (*Helicobacter pylori*)	IS200/IS605 Family	1	2A6M_A
	Rhiz_NXC24	IS200/IS605 insertion sequence (*Rhizobium* sp. NXC24)	IS200/IS605 Family	1	AVA22184.1
	ISDra2	Insertion sequence ISDra2 (*Deinococcus radiodurans*)	IS200/IS605 Family	1	WP_010887312.1
Plasmids					
Eukaryotic	**pPpulchr**	Pyropia pulchra (red algae) plasmid	Gemini_AL1	1	AAF36424.1
Bacterial	**pEcOYNIM**	Onion yellows phytoplasma EcOYNIM_2000	Gemini_AL1	1	YP_006959597.1
	pPASb11	Candidatus Phytoplasma australiense plasmid pPASb11	Gemini_AL1	1	YP_001965310.1
	pPAPh2	Candidatus Phytoplasma australiense plasmid pPAPh2	Gemini_AL1	1	YP_001965305.1
	pPaWBNy	Paulownia witches'-broom phytoplasma plasmid pPaWBNy-1	Gemini_AL1	1	YP_001708784.1
	p4M	Bifidobacterium pseudocatenulatum plasmid p4M	Viral_Rep	1	NP_613078.1
	pFTB14	Bacillus amyloliquefaciens plasmid pFTB14	Rep_1	1	P13963.1
	pUB110	Staphylococcus aureus plasmid pUB110	Rep_1	1	AAA88362.1
	pBC1	Bacillus coagulans plasmid pBC1	Rep_1	1	AAA98048.1
	pKYM	Shigella sonnei plasmid pKYM	Rep_1	1	AAA98159.1
	pSK89	Staphylococcus aureus plasmid pSK89	Rep_1	1	AAB02112.1
	pNost	Nostoc sp. plasmid ('pNost')	Rep_1	1	AAA25513.1
	pTD1	Treponema denticola plasmid pTD1	Rep_1	1	AAA98363.1
	pAYWB	Aster yellows witches'-broom phytoplasma AYWB plasmid pAYWB-II	Rep_2	1	ABC65794.1
	pOYM	Onion yellows phytoplasma plasmid pOYM	Rep_2	1	YP_002600752.1
	pCPa	Candidatus Phytoplasma australiense plasmid pCPa	Rep_2	1	YP_001966814.1
	pLm	Leuconostoc mesenteroides plasmid replication protein	Rep_2	1	WP_002815993.1
	pLa	Lactobacillus acidophilus plasmid replication protein	Rep_2	1	WP_003549058.1
	pQA504	Lactococcus lactis plasmid pQA504	Rep_2	1	AEU41945.1
	pSAP110B	Staphylococcus epidermidis plasmid SAP110B	Rep_2	1	YP_006939186.1
	pMV158	Streptococcus agalactiae plasmid pMV158	Rep_2	1	YP_001586272.1
	pE194	Staphylococcus aureus plasmid pE194	Rep_2	1	P03858.2
	pADB201	Mycoplasma mycoides pADB201	Rep_2	1	NP_040430.2
	pWV01	Lactococcus lactis plasmid pWV01	Rep_2	1	NP_053450.1
	pPhasyl	Phage-plasmid hybrid Phasyl	Phage_GPA	2	P19071.1
	pHT926	Brevibacillus borstelensis plasmid pHT926	PHA00330	2	BAA07788.1
	pUnnamed2	Fusobacterium nucleatum subsp. polymorphum plasmid "unnamed2"	PHA00330	2	ALQ43495.1
	pGL3	Leptolyngbya boryana plasmid pGL3	Unclassified	2	AAA25610.1
	pSA1	Streptomyces cyaneus plasmid pSA1.1	Unclassified	2	BAA34784.1
Archaeal	**pHGN1**	Halobacterium sp. plasmid pHGN1	DUF1424	2	S06780
	pGRB1	Halobacterium salinarum plasmid pGRB1	DUF1424	2	P17565.1
	pZMX201	Natrinema sp. CX2021 plasmid pZMX201	DUF1424	2	YP_232880.1
	pHF2	Haloferax sp. Q22 plasmid pHF2	DUF1424	2	AKN10606.1

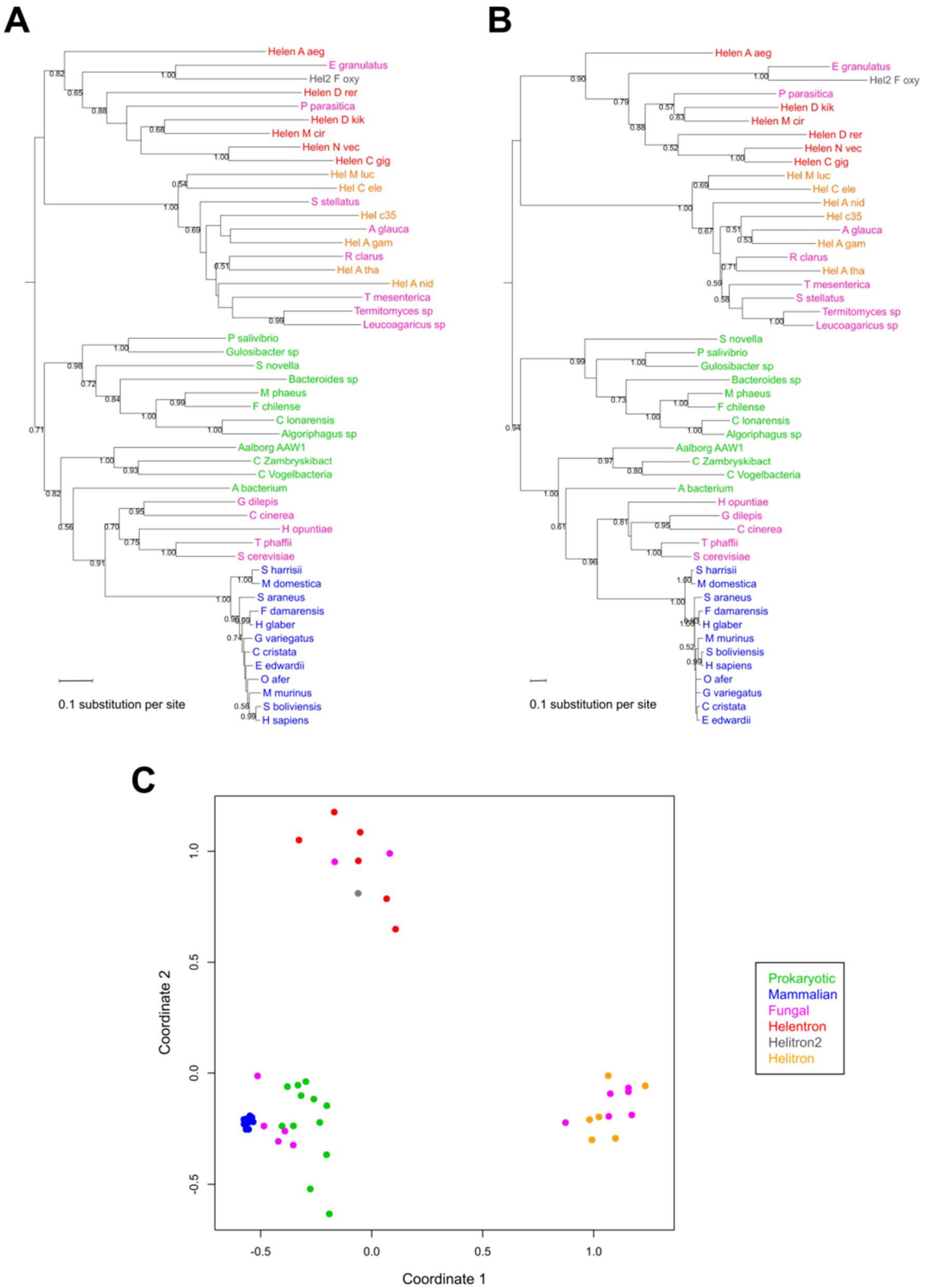
	pHK2	Haloferax lucentense DSM 14919 plasmid pHK2	DUF1424	2	YP_006961960.1
	pNB101	Natronobacterium sp. AS-7091 plasmid pNB101	DUF1424	2	NP_942603.1
	pML	Methanohalophilus mahii plasmid pML	DUF1424	2	NP_976268.1
	pTP2	Thermococcus prieurii plasmid pTP2	PHA00330	2	YP_007974244.1
Helitrons					
	Helen_A_aeg	Helitron-2_Aae (*Aedes aegypti*)	Helentron	2	Helitron-2_Aae [g]
	Helen_D_rer	Helitron-2_DR (*Danio rerio*)	Helentron	2	Helitron-2_DR [g]
	Helen_D_kik	Helitron-1_DK (*Drosophila kikkawai*)	Helentron	2	Helitron-1_DK [g]
	Helen_N_vec	Helitron-1_NV (*Nematostella vectensis*)	Helentron	2	Helitron-1_NV [g]
	Helen_M_cir	Helitron-like sequence (*Mucor circinelloides)*	Helentron	2	EPB86818.1
	Helen_C_gig	Helitron-10_Cgi (*Crassostrea gigas*)	Helentron	2	Helitron-10_CGi [g]
	Hel2_F_oxy	FoHeli1 (*Fusarium oxysporum*)	Helitron2	2	FoHeli1 [g]
	Hel_A_tha	HELITRON1 (*Arabidopsis thaliana*)	Helitron	2	AAD15468.1
	Hel_c35	Hel_c35 (*Cotesia vestalis bracovirus*)	Helitron	2	AEE09607.1
	Hel_M_luc	HELIBAT1 (*Myotis lucifugus*)	Helitron	2	HELIBAT1 [g]
	Hel_A_nid	Helitron-1_AN (*Aspergillus nidulans*)	Helitron	2	XP_662882.1
	Hel_C_ele	HELITRON1_CE (*Caenorhabditis elegans*)	Helitron	2	NP_493834.1
	Hel_A_gam	HELITRON1_AG (*Anopheles gambiae*)	Helitron	2	HELITRON1_AG [g]

Notes:

[a] Plasmids were classified by their RCRE protein family. Helitrons were assigned to their structural variant according to Thomas and Pritham (2015). [b] Number of tyrosines in the unclalytic core. The colors indicate the tyrosine group (Y1 = green, Y2 = red, Yx = blue), as shown in figures 2 and 3C. [c] Sequences representing unclassified viruses were sampled from Zawar-Reza et al. (2014). [d] Family proposed by Kazlauskas et al. (2017). [e] Viral sequence integrated in the genome of indicated taxon. [f] Translated ORF was obtained from nucleotide sequence, according to Wang et al. (2016). [g] Sequences retrieved from Repbase (Bao et al. 2015).



Supplementary Figure S1. Phylogenetic analysis of RCRE domain sequences. Same phylogeny as in Figure 2, with branch support numerical values displayed. Only values above 50% are shown.



Supplementary Figure S2. Phylogenetic and NMDS analysis of helicase sequences. (A) Phylogeny of helicase domain sequences inferred by the Neighbor Joining method (Poisson correction). (B) Phylogeny of helicase domain sequences inferred by the Maximum Likelihood method (LG+G+I). (C) NMDS of evolutionary divergence between helicase domain sequences with scaling representing euclidean distances for three dimensions (stress: 0.08666). See Table S2 for taxa information.

Supplementary Table S2. Taxa used in the helicase domain analysis [a]

Group	Sequence ID	Taxon name	Accession
Prokaryotes			
	M_phaeus	Myroides phaeus	WP_090404604.1
	F_chilense	Flavobacterium chilense	WP_068841780.1
	C_lonarensis	Cecembia lonarensis	WP_009185623.1
	P_salivibrio	Pontimonas salivibrio	WP_104912779.1
	C_Zambryskibact	Candidatus Zambryskibacteria	OHB14600.1
	Algoriphagus_sp	Algoriphagus sp.	WP_100627322.1
	A_bacterium	Alphaproteobacteria bacterium	OJV13697.1
	C_Vogelbacteria	Candidatus Vogelbacteria	OHA59397.1
	Aalborg_AAW1	SR1 bacterium Aalborg_AAW-1	AKH32407.1
	Gulosibacter_sp	Gulosibacter sp.	WP_087008023.1
	Bacteroides_sp	Bacteroides sp.	CDC65823.1
	S_novella	Starkeya novella	PZQ84937.1
Fungi			
	P_parasitica	Parasitella parasitica	CEP10706.1
	G_dilepis	Gymnopilus dilepis	PPQ64766.1
	C_cinerea	Coprinopsis cinerea	XP_001829007.2
	H_opuntiae	Hanseniaspora opuntiae	OEJ83279.1
	T_phaffii	Tetrapisispora phaffii	XP_003684282.1
	E_granulatus	Elaphomyces granulatus	OXV06635.1
	R_clarus	Rhizophagus clarus	GBB91117.1
	T_mesenterica	Tremella mesenterica	XP_007002293.1
	A_glauca	Absidia glauca	SAL95951.1
	Termitomyces_sp	Termitomyces sp.	KNZ79783.1
	Leucoagaricus_sp	Leucoagaricus sp.	KXN86260.1
	S_stellatus	Sphaerobolus stellatus	KIJ35046.1
	S_cerevisiae	Saccharomyces cerevisiae	NP_013650.1
Mammals			
	F_damarensis	Fukomys damarensis	XP_010639595.1
	H_glaber	Heterocephalus glaber	EHA98492.1
	S_boliviensis	Saimiri boliviensis	XP_010349962.1
	S_araneus	Sorex araneus	XP_004619712.1
	M_murinus	Microcebus murinus	XP_012614176.1
	H_sapiens	Homo sapiens	NP_079325.2
	S_harrisii	Sarcophilus harrisii	XP_012398677.2
	M_domestica	Monodelphis domestica	XP_007479627.1
	G_variegatus	Galeopterus variegatus	XP_008566201.1
	C_cristata	Condylura cristata	XP_004687737.1
	E_edwardii	Elephantulus edwardii	XP_006899697.1
	O_afer	Orycteropus afer	XP_007956003.1
Helentron			
	Helen_A_aeg	Helitron-2_Aae (Aedes aegypti)	Helitron-2_Aae [b]
	Helen_D_rer	Helitron-2_DR (Danio rerio)	Helitron-2_DR [b]
	Helen_D_kik	Helitron-1_DK (Drosophila kikkawai)	Helitron-1_DK [b]
	Helen_N_vec	Helitron-1_NV (Nematostella vectensis)	Helitron-1_NV [b]
	Helen_M_cir	Helitron-like sequence (Mucor circinelloides)	EPB86818.1
	Helen_C_gig	Helitron-10_Cgi (Crassostrea gigas)	Helitron-10_CGi [b]
Helitron2			
	Hel2_F_oxy	FoHeli1 (Fusarium oxysporum)	FoHeli1 [b]
Helitron			
	Hel_A_tha	HELITRON1 (Arabidopsis thaliana)	AAD15468.1
	Hel_c35	Hel_c35 (Cotesia vestalis bracovirus)	AEE09607.1
	Hel_M_luc	HELIBAT1 (Myotis lucifugus)	HELIBAT1 [b]
	Hel_A_nid	Helitron-1_AN (Aspergillus nidulans)	XP_662882.1
	Hel_C_ele	HELITRON1_CE (Caenorhabditis elegans)	NP_493834.1
	Hel_A_gam	HELITRON1_AG (Anopheles gambiae)	HELITRON1_AG [b]

Notes: [a] Prokaryotic, fungal and mammalian sequences were retrieved from Genbank (Benson et al. 2017) by using Helitron sequences as a reference. [b] Sequences retrieved from Repbase (Bao et al. 2015).

References

Bao, W.; Kojima, K.K; Kohany, O. Repbase Update, a database of repetitive elements in eukaryotic genomes. *Mob. DNA* 2015, 6, 11, https://doi.org/10.1186/s13100-015-0041-9.

Benson, D.A.; Cavanaugh, M.; Clark, K.; Karsch-Mizrachi, I.; Lipman, D.J.; Ostell, J.; Sayers, E.W. GenBank. *Nucleic Acids Res.* 2017, 45:D37–D42, https://doi.org/10.1093/nar/gkx1094.

Kazlauskas, D.; Dayaram, A.; Kraberger, S.; Goldstien, S.; Varsani, A.; Krupovic, M. Evolutionary history of ssDNA bacilladnaviruses features horizontal acquisition of the capsid gene from ssRNA nodaviruses. *Virology* 2017, 504, 114-121, https://doi.org/10.1016/j.virol.2017.02.001.

Thomas, J.; Pritham, E.J. Helitrons, the eukaryotic rolling-circle transposable elements. *Microbiol. Spectr.* 2015, 3, MDNA3-0049-2014, https://www.doi.org/10.1128/microbiolspec.MDNA3-0049-2014.

Wang, Y.; Sima, L.; Lv, J.; Huang, S.; Liu, Y.; Wang, J.; Krupovic, M.; Chen, X. Identification, characterization, and application of the replicon region of the halophilic temperate sphaerolipovirus SNJ1. *J. Bacteriol.* 2016, 198:1952–1964, http://dx.doi.org/10.1128/JB.00131-16.

Zawar-Reza, P.; Argüello-Astorga, G.R.; Kraberger, S.; Julian, L.; Stainton, D.; Broady, P.A.; Varsani, A. Diverse small circular single-stranded DNA viruses identified in a freshwater pond on the McMurdo Ice Shelf (Antarctica). *Infect. Genet. Evol.* 2014, 26, 132-138, https://doi.org/10.1016/j.meegid.2014.05.018.

Supplementary Data S1. Trimmed amino acid sequences used in the alignment

```
>MSV
VNTFLTYPHCPENPEIVCQMIWELVGRWTPKYIICAQEAHKDGDMHLHALLQTEKPVRITDSRFFDIEGFHPNIQSAKSVNKVRDYILKEPL
>WDV
KYLFLTYPQCTLEPQYALDSLRTLLNKYEPLYIAAVRELHEDGSPHLHVLVQNKLRASITNPNALNLRMFHPNIQAAKDCNQVRDYITKEVD
>BMCTV
KNIFLTYPRCSVIKEDALEILKNIPCPSDKLFIRVSQEKHQDGSLHLHALIQFKGKAQFRNPRHFDITHFHPNFQGAKSASDVKQYIEKDGD
>TYLCSV
KNYFLTYPKCDLTKENALSQITNLQTPTNKLFIKICRELHENGEPHLHILIQFEGKYNCTNQRFFDLVSFHPNIQGAKSSSDVKSYIDKDGD
>CLCGV
KNYFLTFPKCSLTKEEALEQIQKISTASNKKYIKICRELHEDGQPHLHVLLQFEGKFKCQNQRLFDLVSFHPNIQGAKSSSDVKSYIDKDGD
>SsHADV
KYVLLTYAQCELDAFRVMDKLSLLGAECIIGREHHEDGGTHLHCFAEFGRKFRSRKADVFDVDGHHPNITSRGTPEKGYDYAIKDGD
>PFFFGmV
RYALLTYAQCDLDPFAVVNHLAELAAECIIGREDHADGGIHLHAFVDFGKKYRTRNTRTFDVEGYHPNISSRRTPEEGYDYAIKDGD
>HPAGmV
RFCIVTYSQTDFDADAIVRILHRDCRGCIVARESHLDGGTHYHAFVDYGTPRDWTNSRRWDVLGVHPNIKVSRTPFNAYAYVGKDKN
>pPpulchr
RLFFLTYPCGLTKELILRELRKIVVVVSKERESGDGYDHFHVLLEAKTKKNYKDPRCFDILGVHGKYETVRNRKRSLKYICKEGD
>pEcOYNIM
QNIFLTYSQCDLSKEEIKTFIINLCNEKKLQINYLIIGIENHQDKGKHHHVFFQLNKQFRTRDLTIFNIPKYSPHIEPIKDTTDVRNYVKKDGD
>pPASb11
KDIFLTYSKCPLGKEKIHNHIKQLMESKNQKIAYIISNTENHQDKEIHTHVLFQLNKRCNLTSQRFFDLDGYHPKIENTRDVEKAIEYIKKDGD
>pPAPh2
RDIFLTYSKCPLGKEKIHNHLKQLLASKKKEIKYIISNNENHQDKEIHTHVFIQLKKQIEITNQRFFDIEGYHPKIETARDVEKSVSYIKKDKD
>pPaWBNy
KDIFLTYSKCPLGKDKIHNHIKQLMASKKKEIQYLITNQENHKDKEIHSHVLFQLTKSATFNGERFFDIEGFHPEIEVARDIEKSISYIKKDGD
>BBTV
VCWMFTINNPTTLPVMRDEIKYMVYQVERGQEGTRHVQGYVEMKRRSSLKQMRGFFPGAHLEKRKGSQEEARSYCMKEDT
>FBNY
KRWCFTLNYKTAVERESFISLFSRDELNYFVCGDETAPTTNQKHLQGYVSLKKMIRLGGLKKKFGYRAHWEIAKGDDFQNRDYCTKETL
>MVDC2
KRWCFTLNYKTALERETFISLFSRDELNYFVCGDEIAPTTGQKHLQGYVSMKKLIRLGGLKKKFGSIAHWEIAKGDDFQNRDYCTKETL
>FBNS
ICWCFTLNNPLSPIFLHESMKYLVYQTEQGESGNIHFQGYIEMKKRTSLAGMKRLIPGAHFEKRRGTQGEARAYAMKEES
>SCSV
ICWCFTLNNPLAPLSLHESMKYLVYQTEAGDNGTIHYQGYVEMKKRTSLVQMKKLLPGAHLEKRRGSQGEARAYAMKEDS
>PCV
KRWVFTLNNPSEEEKNKIRELPISLFDYFVCGEEGLEEGTPHLQGFANFAKKQTFNKVKWYFGARCHIEKAKGTDQQNKEYCSKEGH
>SGCV
KRYVFTLNNYTTEEYARIDNVGADGLARYMITGKEVGENGTPHLQGFINLKVKKRFSQIKEMLGSRCHIEKARGTDLENRVYCSKEGS
>ZFCV
KRWVFTLNNPTEQEVESVKSLPPSEYHYAIVGKEKGEQGTPHLQGFLHLKKKVRLNQMKQLIPRAHFEIARGSDEDNEQYCSKEGD
>HSCycl
RRFCFTWNNYTELNYALCQEFIKKYCKYGIVGKELAPTTNTPHLQGFCNLQKPMRFSTIKKRLDNGIHIEKSMGSDTQNQTYCSKSGE
>DACycl
RRFVFTWNNYTPSDFETCITFLDNFCKYGIIGKEKCPTTQTPHIQGFCNLSKPMRFNNIKKHLHNSIHIEKANGSDEQNKIYCSKSGE
```

```
>CACycl

RRFVFTWNNYPIEAYDKCEKYLTKFCKYGIVGEEIAPETGTPHLQGFCNLHKPTRFSTIKKHLDNSIHIEKANGSDIDNQKYCSKSGI

>DCircV

RNWVFTLNNYVDADRVVIGERLANDATYVCYQPEIGASGTPHLQGLVVFANPRTLGGVKRLISDRVHLEPMRGTFAEAHAYCSKDDT

>SARCircV

KAWCFTLNNYTENEHGALVQRFSDFDDKYYFIVGCEIGAQGTPHLQGYIEKKVGRFRPLPCFEVLRDGKNAMHFERAKGNRKQNYNYCSKDGD

>MpaCircV1

KHWQFTLNNPTQDERNVLAELGDQPTTQYLIYGDEVGASGTPHLQGHVSFVQRYRFNQVKNWVSPRAHLELVRLLRRHIEYCKKDGA

>MpaCircV2

RCVCVTIHVDNIFWELQKWNQSLTYGIGQLELGLNGSTHWQMYFENNTAISLTQWKQLLGCKRAHVETRKGTALLAIEYCKKEET

>MpaCircV3

RNFVFTWNNYSDASKTYLSTLACKYVAYAEEVAPTTGTRHLQGFIAFTNAKTIQQARSKLPGCHVETMNGSIAQSEDYCSKAGT

>MpaCircV4

KYWVFTWHGPPKDDEGNRASPALWPEPQFDADMMDALQYQMEIAPSTGKYHYQGAVAFKTRKRSDPLREALAIPGAWTQMMRGSDKDQVYTNKEET

>MpaCircV5

KHWCFTVNNYTDEDIHKLSKASLLLQPLVSSCIYQQEVPGQESATPGTPHLQGFISFKTKQSFKFTKNLVSDRAHVEVAKGTPQQNRIYCSKAKD

>RsaCircV

RYYMLTIPYSLFTIPDPLPEGLVWLKGQPERGENGYEHWQLICCTRKKCRASAVKRLFCPQAHVELTRSAAADEYVWKDDT

>BcCircV

RYWLLTIPYEHFTPYLPPNCAYIKGQLEQGSNTSYLHWQLVVYFSQKKSLNYVKLIFGDGIHCEPSKSKAAEEYVWKEDT

>CsalDNAV

SRCIVTFFPKDNDRRWLKPETYFGPNPDNFQCWCGQFEICPRTGALHAHIYFECVRSRRLRFVRTAALFRKYHHRVHIKKARTVSKKQRQSAINYV
LDDAK

>AcrBV1

GRCIVTLFPPDSEPKWLDPSTYYTDPASVVKIWVGQFEITPETNQIHAHIYIEFHHKKRPKFNLFVKMFTDIGKHVNVKSPKKSNNTQRQGAVNYC
MKDET

>AHEaBV

RSGLLTIHPPSSHPSWLKPETWFPQCDDILEIWCAKFEKGEDTGNLHVHIYFKLKHSNTIRFELLQKWITKHVTGFDFKPQRSATKNSTQCVVNYV
LKPET

>p4M

TDWLLTIRRELPDGSERTVDDVVNALQGIFDAAIGQPEKGEGGYRHYQIFAQGKRQRFSTLKKKLTAAGLGDAHVEPRKGSVSEAVGYCSKEKT

>pAYWB

CELVINANKITKSKIENILELKKKAIQNYAYILHDKDTYQNEKEAQLNGKKIGDLKSPHYHIYLRFNYAYDTKHIAQWFNTQDNFVSKIKGRFSDA
LMYMTHANS

>pOYM

CELVINKTLITKTKIETILETKKKAIQNYAYILHDKDIYQNEKEAQLNGKKVGDIKAPHWHIYLRFNYSQDTKHISQWFNTQENFVSKIKGRFSDA
LMYMIHANR

>pCPa

CELVIKADLIKQTEIEKVLESKKKVIQSYAFILHDNDKYLNEKEAKENGKSVGDYKIPHWHIMLRFHQSQEFKYIAKWFNTTENFVSQIKGRFTDA
LLYLTHANR

>pLm

RTFMYTQQLQHLPFQDVAAFQSRLENINVAEYAFIIHDQDTVDGHPVTSHIHAVLRYQNARSVDSVAKQVSDKAQYIEIWNGNYANAYAYLVHKTD

>pLa

RQFMYTQDLDHLPFKKEDLKTLLEKSSAEEWAYILHDKDIGKNGKTIRPHFHVVMKFKDAKTISRVAKLFNDKQEYIEVWRNTIGNAYSYLIHETS

>pQA504

SVFGFTQQFKADMWDWADDEKAVCFPNGVPDTARIMKRVAERLYVYLIGDIKKANAPDRPHAKDLFKYSAIIHDKDMSFAWDTKTNSKVIVPKELH
MHAVIELPSKRDLSFISTAIGIRPEQIEVPRGRYGRENMLAYLVHAKD

>pSAP110B

TKFMYTQQLKYLNLSIEQLKNNLENDAYIQDFAMINHNKDLDENNQNVAEHLHVFIKLNQQKTIDYVADLVDDKAQYIEFFDKSNKSRNEQNGYLY
LLHKTK
```

```
>pMV158

TFLLYPESIPSDWELKLETLGVPMAISPLHDKDKSSIKGQKYKKAHYHVLYIAKNPVTADSVRKKIKLLLGEKSLAMVQVVLNVENMYLYLTHESK

>pE194

TFVLYPESAKAEWLEYLKELHIQFVVSPLHDRDTDTEGRMKKEHYHILVMYEGNKSYEQIKIITEELNATIPQIAGSVKGLVRYMLHMDD

>pADB201

TLLVYPDSAPENWKEILDQNGVEYFGALHDKDVNPDGTIKKPHYHIVLAYSGPTTFNNVKTLCNTLNSPKPLPLDGVGGMWRYMTHKDN

>pWV01

GFLLYPDSIPNDWKEKLESLGVSMAVSPLHDMDEKKDKDTWNSSDVIRNGKHYKKPHYHVIYIARNPVTIESVRNKIKRKLGNSSVAHVEILDYIK
GSYEYLTHESK

>pFTB14

GWIFLTLTVRNVKGERLKPQISEMMEGFRKLFQYKKVKTSVLGFFRALEITKNHEEDTYHPHFHVLLPVKRNYFGKNYIKQAEWTSLWKRAMKLDY
TPIVDIRRVKGRVKIDAEQIESDVREAMMEQKAVLEISKYPVKDTD

>pUB110

RWLFLTLTVKNVYDGEELNKSLSDMAQGFRRMMQYKKINKNLVGFMRATEVTINNKDNSYNQHMHVLVCVEPTYFKNTENYVNQKQWIQFWKKAMK
LDYDPNVKVQMIRPKNKYKSDIQSAIDETAKYPVKDTD

>pBC1

QWLFLTLTVRNTSPESLPETISAMFEGFNRLTKYKAFKTSVKGYFRALEVTKNRDPHSEWFGTYHPHFHVLLCVPSSYFKKKELYITEQEWTDLWK
KAMKLDYTPIVHVQRVKPKEQLEDMETYEEQLKNAIREQNAILEVSKYPVKDTD

>pKYM

RWLFLTLTVRNCEIGELGTVLTAMNAAFKRMEKRKELSPVQGWIRATEVTRGKDGSAHPHFHCLLMVQPSWFKGKNYVKHERWVELWRDCLRVNYE
PNIDIRAVKTKTGEVVANVAEQLQSAVAETLKYSVKPED

>pSK89

QFIFLTLTTPNVTDEHLESEIKNYNHAFQKMFKRKKVNAITKGYVRKLEITYNSKRDDYNPHFHVLMAVNKSYFKDTKAYISQKEWLNLWRDVTGI
SEITQVHVQKIKQNSNKELYEMAKYSGKDSD

>pNost

RWLFVTLTVKNCAITDLRETLTWMNKSFKRFSELKAFPAEGYIKTVEVTRGKTPDGSAHPHFHVLMMVKPSYFGVGYLSQAKWVEMWRKSLRVDYK
PILDVQSLNPQDSLIGLLAEVIKYSVKESD

>pTD1

DFIFITLTVKNCSADELPATLEMMTKGWRRLAMTAMCEFRRSFEGTFKALEITVNKKTGEYHPHYHILAAVKKGYFRKSNPDYISQENLIKLWQKV
CKLDYEPNVDIRRVKNSTYKAVAEVAKYSVKATD

>AAV2

YEIVIKVPSDLDEHLPGISDSFVNWVAEKEWELPPDSDMDLNLIEQAPLTVAEKLQRDFLTEWRRVSKAPEALFFVQFEKGESYFHMHVLVETTGV
KSMVLGRFLSQIREKLIQRIYRGIEPTLPNWFAVTKTRNGAGGGNKVVDECYIPNYLL

>AAV5

YEVIVRVPFDVEEHLPGISDSFVDWVTGQIWELPPESDLNLTLVEQPQLTVADRIRRVFLYEWNKFSKQESKFFVQFEKGSEYFHLHTLVETSGIS
SMVLGRYVSQIRAQLVKVVFQGIEPQINDWVAITKVKKGGANKVVDSGYIPAYLL

>SLP

WELVIKLKYDWIEDLEGSDDPWYDWPEDEIDIYMAILGIKAIKAITRVLRERSKNKTCNYFGQIEQGGEFFHIHLLFEVDGFVSFLLGRMFETLRQ
TLRNSVYFGYPFEVSSEIAITKVKTGGRNKVQDGSYIVNYLL

>SNJ1

HHSVISPPEELYIDAEFPEQELISVAQEFMEEIGMQGIALYHSWSGGDDHDDDIGEWKKRLFADRDWHGDVREELQHRPHVHLIGACPWFPMGDVT
KLTHAETDWVIHRITGKRDGNSSVSLADMRSVARAVVYALSHCA

>H_inordinatus

HHVVFSPPRDWFLQAQDPLDKTFKLIGDILTNHFDAAGRVYYHGWSGGDDLEDDLGEWKNRLFEGRDWETDVRHELEPRPHFHAVVASPFIPGEGV
TDRIHDETGWVIKRIADEKSKRSIDGIDALARVVTYCMSHTS

>H_thailandensis

IHAMFSPEQDWTISRVDGMRSESYELAQEAGVTGGGALLHMWRTTDDLDGEFKKWKYRETYGQGWRQATEVAPHVHQIATAPEFEPEQGDWVAKRV
RTLDAMRSLSHPSSYEDVAGLAMYLLSHTA

>CN_piranensis

LHNIVSIPFELYLTKDGRKKLRAKAIKYLKEFDIDGGVMIDHPYRFSKDLESARLSPHLHLIVTGWLDGQKVKELYEKTGWIVTNVSTIETWNDCY
NLSKYLLSHSA

>Therm_BRNA1
```

```
VHVVVSPPQDLRFMRSKEGFRIMVNKVIRVLKDFQVDTGALVFHPWRQCGDRDGSFPSSSFVWRAGPHFHAVGYGYVPEDRIKEFHERTGWILKVV
HDKSDVVSPTATLAYLLTHAG

>Thaum_SCGC

IHLILAVPENQRELPVKLLRQRMSHILKLGNIKGGSVIFHPFRFSKTQHRWYASPHFHLVGFGKSSDIKNAFGRYGWYVKEAGERESVFQTFCYLL
SHCG

>IS91

QHIVFTLPCQYWSLVFHNRWLLAEMSRIAADVILEICHQTDVEPGIFTVIHTWGRDQQWHPHIHLSTTAGGVTSGHTWKNLHFYARKVMSMWRYRI
TRLLSRKYPELVIPDELAVGNSKRDWNCFLDTYRRGWNVNISRVMDNATHVAVYFGSYLK

>IS801

QHLVFTLPDTLWPLFFYNRWLLDALFRLAADNLIYAAKRRGLRVGIFGALHTYGRRLNWHPHVHLSVTAGGLDEQGVWKNLSFHKEALRRRWMWLV
RDYLLGQPLSQLTMPPPLAHILCESDWRRLILAGGQHWHIHLSKKTKNGRKTVNYLGRYLK

>IS1294

VHLVFTLPDTLWPVFESNRWLLNDVCRLAVENLLYAARKRGLEPGIFCAIHTYGRRLNWHPHVHVSVTCGGLNKHGQWKKLSFLKDAMRSRWMWNM
RQLLLKAWSEGMAMPESLSHITTESQWRSLVLKGGKYWHVYMSKKTAGGRNTARYLGRYLK

>ISCR1

RQWVLSFPFQLRFLLARHPQLLSIVYRTLSTHLIKKAGYTKASAQTGSVTLIQRFGSALNLNVHYHMLFLDGVYAEDDYGKQRFHRKALAHTLSHR
IARCMEKRTLTQLHGASVTYRIAVGPQQGRKVFTAGFSLHAGVMAEAHQRDKLERLCRYIS

>ISCR2

RQWVLSFPFQLRFLFASRPEILGIVYRVIATHLVKKAGHTHQVAKTGAVTLIQRFGSALNLNVHFHMLFLDGVYVEQSHGSARFRWKALTHTIAHR
VGRYLERQPMTPLLGHSITYRIAVGSQAGRKVFTAGFSLHAGVAARADERKKLERLCRYIS

>ISCR3

RQWVLSFPYPLRFLFASKPEALGIVQRVIAGWLADQAGIDRASAQCGAVTLIQRFGSALNLNIHFHMLWLDGVYVEATRRELRLHRRALAATIAHR
VCRHLTRKSMDGLRMSSITYRIATGRDAGCKVVTGGFSLHAGVAAEAHESHKLEKLCRYIT

>IS608

HNVVYSCKYHIVWCPKYRRKVLVGAVEMRLKEIIQEVAKELRVEIIEMQTDKDHIHILADIDPSFGVMKFIKRILRQEFNHLKTKLPTLWTNSCFI
STVGGAPLNVVKQYIENQQN

>Rhiz_NXC24

RIVVPDIPHHVTQRGNGRAQTFFCDDDYALYRDLLAHHCRAADVEVWGWVLMPNHVHLILVPADADGIRRALRVHRAYAGHIHARLRRTGHFWQGR
FGCVPMDEEHLAAALRYVALNPV

>ISDra2

RGYVYQLEYHLIWCVKYRHQVLVGEVADGLKDILRDIAAQNGLEVITMEVMPDHVHLLLSATPQQAIPDFVKRRMFVAYPQLKEKLWGGNLWNPSY
CILTVSENTRAQIQKYIESQHD

>phiX174

FIVFDTLTLADDRLEAFYDNPNALRDYFRDIGRMVLAAEGRKANDSHADCYQYFCVPEYGTANGRLHFHAVHFMRTLPTGSVDPNFGRRVRNRRQL
NSLQNTWPYGYSMPIAVRYTQDAFSRSGWLWPVDAKGEPLKATSYMAVGFYVAKYVN

>phageNC3

FFVFDTLTLADDRLQAFNENPNALRDYFRTVGRAVLRAEGRSVKDSYNDCYRYLCVPEFGGQHGRLHWHVVHMVRTLPLGSHDPNFGRKVRNYRQI
NSFRGMWPYGFTQPIAVRYQHDAYSRKGWLWPVDKSGKAMQSKPYQAVAWYVTKYVA

>ERBP1

YCIFNTLTVNESSIEKVFEKGSRIFSDYVRSLDRGVGIAIHKNWRQAVTKRKEGNEFHTYFAVVERGTKNGRLHIHVIHMMKELPNGCVDPNAGRA
IPNRREVTYLKRYWKYGYSAPIAVRFNTNDAFGKKYWRFPVKEVAKNRFESLECKDAGSIIGYIGKYMT

>P2

VGMFITLTAPSKYHPTRQVGKGESKTVQLNHGWNDEAFNPKDAQRYLCHIWSLMRTAFKDNDLQVYGLRVVEPHHDGTPHWHMMLFCNPRQRNQII
EIMRRYALKEDGDERGAARNRFQAKHLNQGGAAGYIAKYIS

>Sphage_RE2010

CAVFYTITCPSRFHSTLNNGRPNPTWTNATVRQSSDYLVGMFAAFRKAMHKAGLRWYGVRVAEPHHDGTVHWHLLCFMRKKDRRAITALLRKFAIR
EDREELGNNTGPRFKSELINPRKGTPTSYIAKYIS

>phiE122

RGVMFTLTCPSRFHAVTTTDSWVRPNPRYDDVDPRAAQAYLRKVWQRTRAELKREGIVYFGMRVAEPNHDGTPHWHGLVFADKIERFCSVMRKHGL
RDSGDEPGAQRHRVRFEMIDRAKGSAVGYVAKYIS

>phi_Lf

AWYFLTLTYRDGSDSSPRDVSELFKRMRGHFNRLKSGRARWNRESFRYVWVGELTQRFRPHYHVMLWVPQGMFFGKVDQRGWWPHGSSQIEKARNC
VGYLAKYAS

>SVTS2
```

```
NLSFLTLTYAVNEKDVKKCKNDLKLFFNNINRWWNNPIRSKNHKGILKYMYTYEYQKRGAVHFHIILNQKIPNSVVQQYWKHGINKNIKVRAGSNE
DVVKYLAKYIV

>Rhizob_R404

RPAMLTLTYREVGQWNPKHISDLLQRIRVWVRRRGHGLRYVWVAELQQRGALHYHLLLWLPRGLTLPKPDKQGWWTHGSTRIEWARKPAGYLAKYA
S

>RSIBR1

VTHMITLTTRECITDLDWFLGLWDAFRRAMARYSQFHYIAVPELQKRGAWHMHVAVSGRVALNLARRVWLKVVGGRGKGYCHIRNPQGAHFGKQWK
LDALASYVAKYIG

>GkshoV_Hs

SNYFVTLTYRPDALPYTKDGKPTLRPKDLTNFFKRLRKHKKGNEKIRYFACGEYGEKKGRPHYHVALFNLKLDDLKPLGPSQGYMLYKSKTLQNIW
GLGFVVIGELTYKSASYISRYVM

>GkshoV_Bird

ENYFVTLTYDNDNVPLSQMHMNTLKKRDFQLFMKRLRKRGNDGIRFFACGEYGSTTMRPHYHAILFNLHLDDLEKLYEKDGMVYYTSQFLQSVWKK
GFVIITSMTWETCAYVARYVC

>GkshoV_Marine

SSSFITLTYDNKHLPPNNSLDYTHWQKFIRSLKKRNNGKSIRYFGVGEYGENFGRPHFHAILFGHTFNDLIPMHSNISKSQQLLSAWPRGFVSVGD
VTPESISYVCGYVQ

>HRPV1

SGVMVTLTTDPKRYDSMLDGLMDAWQNLHETLNYLEGTRLDFIRALEFGGSGLPHLHVCVFGVPYIDHRWLKHYWSHAEIVHIHGMNKRGNDSWIM
TSGTHAGKSVAGYLGKYLS

>HRPV2

NAVFCTLTTDPKKFDSLYDAVMSINENFHRLMSYLRSVTGRPRETLDYIKVLEFTSAGYPHLHVLFFDVPWLVDKRELSAKWKQGQIVDLYPLVHR
DDDDWVEEQTRSDDVYQSKTAGSYVGKYIS

>H_rubripr

NAVLVTLTTDPKRQDSLLDGIDSINENLNRLLSYFDSVTGRPRDRPDYIKALEFTEKGYPHLHVLFFDVPWLCDKSEVAAKWAQGEIVDVYPLTYR
DDEDWVRERTRDDGHEKESTAGAYLGKYLS

>pPhasyl

NVGFLTLTFRDHVTDPKEAQRRFNSLKTNILAKRYRAYIRVMEPMKSGRIHYHLLVALHSDIRTGFDFPAVYRQDYSSANKAIRSEWSFWRKTAPK
YGFGRTELMPVRSNSEGIGRYVGKYIS

>pGL3

RLSFITLTLPPAVAEDLSGRWAHVVDLMKRRLPTEIIACTEVQEKVALHLHIVMVGRHSRGSPRQLEKMWSECCETAVRNVIEPNERVTSRVTNSR
TESESNGNGNATGNTSSNANSNGNANGNIHTEVNWNAAVNVQRIKKSASAYMGKYLS

>pHT926

KPVFMTLTFAENVTDVDLANKAFKQFIRKLNGHVYGRGRVGLKYVTVIEFQKRGAVHYHCVFFNLPFIDSGVIASLWGQGFIKVNSMKKRDGTNCD
NVGAYVTKYMQ

>pSA1

PRVFATLTAPELGIPLDPATYDASDLWRYFTIYLRRESRVSFKVAEYQKRGAVHFHAVIRFDGAGDQPARTLHWGTQLDVQPIGAFGHGEEITEQA
VASYVAKYTT

>pUnnamed2

KSTFLTLTFKENIQDIERANREFTLFIKRLKRYLKNQQLKYIATWELQQRGAIHYHLVLFSVPYIDNKKLGELWANGFIKINKIKETVKNEAVGVY
ITKYFV

>pHGN1

HTAMVTLTASTTEEDGGPRPLVDHLRDLLSSWSAVYDALRHTLEDREFEYLAIIEPTPAGYAHIHLGVFVKGPVVAEQFQDVLDAHVKNSEGAGRE
AHRAVVEDDEDEAAVSIRRSARPDREDGIENLGAYLAAYMA

>pGRB1

HTGMVTLTASSTDDEGRLRPPLEHFEDLLESWEAVRRALARVLEGREWEYLAILEPHESGYVHIHLGVFVRGPVVAEQFEPVLDAHLRNCPTAGED
AHQVFDENGDEDAVRVRRSSHPSRSGGVENLGAYLAAYMA

>pZMX201

HTAMLTFTASSRPNGQPIPPVDHLDELLASWDALTTALDRVLGDRRYARLGILEPHNNGYLHIHVAVFIDGKVEQEDFAPVIRSHVNNCEYATEDA
HDPTSEDTISIRHAGDPKRDSDVIGELAIYLAEYLG

>pHF2

TTAMLTLTASHRNEKGGWRCPADHMRDIMDGYDAARKQLHQVLSGRKWEYARVWEPHADGYGHLHIAVFVEDDLRADDFEPVMRSHVENCGPAGSK
AHDPAGDSVSVRDDVENLGSYISEYIG

>pHK2
```

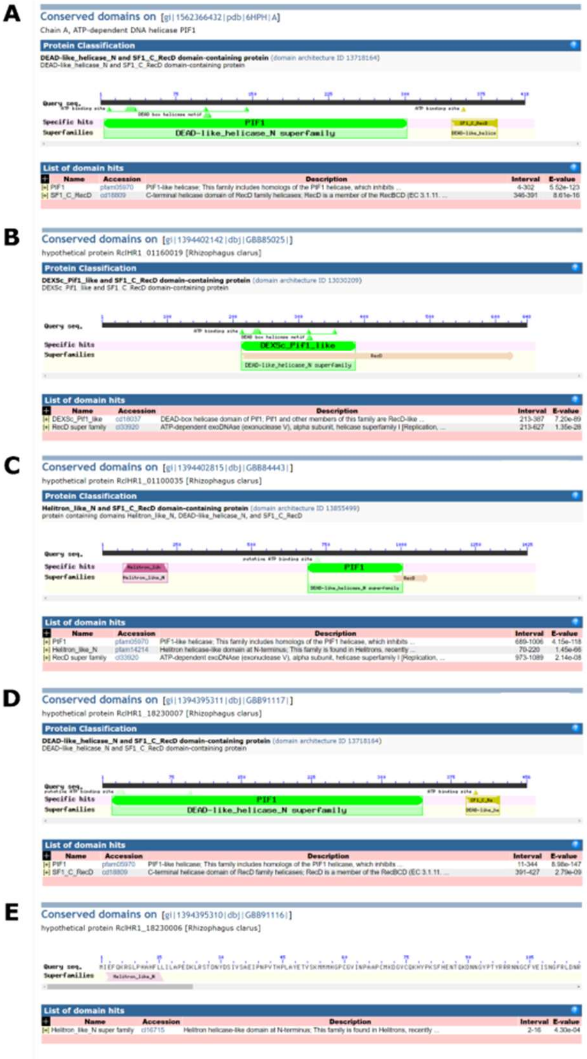
```
ATAMLTFTASSVPNGERLPPVEHTDALHDSYDGVRDTLRNTLDADEWGYWLQAEPHNACYSHLHVGVYFDAAVVGPEFERVIDKHVEECEYASFSA
HDYRNTDYLNDSISLNAGVENMGSYLAAYMG

>pNB101

TMVMVTLSASSENAKGGRRCPADHMRDIARGWNSARKALHRVLRRFEWEYAKVWEPHQSGYGHMHVAVAVDDPIEGETFRPVVRSHVENVEPAGSA
AHGLNAVGMGDTVSVNREVENLGSYISEYIG

>pTP2

DAVFLTLTTDPSRFSNLYEANRQFSHSFNRFMSRLRGYFARRGQHLEYIAVYEFTKSGLLHAHVIIFGVRYVISRWWSQGRVVYIYRLRNVDGRWV
WARRRPRDVRAGEGAEDYLKKYLR

>pML

PITMITLTTYQDSQYSVKKHKVDHEQALEMLVDGFRKLRELITRICEGHTPDYFWILEPHESGYPHMHLCYLEEFTEGEQEHIKSIWGAGEQVDFS
FRKPEDTVRSIRNYLMKYMS

>Helen_A_aeg

PTMFLTLSASETQWPLLLKQLHKLTLVNDDAVTCCLYFNKLVDVLMGILSSPRYVVDFFKRIEFQHRGSPHAHIMLWLANDPNETVSELIRKVCSI
SAIHLSETISHTFTCYKRNEKRCRFNIPYWPMNEERTLYEYYLDVLRSSIQRPTIFLKRSMNEMWTNPFNPWIAEKLRSNMDLQFILDVYSCACYL
AGYVN

>Helen_D_rer

PTFFCTFSAAEMRWPEIVTVIKAQEILRSNPVTVMRMFEKRVDALMAHLLLSPEVEDFFYRVEFQARGSPHIHLLAWVKDAPDPEEDNFIDRYVSC
KLPDPNVDPELHKIVTNHSKSCKKGKVVCRFGFPKLPMPKTMITMDDYLNYAEGLTTGSAVLLKRDPKETWVNGYNPDLLRAWNANMDIQYILDAY
SCIMYMLSYVS

>Helen_D_kik

PTFFITFSAAESKWNELLVTLSRLRLIRSDPVTCSRYFDFRFRQLIKLFKSSETLVHYYWRIEFQHRGSPHSHGMYWFSGAPKLEGPEFIDRFITT
TGDDPELQEVIKHSSSCLREGQEFCRFQMPYPPMPETMVLFEEYKFAIRSSLKKPQVFLRRKFSDRLVNAYNRDILGLHRANMDIQFILDAFACCS
YIINYIN

>Helen_N_vec

ATLFCSFSSAETQWMHLLRILGQLRLIQSDPVTCARHFDYQVNQFLTNFLFSSKISDWFYRVEYQQRGSPHIHMLMWLEDAPQFQIDSFIDKIITC
QKPVDNADLLVLVRHSHTCRKNTSSKCRFNYPQPPMKQTMIIKQNYLLAVSSSINTPTVFLKRNPNELRINNYNPDCLSAWRANMDIQFVLDVYAC
AVYIVNYIS

>Helen_M_cir

PTLFITLSAAESKWTELLAMLKKIWLVQSDPVTCASYFDYRFRELKKTRTAPCNVQEYFFRTEFQHRGSPHIHMLIWLEDAPRILPDSFVDGIITC
EKEWDGSPATWDDIIKHTATCKRKDQIVCRFNIPFLPMDVTRVLVDAYIYSIRSTLKTTKVFLRRTPNQVLTNSYNRKILSMFRSNMDLQFIVDGY
ACCSYVADYIN

>Helen_C_gig

PTWFCSFSAAETKWIPLLKTLGKLRLIKSDPVTCSRYFDYRFQRFLHGVLLHKEVVDYFFRVEFQQRGSPHVHMLLWVKNAPNVSSDSFVDRYVSC
SKSGADPVLVRHAKTCMKKNKPICRFNFPIPPMPKTVTLFETYTLAIRSSLTQSKLFLKRQPYEIRINSYNCTLLKSWLANMDIQFILDPYACATY
IVSYIS

>Hel2_F_oxy

PGAFITFSPADLHWRSLYQHMPQYRLLRQNPHIAAFHFYRRYTLFRDIVLSKKSITDYWDRYEWQGRGSPHNHGLYWMDNCPGADMEDTWGFHVTA
INPEPSRTLRLSQIVEAANVANPERECRFDFPRALRELAAVIGRSYYVFEAARNDSLMNNFNPAIILGWLANIDISPCTSLAVITYAAKYCS

>Hel_A_tha

PDLFITFTCNPKWPHITRYCDKRLNPKDRLDIIARIFKIKLDSLMNDLTVKKKTVASMYTVEFQKRGLPHAHILLFMHAKSKLPTSDDIDKLISAE
IPDKEKEPELYEVINVKSPCMVDGECSKLYPKKHQDITKVGSDGYPIYRRRKIDDYVEKGGIKCDNRYVMPYNKKFSLRYNAHINVEWCNQNDSIK
YLFKYIN

>Hel_c35

PDLFITFTCNPKWIEITQLLLPGQTSSDRHDITARIFRQKIRSLMNFIVKQRDTRCWMYSIEWQKRGLPHAHILIWLVERIQPDQIDDIICAEIPD
YEVDPDLHDVVNPQSPCMVDGKCSKRYPRKLTAETVTGNDGYPLYRRRSPDDKVKRMDFVVDNSWIVPYSPLISKSFKTHCNVEYCNSVKSIKYIC
KYVT

>Hel_M_luc

PDLFITMTCNPKWADITNNLQRWQKVENRPDLVARVFNIKLNALLNDICKFHKVIAKIHVIEFQKRGLPHAHILLILDSESKLRSEDDIDRIVKAE
IPDEDQCPRLFQIVNPNSPCMENGKCSKGYPKEFQNATIGNIDGYPKYKRRSGSTMSIGNKVVDNTWIVPYNPYLCLKYNCHINVEVCASIKSVKY
LFKYIY

>Hel_A_nid

PSLFITFTANPAWDEVTRELRPGETWEDRPDIVSRVFNILRAEMVDELCKKKVAPGRFFTIEYQKRGLPHMHLVLFLEERERFLDAAHIDEMVSAE
LPDPREDLELYKLVNSRAPCCDKNMIYCTKRFPKAEQYETQPIEEGYPLYRRRADPRGAYNDMVRIDNTWVVPYNPYLLKRFRSHINVEVCRGVDV
IKYITKYIY

>Hel_C_ele
```

```
PDIFLTFTCNPAWTEISENLGPRQSASDRPDLIARVFKLKVVDALFDDLLNRDHVAAYISVFEWQKRGLPHVHMLLTMAENSKPRTSEDIDKIVQA
EIPNPDNEPELHRIVNPHSPCMVDGHCSKRYPKDFHPSTTLNVDGYPGYRRRDDGRYVEYGTQHLDNRRVVPYNKWLLLRYNAHMNVEICGFIEAV
KYLFKYVY

>Hel_A_gam

PDLFITVTCNPKWPEITQCLLPRQQAPDRPDVIVRVFRLKLKAILNDLTMGIEVARIHVIEFQKRGLPHAHILVILAEEDKPQTPADYDKIVSAEL
PNPATSSQLFETVNPAAPCMKDGTCEKGFPKSFCEQTRSMDNGYPQYRRRNNGRSVTVKGIELDNRYVVPYNPWFTHKYNCHINVEVCTSISSVKY
LYKYVY
```

9.2 Material suplementar do Capítulo 2

Supplementary Material

Supplementary Figure S1. Conserved domains from sequences containing Pif1 domains. (A) Human Pif1 domain. (B) Best candidate for the genomic Pif1 sequence in the fungal species *Rhizophagus clarus*. (C) Example of a *Helitron* transposase sequence with the Rep and Hel (Pif1) domains. (D) Example of a second candidate genomic Pif1 gene from *R. clarus* structurally similar with the human Pif1 domain. (E) Coding sequence upstream from the ORF in (D) containing a Rep domain, indicating that (D) is part of a broken RepHel ORF starting in (E). Image from the Conserved Domain Database (CDD) search tool (Lu et al. 2020).



Conserved domains on [gi|1562366432|pdb|6HPH|A]

Chain A, ATP-dependent DNA helicase PIF1

Protein Classification

DEAD-like_helicase_N and SF1_C_RecD domain-containing protein (domain architecture ID 13718164)
DEAD-like_helicase_N and SF1_C_RecD domain-containing protein

List of domain hits

Name	Accession	Description	Interval	E-value
[+] PIF1	pfam05970	PIF1-like helicase; This family includes homologs of the PIF1 helicase, which inhibits ...	4-302	5.52e-123
[+] SF1_C_RecD	cd18809	C-terminal helicase domain of RecD family helicases; RecD is a member of the RecBCD (EC 3.1.11. ...	346-391	8.61e-16

Conserved domains on [gi|1394402142|dbj|GBB85025|]

hypothetical protein RclHR1_01160019 [Rhizophagus clarus]

Protein Classification

DEXSc_Pif1_like and SF1_C_RecD domain-containing protein (domain architecture ID 13030209)
DEXSc_Pif1_like and SF1_C_RecD domain-containing protein

List of domain hits

Name	Accession	Description	Interval	E-value
[+] DEXSc_Pif1_like	cd18037	DEAD-box helicase domain of Pif1; Pif1 and other members of this family are RecD-like ...	213-387	7.20e-89
[+] RecD super family	cl33920	ATP-dependent exoDNAse (exonuclease V), alpha subunit, helicase superfamily I [Replication, ...	213-627	1.35e-28

Conserved domains on [gi|1394402815|dbj|GBB84443|]

hypothetical protein RclHR1_01100035 [Rhizophagus clarus]

Protein Classification

Helitron_like_N and SF1_C_RecD domain-containing protein (domain architecture ID 13855499)
protein containing domains Helitron_like_N, DEAD-like_helicase_N, and SF1_C_RecD

List of domain hits

Name	Accession	Description	Interval	E-value
[+] PIF1	pfam05970	PIF1-like helicase; This family includes homologs of the PIF1 helicase, which inhibits ...	689-1006	4.15e-118
[+] Helitron_like_N	pfam14214	Helitron helicase-like domain at N-terminus; This family is found in Helitrons, recently ...	70-220	1.45e-66
[+] RecD super family	cl33920	ATP-dependent exoDNAse (exonuclease V), alpha subunit, helicase superfamily I [Replication, ...	973-1089	2.14e-08

Conserved domains on [gi|1394395311|dbj|GBB91117|]

hypothetical protein RclHR1_18230007 [Rhizophagus clarus]

Protein Classification

DEAD-like_helicase_N and SF1_C_RecD domain-containing protein (domain architecture ID 13718164)
DEAD-like_helicase_N and SF1_C_RecD domain-containing protein

List of domain hits

Name	Accession	Description	Interval	E-value
[+] PIF1	pfam05970	PIF1-like helicase; This family includes homologs of the PIF1 helicase, which inhibits ...	11-344	8.98e-147
[+] SF1_C_RecD	cd18809	C-terminal helicase domain of RecD family helicases; RecD is a member of the RecBCD (EC 3.1.11. ...	391-427	2.79e-09

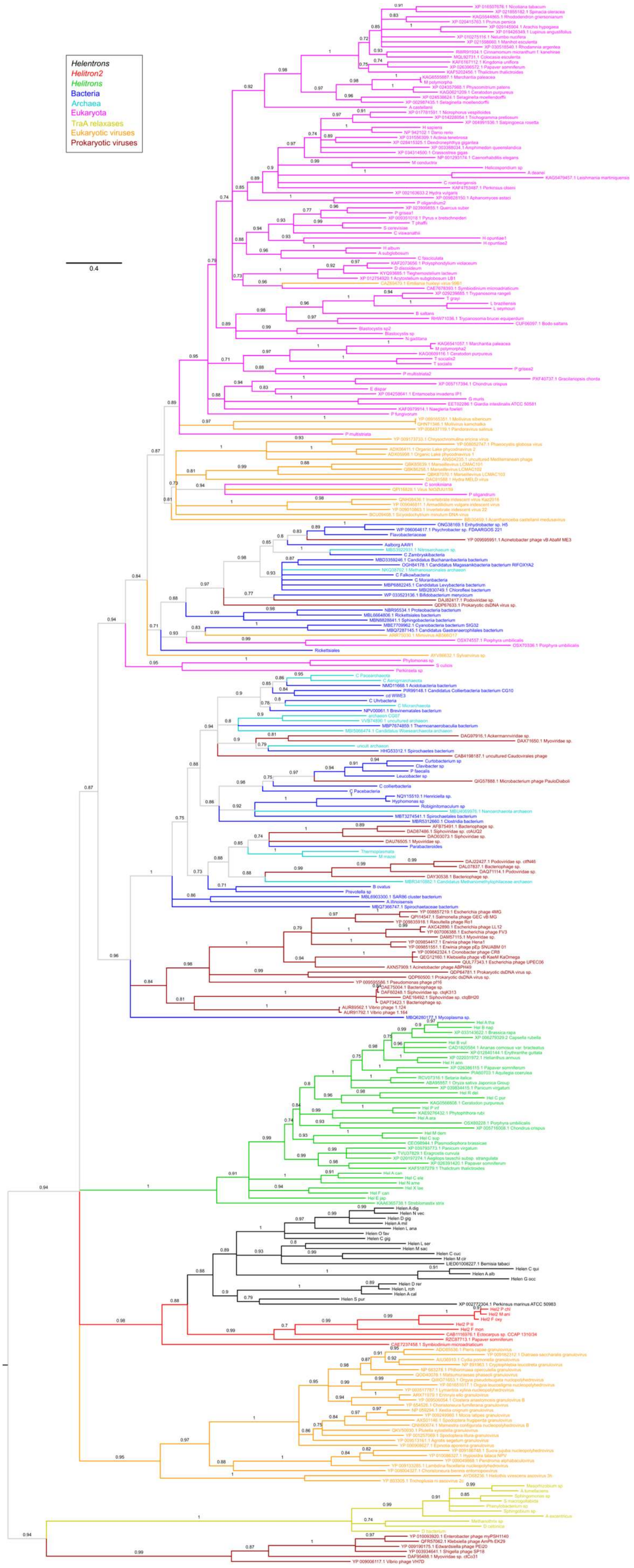
Conserved domains on [gi|1394395310|dbj|GBB91116|]

hypothetical protein RclHR1_18230006 [Rhizophagus clarus]

List of domain hits

Name	Accession	Description	Interval	E-value
[+] Helitron_like_N super family	cl16715	Helitron helicase-like domain at N-terminus; This family is found in Helitrons, recently ...	2-16	4.30e-04

Supplementary Figure S2. Phylogeny of Pif1-like sequences. Same phylogeny as in Figure 3 (main text), displaying branch support values and taxa names (see Table S1 below). Specifications of the procedures used for phylogenetic inference are described in the Materials and Methods.


Helentrons
Helitron2
Helitrons
Bacteria
Archaea
Eukaryota
TraA relaxases
Eukaryotic viruses
Prokaryotic viruses
0.4

Supplementary Table S1

Group	Abbreviation*	Taxon/Host name	Accession
Helentron			
	Helen A alb	*Aedes albopictus*	XP_029715674.1
	Helen D rer	*Danio rerio*	XP_021330385.1
	Helen L ser	*Lucilia sericata*	XP_037823532.1
	Helen N vec	*Nematostella vectensis*	XP_032223796.1
	Helen M cir	*Mucor circinelloides* 1006PhL	EPB86818.1
	Helen C gig	*Crassostrea gigas*	XP_019922950.2
	Helen C qui	*Culex quinquefasciatus*	EDS39572.1
	Helen G occ	*Galendromus occidentalis*	XP_028966621.1
	Helen L roh	*Labeo rohita*	RXN14713.1
	Helen A cal	*Astatotilapia calliptera*	XP_026026780.1
	Helen M sac	*Melanaphis sacchari*	XP_025191627.1
	Helen A mil	*Acropora millepora*	XP_029180665.1
	Helen D gig	*Dendronephthya gigantea*	XP_028394532.1
	Helen L ana	*Lingula anatina*	XP_013378814.1
	Helen C cuc	*Choanephora cucurbitarum*	OBZ82310.1
	Helen S pur	*Strongylocentrotus purpuratus*	XP_011671010.2
	Helen O fav	*Orbicella faveolata*	XP_020609775.1
	Helen A dig	*Acropora digitifera*	XP_015779364.1
		Bemisia tabaci	LIED01008227.1
		Perkinsus marinus ATCC 50983	XP_002772304.1
Helitron2			
	Hel2 F oxy	*Fusarium oxysporum*	AKC01507.1
	Hel2 P lil	*Purpureocillium lilacinum*	OAQ59778.1
	Hel2 P chl	*Pochonia chlamydosporia* 170	XP_018136201.1
	Hel2 F mon	*Fonsecaea monophora*	XP_022510545.1
	Hel2 M ani	*Metarhizium anisopliae*	KFG84029.1
		Ectocarpus sp. CCAP 1310 34	CAB1116976.1
		Papaver somniferum	RZC87713.1
		Symbiodinium microadriaticum	CAE7237458.1
Helitron			
	Hel A tha	*Arabidopsis thaliana*	AAD15468.1
	Hel X lae	*Xenopus laevis*	XP_041421549.1
	Hel C ele	*Caenorhabditis elegans*	NP_493834.1
	Hel A ara	*Anopheles arabiensis*	XP_040164812.1
	Hel B nap	*Brassica napus*	XP_022553550.1
	Hel B vul	*Beta vulgaris subsp. vulgaris*	XP_010692805.1
	Hel C sup	*Chilo suppressalis*	RVE40746.1
	Hel M dem	*Microplitis demolitor*	XP_008549021.2
	Hel F can	*Folsomia candida*	XP_021953640.1
	Hel E jap	*Eumeta japonica*	GBP49736.1
	Hel C pur	*Claviceps purpurea* 20.1	CCE31728.1
	Hel R del	*Rhizopus delemar* RA 99-880	EIE75949.1
	Hel A can	*Ancylostoma caninum*	RCN43056.1
	Hel N ame	*Necator americanus*	XP_013304266.1
	Hel P inf	*Phytophthora infestans* T30-4	XP_002905633.1
	Hel H ann	*Helianthus annuus*	XP_022020320.1
		Brassica rapa	XP_033143622.1
		Capsella rubella	XP_006279329.2
		Ananas comosus var. bracteatus	CAD1820584.1
		Erythranthe guttata	XP_012840144.1
		Helianthus annuus	XP_022031972.1
		Oryza sativa Japonica Group	ABA95557.1
		Setaria italica	RCV07316.1
		Panicum virgatum	XP_039834415.1
		Papaver somniferum	XP_026386115.1
		Aquilegia coerulea	PIA60703.1
		Panicum virgatum	XP_039793773.1
		Eragrostis curvula	TVU37829.1
		Aegilops tauschii subsp. strangulata	XP_020197274.1
		Thalictrum thalictroides	KAF5187279.1
		Papaver somniferum	XP_026391420.1
		Ceratodon purpureus	KAG0566608.1
		Phytophthora rubi	KAE9276432.1
		Plasmodiophora brassicae	CEO98944.1
		Chondrus crispus	XP_005716008.1

		Porphyra umbilicalis	OSX80228.1
		Streblomastix strix	KAA6365738.1
Bacteria			
	C collierbacteria	*Candidatus Collierbacteria* bacterium	KKT34677.1
	Rickettsiales	*Rickettsiales* bacterium	MBO87943.1
	cd WWE3	Candidate division WWE3 bacterium	OGC46700.1
	Prevotella sp	*Prevotella sp.*	EID32542.1
	Curtobacterium sp	*Curtobacterium sp.*	PZF21459.1
	A illinoisensis	*Alkanindiges illinoisensis*	TEU24735.1
	Clavibacter sp	*Clavibacter sp.*	RIJ49579.1
	C Uhrbacteria	*Candidatus Uhrbacteria* bacterium	PIQ67211.1
	C Pacebacteria	*Candidatus Pacebacteria* bacterium	PIR60552.1
	Leucobacter sp	*Leucobacter sp.*	RRD35472.1
	B ovatus	*Bacteroides ovatus*	CDB60500.1
	Hyphomonas sp	*Hyphomonas sp.* Mor2	WP_070961327.1
	C Zambryskibacteria	*Candidatus Zambryskibacteria* bacterium	OHB16576.1
	Flavobacteriaceae	*Flavobacteriaceae bacterium*	QCX39293.1
	Parabacteroides	*Parabacteroides*	WP_075965667.1
	C Falkowbacteria	*Candidatus Falkowbacteria* bacterium	PKM88561.1
	Aalborg AAW1	Candidate division SR1 bacterium Aalborg AAW-1	AKH32407.1
	P faecalis	*Pseudoclavibacter faecalis*	WP_019619849.1
	C Moranbacteria	*Candidatus Moranbacteria* bacterium	PID52462.1
	Robiginitomaculum sp	*Robiginitomaculum sp.*	PHS28547.1
		Candidatus Levybacteria bacterium	MBP6882245.1
		Candidatus Buchananbacteria bacterium	MBD3359246.1
		Candidatus Magasanikbacteria bacterium RIFOXYA2	OGH84178.1
		Chloroflexi bacterium	MBI2830749.1
		Psychrobacter sp. FDAARGOS 221	WP_096064617.1
		Enhydrobacter sp. H5	ONG38169.1
		Bifidobacterium merycicum	WP_033523136.1
		Rickettsiales bacterium	MBL6664806.1
		Proteobacteria bacterium	NBR95534.1
		Sphingobacteriia bacterium	MBN8828841.1
		Candidatus Gastranaerophilales bacterium	MBQ7287145.1
		Cyanobacteria bacterium SIG32	MBE7709962.1
		Mycoplasma sp.	MBQ6280177.1
		Acidobacteria bacterium	NMD11668.1
		Brevinematales bacterium	NPV00061.1
		Spirochaetes bacterium	HHG53312.1
		Candidatus Collierbacteria bacterium CG10	PIR99148.1
		Thermoanaerobaculia bacterium	MBP7674859.1
		Henriciella sp.	NQY15510.1
		Spirochaetales bacterium	MBT3274541.1
		SAR86 cluster bacterium	MBL6903300.1
		Clostridia bacterium	MBR5312660.1
		Spirochaetaceae bacterium	MBQ7366747.1
Archaea			
	uncult archaeon	uncultured archaeon	VVB99669.1
	C Micrarchaeota	*Candidatus Micrarchaeota* archaeon	OIO26558.1
	C Aenigmarchaeota	*Candidatus Aenigmarchaeota* archaeon	OIN88664.1
	archaeon CG07	archaeon CG07	PIU63205.1
	C Pacearchaeota	*Candidatus Pacearchaeota* archaeon	OGJ22063.1
	M mazei	*Methanosarcina mazei*	TAH75514.1
	Thermoplasmata	*Thermoplasmata* archaeon	RLF60972.1
		Nitrosarchaeum sp.	MBS3922931.1
		Methanosarcinales archaeon	NKQ38702.1
		uncultured archaeon	VVB74890.1
		Candidatus Woesearchaeota archaeon	MBI5066474.1
		Candidatus Methanomethylophilaceae archaeon	MBR3410882.1
		Nanoarchaeota archaeon	MBU4069976.1
Eukaryota			
	S cerevisiae	*Saccharomyces cerevisiae*	NP_013650.1
	H opuntiae1	*Hanseniaspora opuntiae*	OEJ88177.1
	H opuntiae2	*Hanseniaspora opuntiae*	OEJ88178.1
	T phaffii	*Tetrapisispora phaffii*	XP_003684282.1
	C viswanathii	*Candida viswanathii*	RCK63232.1
	P grisea1	*Pyricularia grisea*	XP_030977745.1
	P grisea2	*Pyricularia grisea*	XP_030984166.1
	H sapiens	*Homo sapiens*	NP_079325.2

D discoideum	*Dictyostelium discoideum*	XP_642006.1
A subglobosum	*Acytostelium subglobosum*	XP_012757294.1
A castellanii	*Acanthamoeba castellanii*	XP_004352499.1
E dispar	*Entamoeba dispar*	XP_001738818.1
P fungivorum	*Planoprotostelium fungivorum*	PRP79697.1
C fasciculata	*Cavenderia fasciculata*	XP_004367121.1
H album	*Heterostelium album*	XP_020433530.1
T socialis	*Tetrabaena socialis*	PNH12573.1
T socialis2	*Tetrabaena socialis*	PNH01360.1
M conductrix	*Micractinium conductrix*	PSC73053.1
Helicosporidium sp	*Helicosporidium sp.*	KDD76138.1
C sorokiniana	*Chlorella sorokiniana*	PRW33669.1
M polymorpha	*Marchantia polymorpha*	OAE29545.1
M polymorpha2	*Marchantia polymorpha*	OAE19993.1
Blastocystis sp	*Blastocystis sp.*	OAO12860.1
Blastocystis sp2	*Blastocystis sp.*	OAO14610.1
P multistriata	*Pseudo-nitzschia multistriata*	VEU33680.1
P multistriata2	*Pseudo-nitzschia multistriata*	VEU34803.1
C roenbergensis	*Cafeteria roenbergensis*	KAA0155673.1
N gaditana	*Nannochloropsis gaditana*	EWM28750.1
P oligandrum	*Pythium oligandrum*	TMW55246.1
P oligandrum2	*Pythium oligandrum*	TMW67775.1
B saltans	*Bodo saltans*	CUE63209.1
L braziliensis	*Leishmania braziliensis*	XP_001562602.1
T grayi	*Trypanosoma grayi*	XP_009312949.1
L seymouri	*Leptomonas seymouri*	KPI83497.1
Phytomonas sp	*Phytomonas sp.*	CCW70641.1
Perkinsela sp	*Perkinsela sp.*	KNH07790.1
A deanei	*Angomonas deanei*	EPY29325.1
S culicis	*Strigomonas culicis*	EPY23385.1
G muris	*Giardia muris*	TNJ28558.1
	Polysphondylium violaceum	KAF2073656.1
	Tieghemostelium lacteum	KYQ93685.1
	Acytostelium subglobosum LB1	XP_012754920.1
	Trichogramma pretiosum	XP_014228054.1
	Nicrophorus vespilloides	XP_017781591.1
	Danio rerio	NP_942102.1
	Crassostrea gigas	XP_034314500.1
	Dendronephthya gigantea	XP_028415325.1
	Actinia tenebrosa	XP_031556309.1
	Amphimedon queenslandica	XP_003388034.1
	Caenorhabditis elegans	NP_001293174.1
	Salpingoeca rosetta	XP_004991536.1
	Quercus suber	XP_023909855.1
	Pyrus x bretschneideri	XP_009351018.1
	Hydra vulgaris	XP_002163633.2
	Perkinsus olseni	KAF4753487.1
	Rhododendron griersonianum	KAG5544865.1
	Prunus persica	XP_020415763.1
	Nicotiana tabacum	XP_016507676.1
	Manihot esculenta	XP_021598660.1
	Lupinus angustifolius	XP_019426349.1
	Arachis hypogaea	XP_029145904.1
	Rhodamnia argentea	XP_030518540.1
	Papaver somniferum	XP_026396572.1
	Kingdonia uniflora	KAF6167112.1
	Thalictrum thalictroides	KAF5202456.1
	Nelumbo nucifera	XP_010275116.1
	Colocasia esculenta	MQL92731.1
	Cinnamomum micranthum f. kanehirae	RWR91934.1
	Spinacia oleracea	XP_021855182.1
	Marchantia paleacea	KAG6555887.1
	Ceratodon purpureus	KAG0621209.1
	Physcomitrium patens	XP_024357988.1
	Selaginella moellendorffii	XP_002987435.1
	Selaginella moellendorffii	XP_024538624.1
	Entamoeba invadens IP1	XP_004258641.1
	Symbiodinium microadriaticum	CAE7678393.1
	Trypanosoma brucei equiperdum	RHW71036.1

		Trypanosoma rangeli	XP_029239885.1
		Marchantia paleacea	KAG6541057.1
		Ceratodon purpureus	KAG0609116.1
		Aphanomyces astaci	XP_009828150.1
		Naegleria fowleri	KAF0979914.1
		Chondrus crispus	XP_005717394.1
		Gracilariopsis chorda	PXF40737.1
		Giardia intestinalis ATCC 50581	EET02286.1
		Leishmania martiniquensis	KAG5479457.1
		Bodo saltans	CUF06097.1
		Porphyra umbilicalis	OSX74557.1
		Porphyra umbilicalis	OSX70336.1
TraA (plasmids)			
	Methanothrix sp	*Methanothrix sp.*	TFH49976.1
	D bacterium	*Desulfobacteraceae* bacterium	RPI73598.1
	D cetonica	*Desulfosarcina cetonica*	WP_054694573.1
	Sphingobium sp	*Sphingobium sp.* B2	WP_145206887.1
	Sphingomonas sp	*Sphingomonas sp.* AAP5	WP_133192514.1
	Phenylobacterium sp	*Phenylobacterium sp.* CCH9-H3	WP_068876894.1
	S macrogoltabida	*Sphingopyxis macrogoltabida*	WP_054590692.1
	A tumefaciens	*Agrobacterium tumefaciens*	AYM81042.1
	Mesorhizobium sp	*Mesorhizobium sp.* B4-1-1	WP_140901472.1
	A excentricus	*Asticcacaulis excentricus*	WP_013478970.1
Eukaryotic viruses			
		Emiliania huxleyi virus 99B1	CAZ69470.1
		Marseillevirus LCMAC101	QBK85639.1
		Marseillevirus LCMAC102	QBK86258.1
		Marseillevirus LCMAC103	QBK87070.1
		Sicyoidochytrium minutum DNA virus	BCU09408.1
		Organic Lake phycodnavirus 1	ADX05998.1
		Organic Lake phycodnavirus 2	ADX06411.1
		Chrysochromulina ericina virus	YP_009173733.1
		Phaeocystis globosa virus	YP_008052747.1
		uncultured Mediterranean phage	ANS04235.1
		Virus NIOZUU159	QPI16828.1
		Invertebrate iridescent virus 22	YP_009010863.1
		Invertebrate iridescent virus Kaz2018	QNH08436.1
		Armadillidium vulgare iridescent virus	YP_009046811.1
		Hydra MELD virus	DAC81588.1
		Mimivirus AB566O17	ARR75030.1
		Mollivirus kamchatka	QHN71346.1
		Mollivirus sibericum	YP_009165351.1
		Erinnyis ello granulovirus	ARX71979.1
		Clostera anastomosis granulovirus B	YP_009506054.1
		Choristoneura fumiferana granulovirus	YP_654526.1
		Phthorimaea operculella granulovirus	NP_663278.1
		Cydia pomonella granulovirus	AIU36910.1
		Pieris rapae granulovirus	ADO85536.1
		Cryptophlebia leucotreta granulovirus	NP_891963.1
		Matsumuraeses phaseoli granulovirus	QOD40078.1
		Diatraea saccharalis granulovirus	YP_009182312.1
		Lymantria xylina nucleopolyhedrovirus	YP_003517787.1
		Orgyia pseudotsugata nuclopolyhedrovirus	QWO71653.1
		Orgyia leucostigma nucleopolyhedrovirus	YP_001651017.1
		Epinotia aporema granulovirus	YP_006908627.1
		Agrotis segetum granulovirus	YP_009513161.1
		Spodoptera litura granulovirus	YP_001257069.1
		Spodoptera frugiperda granulovirus	AXS01146.1
		Mocis latipes granulovirus	YP_009249960.1
		Xestia cnigrum granulovirus	NP_059294.1
		Mamestra configurata nucleopolyhedrovirus B	QNH90674.1
		Plutella xylostella granulovirus	QKV50030.1
		Sucra jujuba nucleopolyhedrovirus	YP_009186748.1
		Hyposidra talaca NPV	YP_010086327.1
		Lambdina fiscellaria nucleopolyhedrovirus	YP_009133285.1
		Peridroma alphabaculovirus	YP_009049868.1
		Choristoneura biennis entomopoxvirus	YP_008004327.1
		Trichoplusia ni ascovirus 2c	YP_803305.1
		Heliothis virescens ascovirus 3h	AYD68236.1

	Pandoravirus salinus	YP_008437119.1
	Acanthamoeba castellanii medusavirus	BBI30459.1
	Sylvanvirus sp.	AYV86632.1
Prokaryotic viruses		
	Acinetobacter phage vB AbaM ME3	YP_009595951.1
	Podoviridae sp.	DAJ82417.1
	Prokaryotic dsDNA virus sp.	QDP67633.1
	Microbacterium phage PauloDiaboli	QIG57888.1
	Myoviridae sp.	DAU76505.1
	Bacteriophage sp.	AFB75491.1
	Siphoviridae sp. ctAUQ2	DAD87486.1
	Siphoviridae sp.	DAO03073.1
	Podoviridae sp.	DAQ71114.1
	Podoviridae sp. ctfN46	DAJ22427.1
	Bacteriophage sp.	DAL07837.1
	Bacteriophage sp.	DAY30538.1
	uncultured Caudovirales phage	CAB4198187.1
	Myoviridae sp.	DAM57115.1
	Escherichia phage FV3	YP_007006388.1
	Escherichia phage LL12	AXC42890.1
	Erwinia phage pEp SNUABM 01	YP_009851551.1
	Erwinia phage Hena1	YP_009854417.1
	Escherichia phage 4MG	YP_008857219.1
	Salmonella phage GEC vB MG	QPI14547.1
	Raoultella phage Ro1	YP_009835918.1
	Acinetobacter phage ABPH49	AXN57909.1
	Cronobacter phage CR8	YP_009042324.1
	Klebsiella phage vB KaeM KaOmega	QEG12160.1
	Escherichia phage UPEC06	QUL77343.1
	Pseudomonas phage pf16	YP_009595586.1
	Prokaryotic dsDNA virus sp.	QDP60500.1
	Prokaryotic dsDNA virus sp.	QDP64781.1
	Vibrio phage 1.164	AUR91792.1
	Vibrio phage 1.124	AUR89562.1
	Bacteriophage sp.	DAE75004.1
	Siphoviridae sp. ctqK313	DAF60248.1
	Bacteriophage sp.	DAP73423.1
	Siphoviridae sp. ctqBH20	DAE16492.1
	Ackermannviridae sp.	DAG97916.1
	Myoviridae sp.	DAX71650.1
	Klebsiella phage AmPh EK29	QFR57062.1
	Enterobacter phage myPSH1140	YP_010093920.1
	Edwardsiella phage PEi20	YP_009190175.1
	Shigella phage SP18	YP_003934641.1
	Myoviridae sp. ctCo31	DAF95488.1
	Vibrio phage VH7D	YP_009006117.1

*Sequences with abbreviated names were selected in the first round of the analysis. The ones without abbreviation were selected afterwards, without filtering taxa with *Helitrons* in their genomes (see text).

Supplementary Table S2

Species	Accession	Classification	Hit from Blastp
Oryza sativa	ABB47755	RepHel protein	-
Arabidopsis thaliana	CAB91581	RepHel protein	-
	NP_190738	79% cover, 59.50% identity to RepHel	RIA05759.1
	CAB63155	67% cover, 49.65% identity to RepHel	XP_018453621.1

Supplementary Table S3

Accession	Identity	Cover	E-value	Classification	Hit from 2nd Blastp
*XP_007819664.1	41.68%	98%	7.00E-106	No significant identity to RepHel	-
*XP_007816514.1	29.64%	92%	9.00E-49	No significant identity to RepHel	-
*XP_007826535.2	28.97%	92%	3.00E-26	RepHel	-
*XP_007816691.2	26.38%	88%	6.00E-26	RepHel	-
*XP_007825309.2	28.26%	87%	6.00E-20	RepHel	-
*XP_007825293.2	28.42%	37%	3.00E-17	No significant identity to RepHel	-
*XP_007816587.2	26.48%	87%	1.00E-16	RepHel	-
*XP_007817134.1	29.06%	46%	5.00E-16	98% cover, 97.90% identity to RepHel (Best hit)	EXU95784.1
*XP_011411820.1	31.74%	38%	2.00E-12	95% cover, 74.31% identity to RepHel (second best hit)	KJK85320.1
*XP_007816591.2	29.35%	42%	4.00E-11	RepHel (cryptic)	-
*XP_007826337.2	31.29%	31%	7.00E-11	RepHel	-
*XP_007816573.2	30.07%	30%	3.00E-10	RepHel	-
*XP_007816551.2	30.07%	30%	5.00E-10	RepHel	-
*XP_007826745.1	29.79%	33%	2.00E-09	100% cover, 76.55% identity to RepHel (Best hit)	EXU95911.1
*XP_007817473.2	30.54%	35%	1.00E-08	97% cover, 78.37% identity to RepHel (second best hit)	XP_007816587.2
*XP_007817117.2	22.46%	66%	3.00E-08	RepHel	-
*XP_007826647.1	22.75%	47%	3.00E-08	100% cover, 85.49% identity to RepHel (second best hit)	KJK73666.1
*XP_007825291.2	31.62%	40%	2.00E-07	No significant identity to RepHel	-
*XP_007817091.1	26.67%	66%	3.00E-07	RepHel (cryptic)	-
*XP_011411726.1	26.70%	44%	4.00E-05	68% cover, 96.19% identity to RepHel (Best hit)	EXU95304.1
*XP_007826148.1	40.00%	16%	1.00E-04	78% cover, 89.06% identity to RepHel	XP_007816591.2
*XP_007817793.2	30.66%	30%	2.00E-04	46% cover, 62.30% identity to RepHel (second best hit)	EXU94892.1
*XP_007826598.2	43.18%	10%	0.02	Not Pif1 helicase	-
**XP_007826758.1	35.00%	8%	5.00E-05	89% cover, 76.07% identity to RepHel (Best hit)	KID81362.1
**XP_007816555.1	31.96%	10%	4.00E-04	RepHel	-

*Result of Blastp search using the human Pif1 domain (accession 6HPH_A) as a query against *Metarhizium robertsii* ARSEF 23.

**Result of Blastp search using the yeast Pif1 (accession NP_013650.1) as a query against *Metarhizium robertsii* ARSEF 23 (only hits that did not overlap with the ones from the search using the human Pif1 domain).

Supplementary Table S4

Accession*	Identity*	Cover*	E-value*	Classification	Hit from 2nd Blastp
*AAG52281.1	27.23%	91%	2.00E-27	RepHel	-
*AAM15154.1	26.46%	93%	4.00E-25	RepHel	-
*AAD25596.1	27.19%	93%	7.00E-24	RepHel	-
*BAB02793.1	25.39%	90%	3.00E-21	RepHel	-
*CAB91581.1	24.72%	90%	3.00E-21	RepHel	-
*AAD32757.1	25.92%	93%	1.00E-20	RepHel	-
*AAG51081.1	25.73%	94%	4.00E-20	RepHel	-
*BAB01023.1	26.91%	79%	6.00E-20	RepHel	-
*AAD15468.1	24.36%	94%	2.00E-19	RepHel	-
*AAG52315.1	25.73%	90%	2.00E-19	RepHel (cryptic)	-
*BAB11364.1	27.73%	86%	4.00E-19	RepHel	-
*CAB81576.1	24.71%	89%	2.00E-16	RepHel (cryptic)	-
*AAC28215.1	30.93%	41%	5.00E-16	93% cover, 69.83% identity to RepHel (best hit)	CAB91581.1
*AAC62789.1	24.79%	82%	8.00E-16	80% cover, 78.30% identity to RepHel (best hit)	AAD15468.1
*OAP18984.1	37.69%	29%	1.00E-12	RepHel	-
*AAD25621.1	37.50%	28%	1.00E-12	RepHel	-
*BAB02227.1	36.72%	28%	4.00E-12	RepHel (cryptic)	-
*AAD20107.1	24.52%	85%	1.00E-11	RepHel	-
*CAA0384207.1	33.57%	31%	6.00E-11	99% cover, 75.32% identity to cryptic RepHel (second best hit)	XP_010421223.1
*AAD15325.1	25.57%	79%	9.00E-10	RepHel	-
*AAG51717.1	30.06%	38%	7.00E-09	RepHel	-
*OAP08664.1	28.14%	57%	1.00E-07	RepHel (cryptic)	-
*NP_190738.1	30.93%	40%	2.00E-07	85% cover 44.56% identity to RepHel (best hit outside Brassicaceae)	XP_030934889.1
*CAA0385759.1	30.93%	40%	2.00E-07	85% cover 44.56% identity to RepHel (best hit outside Brassicaceae)	XP_030934889.1
*VYS60096.1	30.93%	40%	2.00E-07	85% cover 44.56% identity to RepHel (best hit outside Brassicaceae)	XP_030934889.1
*AAF06079.1	32.79%	27%	1.00E-06	RepHel (cryptic)	-
** Same hits	-	-	-	-	-

* Result of Blastp search using the human Pif1 domain (accession 6HPH_A) as a query against *Arabidopsis thaliana*.

** Result of Blastp search using the yeast Pif1 (accession NP_013650.1) as a query against *Arabidopsis thaliana* (only hits that did not overlap with the ones from the search using the human Pif1 domain).

Supplementary Table S5

Accession	Identity	Cover	E-value	Classification	Hit from 2nd Blastp
*AAK54302.1	27.39%	93%	5.00E-26	RepHel	-
*ABF97674.1	26.19%	93%	2.00E-25	RepHel	-
*XP_025876548.1	26.17%	93%	3.00E-25	RepHel (cryptic)	-
*AAP52492.2	27.25%	91%	8.00E-25	RepHel	-
*AAM92800.1	27.25%	91%	8.00E-25	RepHel	-
*XP_015613561.1	27.25%	91%	8.00E-25	RepHel	-
*AAN09850.1	27.25%	91%	1.00E-24	RepHel	-
*AAP52578.2	27.25%	91%	1.00E-24	RepHel	-
*XP_015613597.1	27.25%	91%	1.00E-24	RepHel	-
*XP_025879680.1	26.79%	92%	1.00E-24	RepHel (cryptic)	-
*BAF26194.2	27.25%	91%	1.00E-24	RepHel	-
*BAH91204.1	26.37%	92%	1.00E-22	RepHel	-
*EEC77075.1	25.93%	91%	2.00E-22	RepHel (cryptic)	-
*XP_015624412.1	25.44%	92%	3.00E-22	RepHel (cryptic)	-
*BAD81603.1	25.89%	88%	5.00E-22	RepHel	-
*BAF04484.1	25.89%	88%	6.00E-22	RepHel	-
*BAH93748.1	27.51%	93%	3.00E-21	RepHel	-
*XP_025879790.1	25.96%	93%	7.00E-21	RepHel (cryptic)	-
*AAO34493.1	26.45%	93%	2.00E-20	RepHel	-
*BAF08763.2	24.59%	89%	2.00E-19	RepHel	-
*AAX95750.1	25.32%	93%	3.00E-19	RepHel	-
*XP_025879706.1	25.24%	93%	2.00E-18	RepHel (cryptic)	-
*XP_015621010.1	26.68%	84%	3.00E-18	RepHel (cryptic)	-
*BAH92578.1	25.11%	93%	3.00E-18	RepHel	-
*CAD40309.2	25.11%	93%	3.00E-18	RepHel	-
*XP_025878111.1	26.52%	69%	5.00E-18	100% cover, 59.24% identity to RepHel (best hit)	BAD81603.1
*AAP54489.2	29.12%	62%	2.00E-17	100% cover, 90.36% identity to RepHel (best hit)	BAF04484.1
*BAC55632.1	26.88%	69%	4.00E-17	RepHel	-
*AAM93454.1	26.52%	66%	2.00E-16	100% cover, 100% identity to RepHel (best hit)	AAM92800.1
*BAH93891.1	36.31%	36%	5.00E-16	RepHel	-
*ABA99439.1	24.08%	81%	6.00E-16	99% cover, 70.98% identity to cryptic RepHel (best hit)	EEC77075.1
*ABA95256.2	25.32%	81%	7.00E-16	RepHel	-
*XP_025878227.1	38.46%	32%	2.00E-15	99% cover, 81.39% identity to cryptic RepHel (best hit)	XP_015637912.1
*XP_015620800.1	39.31%	74%	2.00E-15	RepHel (cryptic)	-
*CAE76063.1	34.93%	32%	2.00E-15	RepHel (cryptic)	-
*CAE76056.1	34.93%	32%	3.00E-15	RepHel	-
*XP_015637912.1	38.19%	68%	8.00E-15	RepHel (cryptic)	-
*BAC84865.1	37.50%	68%	1.00E-14	RepHel	-
*BAF22399.2	37.50%	68%	2.00E-14	RepHel	-
*BAD01692.1	37.50%	68%	2.00E-14	RepHel	-
*CAH66128.1	37.50%	68%	2.00E-14	RepHel	-
*AAX95983.1	37.50%	68%	2.00E-14	RepHel	-
*AAU44208.1	35.66%	67%	3.00E-14	RepHel	-
*ABA94634.1	35.66%	67%	3.00E-14	RepHel	-

*ABA94947.1	35.66%	67%	3.00E-14	RepHel	-
*BBD82308.1	35.66%	67%	3.00E-14	RepHel	-
*ABA95236.1	35.66%	67%	3.00E-14	RepHel	-
*AAT85173.1	34.93%	74%	3.00E-14	RepHel (cryptic)	-
*AAV44035.1	34.97%	67%	7.00E-14	RepHel	-
*AAK13103.1	26.43%	71%	9.00E-14	RepHel	-
*BAS88751.1	24.79%	71%	1.00E-13	RepHel (cryptic)	-
*XP_015627019.1	23.96%	69%	1.00E-13	RepHel (cryptic)	-
*BAF14458.1	24.79%	71%	1.00E-13	RepHel (cryptic)	-
*XP_025880731.1	24.58%	72%	2.00E-13	RepHel	-
*ABA94881.2	37.06%	68%	2.00E-13	RepHel	-
*AAK54292.1	33.95%	35%	2.00E-13	RepHel	-
*ABB47755.2	33.95%	35%	2.00E-13	RepHel	-
*KAB8095338.1	24.79%	71%	2.00E-13	RepHel (cryptic)	-
*EEC77085.1	24.79%	71%	3.00E-13	RepHel	-
*BAH94916.1	33.95%	35%	3.00E-13	RepHel (cryptic)	-
*CAD40616.1	24.79%	71%	4.00E-13	RepHel	-
*BAD68127.1	31.21%	32%	3.00E-12	RepHel	-
*EEC82986.1	32.81%	29%	1.00E-11	RepHel	-
*BAF04591.1	34.42%	32%	8.00E-11	RepHel (cryptic)	-
*ABA93595.1	37.80%	28%	1.00E-10	RepHel (cryptic)	-
*BAF29741.2	35.38%	65%	1.00E-09	RepHel (cryptic)	-
*BAG93269.1	27.44%	48%	2.00E-09	RepHel (cryptic)	-
*XP_025880729.1	24.93%	67%	2.00E-09	96% cover, 97.98% identity to RepHel (best hit)	BAH91022.1
*BAD68018.1	27.44%	48%	2.00E-09	100% cover, 100% identity to RepHel (best hit)	BAG93269.1
*ABA98117.1	35.38%	65%	2.00E-09	RepHel (cryptic)	-
*BAF08718.2	31.88%	35%	1.00E-08	RepHel	-
*ABA99343.2	50.88%	13%	3.00E-08	RepHel (cryptic)	-
*BAF24192.1	31.86%	25%	3.00E-08	RepHel	-
*ABA95557.1	25.94%	46%	4.00E-08	RepHel (cryptic)	-
*BAH94086.1	28.03%	35%	3.00E-07	RepHel	-
*AAT85232.1	31.43%	23%	3.00E-07	RepHel	-
*XP_015650422.1	28.03%	35%	3.00E-07	RepHel	-
*EEE67922.1	28.03%	35%	5.00E-07	RepHel	-
*BAH91022.1	24.44%	67%	3.00E-06	RepHel	-
*BAH94330.1	25.77%	52%	4.00E-06	RepHel	-
*XP_025880332.1	27.50%	34%	6.00E-06	RepHel (cryptic)	-
*ABA96519.2	28.17%	31%	0.016	RepHel	-
*BAF05239.1	23.11%	56%	0.02	RepHel (cryptic)	-
*AAQ56555.1	41.51%	9%	0.036	RepHel	-
*AAL75753.1	27.42%	24%	0.038	RepHel	-
**XP_025877503.1	26.47%	34%	2.00E-06	RepHel	-
**ABA97607.1	24.17%	33%	3.00E-05	RepHel (cryptic)	-
**KAB8082674.1	27.42%	17%	1.00E-04	RepHel (cryptic)	-
**BAH92476.1	31.47%	16%	0.047	RepHel (cryptic)	-

* Result of Blastp search using the human Pif1 domain (accession 6HPH_A) as a query against *Oryza sativa*.

** Result of Blastp search using the yeast Pif1 (accession NP_013650.1) as a query against *Oryza sativa* (only hits that did not overlap with the ones from the search using the human Pif1 domain).

Supplementary Table S6

Group*		Species	Accession	Identity	Cover	E-value	Classification	Hit from 2nd Blastp
Brassicales	*	Brassica napus	XP_022547407.1	26.84%	93%	8.00E-33	99% cover, 88.42% identity to RepHel	KAF8111651.1
	*	Camelina sativa	XP_010436751.1	27.27%	94%	1.00E-31	97% cover, 58.80% identity to RepHel	RID40682.1
	*	Brassica napus	XP_022551638.1	26.88%	93%	1.00E-30	RepHel	-
	*	Brassica rapa	XP_033148559.1	26.88%	93%	1.00E-30	RepHel	-
	*	Brassica napus	XP_013725746.1	26.88%	93%	1.00E-30	RepHel	-
	*	Brassica napus	XP_013719709.1	26.88%	93%	1.00E-30	RepHel	-
	*	Raphanus sativus	XP_018453621.1	26.67%	93%	5.00E-30	RepHel	-
	*	Brassica rapa	XP_033143195.1	27.90%	94%	9.00E-30	RepHel	-
	*	Arabidopsis thaliana x Arabidopsis arenosa	KAG7586339.1	26.92%	94%	9.00E-30	RepHel	-
	*	Eutrema salsugineum	XP_024013997.1	27.10%	93%	1.00E-29	RepHel	-
	**	Brassica napus	CAF2097984.1	26.52%	40%	3.00E-12	RepHel	
	**	Microthlaspi erraticum	CAA7047626.1	26.77%	42%	1.00E-11	RepHel	
	**	Microthlaspi erraticum	CAA7039386.1	26.24%	44%	1.00E-11	RepHel (cryptic)	
	**	Microthlaspi erraticum	CAA7015018.1	26.77%	35%	3.00E-11	RepHel (cryptic)	
	**	Brassica napus	XP_022544095.1	25.07%	35%	6.00E-11	RepHel	
	**	Raphanus sativus	XP_018460436.1	26.18%	33%	8.00E-11	RepHel (cryptic)	
	**	Brassica napus	XP_013694041.1	25.76%	33%	1.00E-10	RepHel	
	**	Brassica rapa	RID62868.1	25.93%	40%	1.00E-10	RepHel	
	**	Brassica napus	XP_022564371.1	26.10%	33%	1.00E-10	RepHel	
	**	Brassica napus	XP_013694540.1	27.87%	33%	1.00E-10	100% cover 95.47% identity to RepHel	XP_022548462.1
Commelinids	*	Zea mays	ONM60906.1	29.15%	92%	4.00E-28	RepHel	-
	*	Zea mays	ONM39160.1	28.05%	92%	2.00E-27	RepHel	-
	*	Sorghum bicolor	XP_002446095.2	27.16%	93%	1.00E-26	RepHel	-
	*	Musa acuminata	ABF70031.1	26.94%	93%	2.00E-26	RepHel	-
	*	Zea mays	AQK52428.1	28.33%	92%	2.00E-26	RepHel	-
	*	Zea mays	AQK60686.1	27.53%	92%	2.00E-26	RepHel	-
	*	Zea mays	PWZ05004.1	26.82%	95%	3.00E-26	RepHel	-
	*	Sorghum bicolor	XP_021314672.1	26.94%	93%	3.00E-26	RepHel	-
	*	Zea mays	PWZ25377.1	28.33%	92%	3.00E-26	RepHel	-
	*	Zea mays	ONM39853.1	27.53%	92%	5.00E-26	RepHel	-
	**	Zea mays	AQK64577.1	26.61%	35%	1.00E-13	RepHel	-
	**	Oryza sativa Japonica Group	XP_025876548.1	26.65%	33%	8.00E-14	RepHel (cryptic)	-
	**	Zea mays	AQK84207.1	26.61%	35%	2.00E-13	RepHel	-
	**	Zea mays	PWZ13396.1	27.27%	35%	2.00E-13	RepHel	-
	**	Zea mays	PWZ06906.1	25.84%	35%	2.00E-13	RepHel	-
	**	Zea mays	AQK97791.1	26.33%	35%	3.00E-13	RepHel	-
	**	Zea mays	ONM55810.1	26.61%	35%	4.00E-13	RepHel	-
	**	Zea mays	PWZ04632.1	26.99%	35%	5.00E-13	RepHel	-
	**	Zea mays	PWZ11828.1	26.61%	35%	8.00E-13	RepHel	-
	**	Oryza sativa Japonica Group	AAK13103.1	27.24%	33%	8.00E-13	RepHel	-

Malvids	*	Theobroma cacao	EOX92974.1	43.82%	98%	2.00E-102	No significant identity to RepHel	-
	*	Theobroma cacao	XP_017972716.1	43.59%	98%	1.00E-100	No significant identity to RepHel	-
	*	Durio zibethinus	XP_022774647.1	43.09%	98%	1.00E-99	No significant identity to RepHel	-
	*	Herrania umbratica	XP_021274232.1	42.99%	98%	2.00E-99	No significant identity to RepHel	-
	*	Corchorus olitorius	OMO60853.1	42.12%	97%	7.00E-94	No significant identity to RepHel	-
	*	Punica granatum	XP_031384248.1	40.79%	97%	1.00E-92	No significant identity to RepHel	-
	*	Rhodamnia argentea	XP_030518540.1	41.07%	98%	2.00E-92	No significant identity to RepHel	-
	*	Eucalyptus grandis	XP_010035891.2	40.75%	98%	3.00E-92	No significant identity to RepHel	-
	*	Rhodamnia argentea	XP_030518284.1	40.93%	98%	9.00E-92	No significant identity to RepHel	-
	*	Punica granatum	PKI33626.1	40.56%	97%	1.00E-91	No significant identity to RepHel	-
	**	Corchorus capsularis	OMO61479.1	40.06%	52%	7.00E-66	No significant identity to RepHel	-
Liliopsida	*	Colocasia esculenta	MQL92731.1	41.90%	96%	1.00E-89	No significant identity to RepHel	-
	*	Asparagus officinalis	ONK72744.1	36.41%	92%	1.00E-74	No significant identity to RepHel	-
	*	Asparagus officinalis	XP_020262994.1	37.87%	83%	2.00E-73	No significant identity to RepHel	-
	*	Zostera marina	KMZ75646.1	34.50%	70%	2.00E-44	No significant identity to RepHel	-
	*	Zostera marina	KMZ70362.1	35.32%	59%	4.00E-39	No significant identity to RepHel	-
	*	Zostera marina	KMZ65819.1	36.67%	55%	1.00E-36	No significant identity to RepHel	-
	*	Zostera marina	KMZ67804.1	43.62%	33%	3.00E-32	No significant identity to RepHel	-
	*	Zostera marina	KMZ56065.1	36.00%	39%	2.00E-30	No significant identity to RepHel	-
	*	Zostera marina	KMZ65715.1	32.45%	60%	1.00E-29	No significant identity to RepHel	-
	*	Zostera marina	KMZ68271.1	43.80%	31%	4.00E-28	No significant identity to RepHel	-
	**	Same hits	-	-	-	-	-	-

* Results of Blastp searches using the human Pif1 domain (accession 6HPH_A) as a query against the corresponding group. The 10 best hits from each search are shown.

** Results of Blastp searches using the yeast Pif1 (accession NP_013650.1) as a query against the corresponding group. The 10 best hits from each search are shown (only hits that did not overlap with the ones from searches using the human Pif1 domain).

Supplementary Table S7

Group	Species*	Accession*	Identity*	Cover*	E-value*	Classification	Hit from 2nd Blastp
Brassicales	*Brassica oleracea*	XP_013639271.1	28.21%	84%	7.00E-22	RepHel	-
	Raphanus sativus	XP_018465781.1	26.99%	90%	9.00E-21	RepHel (cryptic)	-
	Brassica rapa	XP_033147243.1	26.82%	84%	3.00E-20	RepHel	-
	Brassica oleracea	XP_013629542.1	26.72%	90%	2.00E-19	RepHel (cryptic)	-
	Eutrema salsugineum	XP_024006484.1	25.64%	88%	7.00E-19	RepHel (cryptic)	-
	Eutrema salsugineum	XP_024007971.1	28.40%	81%	3.00E-19	100% cover, 95.18% identity to RepHel (best hit)	XP_024004792.1
	Eutrema salsugineum	XP_024014429.1	25.46%	88%	8.00E-19	RepHel (cryptic)	-
	Capsella rubella	EOA12259.1	26.56%	82%	5.00E-19	RepHel (cryptic)	-
	Capsella rubella	XP_023633617.1	26.69%	84%	2.00E-18	RepHel	-
	Capsella rubella	EOA12327.1	27.17%	84%	2.00E-18	RepHel (cryptic)	-
Commelinids	*Oryza sativa*	XP_025876548.1	27.31%	84%	2.00E-26	RepHel (cryptic)	-
	Oryza sativa	AAK54302.1	26.81%	82%	7.00E-25	RepHel	-
	Oryza sativa	BAH94916.1	26.38%	82%	7.00E-25	RepHel (cryptic)	-
	Sorghum bicolor	OQU91688.1	25.93%	80%	2.00E-24	RepHel	-
	Zea mays	AQK95425.1	28.15%	84%	3.00E-24	RepHel (cryptic)	-
	Triticum dicoccoides	XP_037419736.1	24.84%	82%	3.00E-24	RepHel	-
	Triticum dicoccoides	XP_037474121.1	26.05%	83%	3.00E-24	RepHel	-
	Triticum urartu	EMS67201.1	24.84%	82%	4.00E-24	RepHel	-

Sorghum bicolor	XP_021305262.1	27.43%	80%	2.00E-24	100% cover, 99.19% identity to cryptic RepHel (best hit)	XP_002444425.2
Sorghum bicolor	XP_002452524.1	28.15%	80%	3.00E-24	99% cover 94.29% identity to cryptic RepHel (best hit)	XP_002444425.2

* Results of Blastp using the best hits from searches in malvids (EOX92974.1) and Liliopsida (MQL92731.1) (see Table S6) as queries against Brassicales and commelinids, respectively. The 10 best hits from each search are shown.

Supplementary Data S1. Consensus sequences of the Hel and Rep domains from *Helentron* (including *Helitron2*) and *Helitron* variants.

```
>Helentron_Hel_consensus
LNKEQREFFYHVLHLIKTSPEPFYLFLSGGAGVGKSHLIKALYQALLKYLNSLPGFRGPKVLLAAPTGKAAFNISGTTLHSLLKLPISQSPYKPLSASRLNTLRCKLRDLKLLIIDEI
SMVGSRMFNWINNRLRDIKGSDEPFGGISIIAVGDLFQLPPVGDKPIFKDPENYILARNLWWEFFKMFELTEIMRQRDDKAFAEALNRLREGQLTDEDIKLLKQRVVTEKNRPSDALH
LFATNDEVNEYNNEVLDRLKGEKIQIKAIDVVIGARTKADTRKTGGLAKLLQLAVGARYMLTRNLDVEDGLVNGAGTVK

>Helitron_Hel_consensus
QLNEEQRRAYDTILAAVSDGSGGLFFLDGPGGTGKTFLYKTLLAAIRSQGKIVLCVASSGIAALLLPGGRTAHSRFKLPLNLNETSVCGIKKQSKLARLLKEAKLIIWDEAPMAHKHA
LEAVDRLLKDIMNNDQPFGGKVVLLGGDFRQILPVVPRGTRADIVNACLKSSYLWPNFKTLKLTKNMRVTSGEDQEFSEWLLKIGDGNLNVDGEGLIEIPEDFLIIEEIIEEIYPDII
DAQNPEFFSERAILAPKNEDVDELNEYILDRLPGEERIYLSIDSVVTDDSEAENYPTEFLNSLNPSGLPPHELRLKVGAPVMLLRNLNPKRGLCNGTRLVITKL

>Helentron_Rep_consensus
PTLFCTFSAAETKWPHLLKILGKLVDNKYTEDELENLDWDEKCRLIQSDPVTCARYFDKRVDALLTTLLLSPAQPFGKVVDYFYRVEFQQRGSPHIHMLLWLEDAPKFGVDSDEEVIE
FIDKIITCQKPDLNELKDLVNRQTHRHTHTCKKKNKKSCRFNIPQPPMPKTMILYPLEDDDSERKELKEKWKKIKDLLNDKEGSFDTFEEFLAKLNLSEEDYLLAVRSSLKRPTVFLK
RQPNELRINNYNPDILKAWRANMDIQFILDVYACAMYIVSYIS

>Helitron_Rep_consensus
PDLFITFTCNPKWPEITENLLPGQTAEDRPDIVARVFKLKLKSLLNDLTKKHVLGKVRAWIYVIEFQKRGLPHAHILLILKEEDKPRTPEDIDKIISAEIPDKETDPELYEIVTSNMI
HGPCGAANPSSPCMVDGKCSKRFPKKFQEETVINVDGYPLYRRRNNGRGVEKGGIELDNRWVVPYNPYLSLKYNAHINVEVCNSIKSIKYLFKYVY
```

Supplementary Data S2. Final alignment of Pif1-like amino acid sequences used in the phylogenetic and NMDS analyses.

```
>Helen_A_alb
-----TTNAEQRD-LILQM------IHS----L-HSY-------------DE----SSKP-----MQIFFTGPAGCGKTFTLRILMET---I-N--R--YSQ----------------
AHNAQK-NAYVACASTGKAAVAI-------GG------TTVHSA-FRI---T--M------SR-R---A---N-------SK---LS------------------FEML---------
------Q---L-------------YRNAF--AN--IKAVIIDEVSM--IGADILNT-IHARLQDIS----------------G---------N----------------------YD-
--------------------------DPFGG-----INIVFCGDLRQLPPVN-----------------AR-----------------------------------------------
--------------------------------------------PVYKP--TG---NSF----HGA-VLWQA--------L-DFLPLVKVMR----Q-T-----------D--V----
---------EFSSILTKIGNG---QQM-T------A--------------EE------------------------------TK--LI-ES--RF---------RTVE----------
-------------------------------------------------------------------WCK-QN---------------------A---------P-G-----------
A-------------------IRLYHRNADVEAYNNEVLH-NQ----D---A----------------------------------------------------L-D-C----------
----------I---------------------ADD------VFA--------------------G--------------------YK-------DAGQLASSRI-----------K--
--------------------------------------------------------------LYK-------------MS--------------------VV-----------E-T--
------GGL----------PYLL-----RLSVGMPYMITTNVD----------V----E--D-----G-----VVNGAIGELKYI
>Helen_C_qui
-----ATNAEQRA-LILHV------IHL----M-HCYE------------------EHEP-----LQVFLTGPAGSGKTFVLRALMET---I-N--R--YSQ----------------
THNSRD-NAYVASASTGKAASAI-------GG------TTLHHA-YHI---T--M------SR-Q---A-----------AK---MN------------------FETL---------
------Q---M-------------YRNEM--QN--IKFHIIDEVSM--VGAHTLNT-AHIRLQDVY----------------M---------I----------------------YD-
--------------------------VPYGG-----VNVLVSGDVMQLPPVN-----------------AR-----------------------------------------------
--------------------------------------------AVFKP--PS---NTI----CGA-VVWQS--------L-MFHELKRVMR----Q-A-----------D--K----
---------QFSDILTKIGNG---LKL-T------A--------------DE------------------------------TK--LI-ES--RF---------FTKE----------
------------------------------------------------------------------DLSK-ED---------------------T---------G-G-----------
A-------------------VRLFHRNIDVTSYNNEALR-NI----D---G----------------------------------------------------L-D-Y----------
----------T---------------------ADD------TFA--------------------G--------------------YK-------TAEQLATARI-----------K--
--------------------------------------------------------------LYK-------------MS--------------------LA-----------E-T--
------AGL----------QYTT-----KFCPGMPYMVTTNVN----------V----E--D-----G-----IVNGAIGDLMYV
>Helen_G_occ
-----KLNAEQRE-LILEV------IHR----L-HDP-------------------NSEA-----IQIFLTGPAGCGKTYTLKALMET---Y-N--R--YAQ----------------
EHNNMN-NAYISTATTGKAATAL-------NG------VTVHSA-FKL---A--L------SN-R---A-----------HQ---LS------------------NDVL---------
------Q---T-------------YRHHL--RN--VRCIIIDEISM--CSSHVFHG-VNTRLQAMT----------------G---------E----------------------FD-
--------------------------ANFGG-----LDLFACGDLKQLPPVR-----------------AA-----------------------------------------------
--------------------------------------------PVFTA--TK---SSI---GGKA-ILWQS--------L-NYFPLVQVVR----Q-S-----------D--I----
---------GFSTLLTKIGCG---EAL-S------Q--------------AE------------------------------TD--KI-QS--RF---------RTRR----------
-------------------------------------------------------------------WCD-AN--------------------LS---------T-E-----------
V-------------------MRLYHTSADVQSYNDSAIP-VT----E---S--------------------------------------------------T-H-N-H----------
```

```
----------I---------------------ATD------IYT--------------------G--------------------YR-------TEAERRNAIG----------K--
--------------------------------------------------------------MHR-------------KD--------------------TR-----------D-T--
------GNL----------PYTI-----TLAEGYPYMLTVNVD----------V----E--D-----G-----LVNGAIGQLRHV
>LIED01008227.1_Bemisia_tabaci
-----RLNVKQRE-ICMHV------LHC----A-KIDR--------------------GL-----QHIFVSGAGGVGKSTVIKTIFQS---V-T--R--HYDNNL-------------
NFAPDS-VKVLLCSYAGKAAYII-------GG------VTVHTA-FVL---P--V------TK-Y---G---G-----QMPP---LN------------------PASA---------
------S---H-------------LITEL--RD--CQWIIIDEVSM--LGGKMATY-IEQRCREIK----------------R---------N-----------------------T-
--------------------------DPFGG-----INIIVVGDFGQIPPVQ-----------------DS-----------------------------------------------
--------------------------------------------MIFKP-PNITELSLL---LESN-FNWRD--------F-KMFELTEIMR----Q-K----------GE--K----
---------AFIEALNNLCWG---T-L-T------E--------------ED------------------------------MK--LF-KS--RQ---------VKNE----------
-----------------------------------------------------------------------SD--------------------VP---------K-E-----------
A-------------------IRLYCFNKQVDAFNAAKIF-EC----P---E--------------------------A-----------------------E-S-V-S----------
----------E---------------------AKN------TVL--------------------G--------------------KA-------TQSRIDYA----------VR-S--
-----L---------------------------------------------------------MN-------------KK--------------------LD-----------D-L--
------KGL----------PHRV-----VFKVGIKYMITINVS----------V----R--D-----G-----LVNGAVGTLVSI
>Helen_L_ser
-----SLNNKQKL-YLSHL------LH-------NIKLG-------------------HS-----FYEFVGGGAGVGKSRLISAIYQS---L-N--Y--RLNFIP-------------
GTDPSL-IKVLLCAPTGKAAFGI-------GG------ATLHSM-FSL---P--I------NQ-S---A---A-----ELRP---LS------------------SDTA---------
------N---A-------------LYSKF--LN--LKLLIIDEISM--VGSKMLRY-LDARLKQVF----------------K---------S-----------------------T-
--------------------------APFGG-----ISLVVFGDLRQLPPVG-----------------DS-----------------------------------------------
--------------------------------------------WIFSA-PSNDPYSVI----YGS-TLWDM--------F-KYFELNEIMR----Q-R----------ED--Q----
---------AFARALNNMACG---K-M-S------E--------------QD------------------------------VS--LM-KS--RE---------VLE-----------
-----------------------------------------------------------------------SN--------------------VP---------A-E-----------
A-------------------IHLLCTNAEVDNWNAIKLN-SI----T---T--------------------------E-----------------------E-S-I-S----------
----------Y---------------------ADD------QVK--------------------S--------------------VG-------LSRENRAS---------ILE-N--
-----V---------------------------------------------------------KV-------------FK--------------------TS-----------E-T--
------QGL----------RYDL-----KLKTTAKYMVTVNIN----------T----S--D-----G-----LVNGATGQLMQI
>Helen_M_sac
-----IFNCDQRH-FVLHV------GHI----F-TQE-------------SP------AP-----FYYFVSGGAGVGKSLLIKGLYQY---L-M--F--MFNRVP-------------
GINPDD-ARILLCAYTGKAAFGI-------GG------QTVHST-FGL---P--V------SQ-C---G---Q-----TMPE---LS------------------ASTA---------
------N---T-------------LACKL--AK--VRLIILDEISM--LGSRTLNQ-INRRLQQVF----------------H---------T-----------------------D-
--------------------------APFAG-----ISIISVGDFNQLPPVG-----------------DN-----------------------------------------------
--------------------------------------------WVFQPNSSRNPLAPL----AGA-PLWEP--------F-RLFMMTKIMR----Q-R----------DD--L----
---------AFAVALNNMAVG---R-M-T------P--------------MD------------------------------IE--LI-NS--RC---------YSIN----------
--------------------------------------------------------------------------------------------TLPIEV------H-G-----------
A-------------------IHLFATNNEVDKYNNQVLS-RM----N---T--------------------------E-----------------------G-C-S-V----------
----------K---------------------ALD------VVS--------------------G--------------------AP-------NPQAAKKA----------LQ-S--
-----V---------------------------------------------------------KA-------------LA--------------------TM-----------Q-T--
------YGL----------PKNL-----FLRVGARYMVTVNMD----------T----T--D-----G-----LVNGTTGILKAI
>Helen_M_cir
-----RLNSKQRL-LLTHI------IHH----I-RNSRDESRYRYFDTAIAPPTNTTFAP-----LHLLVTGGAGTGKSMLINTLYQS---L-I--R--EFDSDR-------------
DRDMAS-PSVLLCAPTGIAAFNI-------GG------QTIHSI-FDL---P--I------SQ-G---T----------LST---LS------------------ASVS---------
------H---S-------------MSVAL--RD--LRVVIIDEISM--VGSLQFGW-IDKRLRDIF----------------D---------S-----------------------Q-
--------------------------KPFGG-----ISIFVFGDFLQLPPVM-----------------AA-----------------------------------------------
--------------------------------------------PVYSR--AVSPGNAIPG-LLSF-NLWSL--------F-EPYKLTQIMR----Q-R----------DD--L----
---------RFAVALNHLAIG---E-L-T------D--------------AD------------------------------RS--LF-QS--RV---------VNLS-----SEQM-
-------------------------------------------------------------------QQI-KDFA---------------AFPLPADENTNDGAT-Q-----------
P-------------------IILCRTNNEVENFNRLILD-GI----Q---G--------------------------E-----------------------E-A-V-S----------
----------V---------------------AFD------VSM--------------------G--------------------VE-------SQFDQNAI----------ER-NT-
--------------------------------------------------------------GDN-------------AN--------------------PR-----------N-V--
------KGL----------IKNL-----RLKVGGKYIISRNVK----------T----S--D-----G-----IVNGAGCILKRI
>Helen_C_cuc
-----KLNNLQRD-FLNHV------INH----I-RHNTD--------GVDGP------MP-----LKLFVTGGAGTGKSLLIKTLYQA---L-V--R--FYDEDP-------------
HRDYNS-PTILLTAPTGKAAFNI-------KG------QTINSA-FLL---P--I------NQ-S---D----------INQ---LS------------------PEIS---------
------H---S-------------MTVAL--AE--LRVVIIDEISM--ISSRVFLW-IDKRLRDIF----------------D---------S-----------------------E-
--------------------------QPFGG-----RHVILFGDFLQLPPVK-----------------GQ-----------------------------------------------
--------------------------------------------SIFAK--PTDNLDLLTSTLRVQ-EIWHS--------F-KVHRLTEIMR----Q-R----------DD--A----
---------LFAKALNNMAIG---A-M-T------P--------------SD------------------------------VA--LF-EG--RL---------VASL-----PDDV-
---------------------------------------------------------------------Q-DN--------------------RD-----------D-------C---
V-------------------VRLYHTNDSANHCNTTILF-NI----E---D--------------------------E-----------------------S-Y-E-S----------
----------A---------------------CFD------KVV--------------------G--------------------NSV------DASTKRRY----------LD-A--
-----IR-------------------------------------------------------SRN-------------LP--------------------AH-----------E-T--
------MGL----------HELL-----MLKVGARYMVINNLD----------T----S--D-----G-----LINGTTGTLKKI
>Helen_D_rer
-----SLNKTQAA-IFYTI------RQWCQNRV-WGL-------------NP------EQ-----FFYFVSGGAGCGKSHVIKCVYTE---A-T--K--ILRQLPQ---------
LREDGDLSI-PTVLLSAFTGTAAFNI-------SG------KTLHSI-LKL---P--K------NL-K---P---P------YQG---L-------------------GNSL------
---------D---D-------------VRAEL--RH--VEILIIDEISM--ISKDFFAY-INWRFQQIR----------------G---------S----------------------
-K---------------------------KPFGG-----ISVIVVGDFYQLPPPG-----------------KA---K--P--------------------------------L----
-----------------------------------------------CVYEE--------------DVL-DFWKD-----H--F-QIVTLTEIMR----Q-K----------ED--L-
------------SFAQLLNRLRVK---R-K-SDA--LKE--------------ED------------------------------RA--LL-LQ--AV---------KNPQ-------
-----------------------------------------------------------------------------------------------DCP---------R-D--------
---A-------------------LHIFATNKEVHSHNCETVN-AL----H---A--------------------------D-----------------------I-V-T-I-------
-------------D---------------------AED------YRK--------------------D--------------------PR-------TGGMKRQT-------------
KP------V----------------------------------------------------------------------------------------------MG-----------K-
K--------DNL----------LDTI-----QVAVGVRIMVIRNLD----------V----E--D-----G-----LVNGCFGKIGNI
>Helen_L_roh
```

```
-----SLNETQAA-IFYTV------RQWCQKRV-WGH-------------NP------EQ-----FFYFLSGGAGCGKSHVIKCIHTE---A-T--K--ILRQLPR---------
LREEGDLSV-PTVLLSAFTGTAAFNI-------SG------KTLHSL-LKL---P--R------SL-K---P---P------YQG---L-------------------GNAL------
---------D---E-------------VRAGL--RD--VEILIIDEVSM--ISKDMFAY-INWRLQQIK----------------G---------S----------------------
-K---------------------------KPFGG-----ISCLVVGDFYQLPPLG-----------------KA---K--P--------------------------------L----
-----------------------------------------------CVFEE--------------DVL-DFWKD-----S--F-QIITLTEIMR----Q-K----------ED--L-
------------AFAELLNRLRVK---Q-K-TEA--LRE--------------DD------------------------------RA--LL-LQ--AV---------KKPE-------
-----------------------------------------------------------------------------------------------DCP---------R-D--------
---A-------------------LHIFATNKEVQKYNTETVQ-AL----Y---T--------------------------D-----------------------I-I-T-I-------
-------------D---------------------AED------YRK--------------------D--------------------PK-------TGGMKRLN-------------
KP------V----------------------------------------------------------------------------------------------TG-----------K-
K--------DDL----------LDLI-----EVAVGVRVMITRNLD----------V----E--D-----G-----IVNGCFGKIGNI
>Helen_A_cal
-----SLNETQAS-IFYAV------REWCFKLV-WGH-------------CP------EQ-----FFYFVSGGAGCGKSHVIKCIYEE---A-T--K--ILHQLPR---------
FRDQADMSY-PAVLLTAFTGTAAFNI-------SG------KTLHSL-LKL---P--R------SL-K---P---P------YQG---L-------------------GNAL------
---------D---E-------------VRASL--SN--AEILIIDEISM--VSKDLFAY-IHWRLQQIK----------------G---------N----------------------
-K---------------------------KPFGG-----MSILAVGDFYQLPPLG-----------------KA---K--P--------------------------------L----
-----------------------------------------------CVYED--------------NVL-DLWKD-----Y--F-HMVNLTEIMR----Q-K----------DD--H-
------------SFAEVLNRIRVK---Q-K-TDS--LEA--------------DD------------------------------KA--LL-TQ--AI---------HDIK-------
-----------------------------------------------------------------------------------------------DCP---------S-N--------
---V-------------------LHIYATNKEVDKHNSATVT-AL----H---S--------------------------D-----------------------I-I-N-I-------
-------------Q---------------------AED------YRK--------------------D--------------------RR-------TGDMVLLA-------------
EM------M----------------------------------------------------------------------------------------------KG-----------N-
K--------GDL----------PDNI-----QAAPGVRVMIIRNLD----------V----E--D-----G-----LVNGTFGTITNI
>Helen_A_mil
-----TLNKKQKE-FFYHI------LHL----I-KTSD--------------------KP-----FYYFLSGGAGVGKSHLVKSLYQA---A-L--K--YYNSKA-------------
GEDFNE-VKILLLAPTGKAAFGI-------KG------NTIHST-FAI---P--V------CQ-S---L---K-----NYKP---LD------------------SSRL---------
------N---T-------------LRCKL--HA--VKLIFLDEISM--VGNTMFNIQINNRLKDIK----------------G---------S-----------------------R-
--------------------------EFFGG-----VSIIALGDLFQLEPVM-----------------DS-----------------------------------------------
--------------------------------------------YVFKN-MKNAEYAAL-----AP-NIWQE-----L--F-TMFELDEIMR----Q-R----------DS--K----
---------AFAEILNRLREG---N-H-T------P--------------ED------------------------------IA--KL-KQ--RC---------ISEN-----C----
-----------------------------------------------------------------------PN--------------------YP---------L-D-----------
I-------------------PHLFIQNSKVDEFNNKVHL-AA----T---G--------------------------D-----------------------K-Y-N-I----------
----------R---------------------AID------SVI--------------------G--------------------AN-------SAELRDKI----------LK-Q--
-----I----------------------------------------------------------P-------------LD--------------------PR-----------K-T--
------KQL----------ASNL-----QLAAGERTELVVNLR----------T----D--D-----G-----MTNGAGNIIKRI
>Helen_N_vec
-----TLNKEQKE-FFYHV------LHL----I-KTSG--------------------EA-----FYCFLSGGAGVGKSHVTKALYQA---A-L--K--YYNTRP-------------
GVNFAE-TKILMLAPTGKAAYNI-------KG------NTIHSA-LAV---P--A------CQ-S---L---K-----NYKK---LD------------------SSRL---------
------N---T-------------LRCQI--GG--LKLIFVDEISM--VGNTMFNVQFNNRLKDIK----------------G---------S-----------------------S-
--------------------------LPFGG-----VSIVAIGDLFQLQPVM-----------------DG-----------------------------------------------
--------------------------------------------YIFKD-MDNDEYGVL-----AP-NVWQE-----L--F-KMFELKEIMR----Q-R-G--------ES--K----
---------DFAELLNRLREG---N-H-T------K--------------EG------------------------------II--KL-KE--RI---------LNHT-----S----
-----------------------------------------------------------------------AQ--------------------YP---------K-D-----------
A-------------------PHLFIQNAKVNDFNYKAHN-AL----Q---G--------------------------P-----------------------K-Y-S-I----------
----------K---------------------AHD------TVI--------------------G--------------------TD-------SEELRDKI----------LK-Q--
-----I----------------------------------------------------------P-------------KD--------------------PR-----------K-T--
------KQL----------HSVL-----HLAIGERTEISLNTR----------N----D--D-----G-----MTNGAGSVINVS
>Helen_A_dig
-----TLNKEQKE-FFYHV------LHL----V-KTSD--------------------EP-----FYCFLSGGAGVGKSHVTKALYQA---A-L--K--YYNSRA-------------
GDSFAQ-IRVLMLAPTGKAAYII-------KG------NTIHSA-LAI---P--A------CQ-S---L---K-----TYKR---LD------------------SNRL---------
------N---S-------------LRTQL--GG--VKLIFIDEISM--VGNTMFNVQIDNRLKDIK----------------G---------S-----------------------P-
--------------------------LPFGG-----VSIIAIGDLFQLQPVM-----------------DD-----------------------------------------------
--------------------------------------------YIFND-LK-TEYGIL-----AP-NLWQE-----L--F-KMFELKEIMR----Q-R----------ES--K----
---------QFAELLNRLREG---K-Q-T------N--------------ED------------------------------IR--VL-KQ--RT---------LQPS-----G----
-----------------------------------------------------------------------SN--------------------YP---------V-D-----------
A-------------------PHLFIQNAKVNDFNDKVHQ-AS----Q---G--------------------------T-----------------------K-Y-N-I----------
----------R---------------------AHD------SVI--------------------G--------------------AT-------SQEVRDKI----------LK-Q--
-----I----------------------------------------------------------P-------------LD--------------------PR-----------K-T--
------KQL----------HGLL-----NIAVGERTEISLNTR----------I----D--D-----G-----MTNGAGNVIKLI
>Helen_D_gig
-----SLNKKQKE-FFYHT------LHV----I-KTSD--------------------KP-----FYCFLSGGGGVGKSHLTRSVYQA---A-L--K--YYNTRA-------------
GEDFHQ-VKILLLAPTGKAAYLI-------NG------NTIHST-LAI---P--A------SQ-S---L---R-----HYKP---LD------------------ASRL---------
------N---T-------------LRSRL--GG--VKLILLDEVSM--VGNNMFTVQINNRLKDIK----------------G---------S-----------------------K-
--------------------------EDFGG-----VSIIGIGDLFQLPPVF-----------------DG-----------------------------------------------
--------------------------------------------YIFND-IQNSEYSIL-----SP-NLWNE-----H--F-RMFELTEIMR----Q-R----------EN--K----
---------EFAEILNRLREG---N-Y-T------N--------------DD------------------------------LL--KI-KT--RC---------VTET----------
------------------------------------------------------------------------E--------------------CP---------P-D-----------
A-------------------PRLFIRNDNVDKYNEAVYN-RA----T---G--------------------------N-----------------------K-Y-S-I----------
----------K---------------------AQD------SVI--------------------G--------------------TN-------TVELRDKI----------LN-Q--
----VI----------------------------------------------------------K-------------MT--------------------LR-----------N-T--
------KQL----------ARTL-----QLAVGLRTEMVLNVR----------T----D--D-----G-----LTNGASNIIKLI
>Helen_L_ana
-----NMNRQQFE-FFHHV------LHL----I-KSNS--------------------DP-----FHIFLSGGAGVGKSFVTRALYQG---I-L--K--YLSSLP-------------
GEDFRT-IRVALVAPTGKAAYNI-------GG------HTIHSL-LKI---P--A------NQ-S---L---R------YKR---LS------------------ADVL---------
------N---S-------------FRYKL--GS--LKVLFIDEVSM--VGSKMLSF-INERLKELK----------------N---------N-----------------------D-
--------------------------RLFGG-----VSIVAIGDLFQLKPVF-----------------DN-----------------------------------------------
--------------------------------------------WIFEN--PNNDYYPL-----AT-NLWQK-----H--F-HMYELTEIMR----Q-K----------DS--K----
---------EFAEILNRLREG---H-H-T------D--------------KD------------------------------IK--VL-SE--RQ---------LDQS----------
```

```
-----------------------------------------------------------------------KT--------------------SQ---------K-P-----------
L-------------------LHVFQTNSLVENFNHASYQ-NA----K---G--------------------------E-----------------------K-F-Q-I----------
----------A---------------------ATD------TIT--------------------G--------------------PV-------PKHLENNI----------KK-Q--
-----I----------------------------------------------------------P-------------LD--------------------HK-----------K-T--
------MNL----------RRIL-----HIAVGERTEVVLNVD----------T----E--D-----G-----ITNGAPNVVKLV

>Helen_O_fav

-----SLNEKQRQ-FFYHV------LHS----I-KTRD--------------------DP-----LRLFLSGGAGVGKSTVTNALYEA---L-I--R--YLNSIA-------------
GENPDD-VKVVKAAPTGKAAFNI-------KG------NTLHSA-FKI---P--A------NR-G---F---E------YCA---LD------------------SDRL---------
------N---T-------------IRAQL--KK--LKTIFIDEISM--VGSGMFNF-LNARLQQIM----------------G---------T-----------------------K-
--------------------------ELFGG-----ISLITVGDLFQLKPVF-----------------DK-----------------------------------------------
--------------------------------------------WIFEN--SAIGYSAL-----AS-NIWTE-----N--F-TLFELTEIMR----Q-K----------DD--R----
---------EFAELLNRLREG---K-H-S------E--------------DD------------------------------VA--IL-KQ--RL---------LKVT-----PQE--
-----------------------------------------------------------------------DN--------------------YP---------M-N-----------
M-------------------THLFTTNASVDAHNNALYT-IS----K---T--------------------------D-----------------------K-A-Q-V----------
----------K---------------------AVD------IVV--------------------G--------------------DI-------ADDLKQQM----------KN-K--
-----I----------------------------------------------------------P-------------ED--------------------PT-----------K-T--
------MGL----------YSLV-----SLATMAKYDLTTNID----------V----T--D-----G-----LTNGAECMIENI

>Helen_C_gig

-----DLNEQQKE-FFNHV------LHW----L-KTKT--------------------EP-----LYAFLSGGAGVGKSVLTRALYQA---L-L--K--YYSHRI-------------
HENPDN-IHVMLCAPTGKAAHNI-------NG------TTLHSA-FCI---P--V------GR-G---F---A------YKP---LD------------------MQQL---------
------N---T-------------LRTKF--IS--LKVIFIDEISM--VGHNMFNF-INLRLQEIK----------------G---------C-----------------------T-
--------------------------LPFGG-----TSIVTVGDLFQLRPVM-----------------DN-----------------------------------------------
--------------------------------------------WIFTQ--SNKGYGPL-----AA-NLWRD-----N--F-KLFELTVIMR----Q-R----------DD--K----
---------IFAELLNRIREG---N-Q-T------E--------------ED------------------------------LS--LL-KT--CV---------KEEC-----QEI--
--------------------------------------------------------------------------------------------------------S-N-----------
V-------------------PHLFTTRNEVTQYNYDIYN-KA---DN---S--------------------------E-----------------------K-V-C-I----------
----------K---------------------AID------WVI--------------------S--------------------SC-------DENVKAKV----------LS-R--
-----I----------------------------------------------------------P-------------DD--------------------YA-----------K-T--
------MGL----------SAEL-----FLVIGIAAEITSNVN----------V----Q--D-----G-----ITNGASCVIKQF

>Helen_S_pur

-----TLNTGQYK-VFSYI------NNWCVDLV-KSRK---------THVDL------QP-----VQLCVTGGAGTGKSHLISTIYQM---A-I--R--TLKHE--------------
GSNPEA-VRVLLTAPTGTAAFNI-------QA------STLHST-FLL---P--L------GQ-T---K---V------YKK---LS------------------DQKR---------
------N---T-------------LRCKL--AD--LDILIIDEVSM--VGCDLLMT-VDQRLREIK----------------G---------V-----------------------N-
--------------------------KIFGG-----ISVLAFGDLYQLAPVC-----------------QK-----------------------------------------------
--------------------------------------------FVFED--AADLFARL-----AG-SLWQD-----N--F-QFAELDEIMR----Q-K----------DD--R----
---------AFAELLNRIRVG---E-Q-T------Q--------------ED------------------------------MT--TL-EQ--CI---------ISPS----------
---------------------------------------------------------------------D-DN--------------------YP---------S-D-----------
A-------------------LHVFATNARVNEYNTEKLS-KV----E---G--------------------------P-----------------------I-R-R-C----------
----------I---------------------AVD------KKP--------------------S----------------------------------------------CLK-SH-
-----V---------------------------------------------------------TS-------------TD--------------------AR-----------F-T--
------GGL----------PHVL-----ELKVGSRVMLTRNMD----------V----T--D-----G-----LVNGALGTVVDF

>XP_002772304.1_Perkinsus_marinus_ATCC_50983

-----KLNEDQAR-IVDEV------RQQAR-NIYAATEE--------IAARP------KP-----IQWFLTGGAGVGKSFVIHVIRNL---V-Q--R--ELHL---------------
MDFPKR-VGCLVTATTGCAAFAI-------QG------ATLHTT-FHL---P--L------TV-G---T---YQ----SMEP---LS------------------QAKV---------
------E---E-------------VRESF--LG--VEFLIIDEVSM--LGYPGLVA-VHQRLQQIR----------------D---------C-----------------------E-
--------------------------DWFGG-----VNVICVGDMFQLPPVM-----------------QT-----------------------------------------------
--------------------------------------------PVYGQLRGLAGMKQL-----AV-HLWKD-----L--F-EIRELREIMR----Q-Q----------NG--S----
---------AFAEALNRLRLG---E-S-T------E--------------DD------------------------------LR--LF-RS--RI---------VNSA----------
----------------------------------------------------------------------------------------------P---------P-D-----------
C-------------------LRLFRTNAACDAYNTEMLS-KT----R---G--------------------------V-----------------------A-Y-E-I----------
----------V---------------------AKT------T---------------------PG-----------------------------------------------------
-----I----------------------------------------------------------T-------------IT--------------------DV-----------Q-A--
------GGV----------REVL-----TLKTGCRIMIVRNVD----------I----E--R-----G-----IVNGATGTLVKI

>Hel2_F_oxy

-----SLNPEQRI-VYDTV------MGH----F-LTQ-------------DP------SQ-----LLLHVDGGGGTGKSYLINLLSAH---L-Q--S--ATGG---------------
----RG-TPVWRAAPTGVAGNQI-------SG------TTLHSL-LHL---P--I------NK-D--------------FKP---LS------------------PVDK---------
------T---Q-------------LQKKL--KD--IKYLIIDEKSM--LRLRQLSW-IDDRLREAFPN--------------R---------N-----------------------E-
--------------------------EFFGG-----LNILLVGDFFQLPPVL-----------------QK-----------------------------------------------
--------------------------------------------PLCYD-KEVQGV-EI----KGR-NAYRR--------FDKSVFLKVVQR----Q-R----------GDDQE----
---------AFRTALGELRLL---Q-L-S------M--------------ES------------------------------WK--LL-ST--RV---------QAKL-----DD---
-------------------------------------------------------------------REV-ARF------------------------------S-S-----------
A-------------------LRVYATKDRVNEYNHYHLD-RL----G---R--------------------------P-----------------------V-V-Q-V----------
----------K---------------------AKN------VGP--------------------G-----------------------------------------------------
--------------------------------------------------------------AAA-------------AP--------------------DD-----------K-A--
------GNL----------AKQI-----PICIGARLMLTSNLW----------Q----P--V-----G-----LCNGARGTVYDI

>Hel2_M_ani

-----SLNRNQRL-VYDTV------MDH----F-LTK-------------VS------SQ-----LLLHVDGGGGTGKSYLINLLSAH---L-Q--A--AAAG---------------
----RG-TPVWRAAPTGVAGNQI-------SG------TTLHSL-LHL---P--I------NK-D--------------FKP---LL------------------PTDM---------
------A---Q-------------LQKKL--KD--IKYLIIDEKSM--LGLRQLSW-IDDRLREAFPN--------------K---------N-----------------------E-
--------------------------EFFGG-----LSILLVGDFFQLPPVL-----------------QK-----------------------------------------------
--------------------------------------------PLYYD-KEVQGV-EI----KGR-NAYRR--------FDKSVFLKVVQR----Q-R----------GDDQK----
---------AFRTALGELRLL---Q-L-S------V--------------ES------------------------------WK--LL-SG--RV---------QAKL-----DD---
-------------------------------------------------------------------QEV-ARF------------------------------A-N-----------
A-------------------LRVYATKDRVNEYNHYHLD-RL----S---R--------------------------P-----------------------V-I-Q-V----------
----------K---------------------AKN------VGL--------------------G-----------------------------------------------------
--------------------------------------------------------------AAA-------------AP--------------------DD-----------K-A--
------GNL----------AKQI-----PICIGSRLMLTSNLW----------Q----P--V-----G-----LCNGARGTVYDI

>Hel2_P_chl
```

```
-----SLNRDQRL-VYDTV------MDH----F-LNQ-------------EP------SQ-----LLLHVDGGGGTGKSYLINLLSAH---L-Q--A--AAGG---------------
----RG-TPVWRAAATGVAGNQI-------SG------TTLHSL-LHL---P--I------NK-D--------------FKP---LS------------------AIDK---------
------A---Q-------------LQKKL--KD--IKYLIIDEKSM--LGLRQLSW-VDDRLREAFPS--------------R---------N-----------------------D-
--------------------------EFFGG-----LIILLVGDFFQLPPVL-----------------QK-----------------------------------------------
--------------------------------------------PLYYD-KEVQGV-EI----KGR-NAYRR--------FDKSVFLKVVQR----Q-R----------GDDQK----
---------AFRTALGELRLL---Q-L-S------A--------------ES------------------------------WK--LL-SS--RV---------QAEL-----DD---
-------------------------------------------------------------------QEV-ARF------------------------------A-K-----------
A-------------------LRVYATKDRVNEYNHYHLD-RL----S---R--------------------------P-----------------------V-I-Q-V----------
----------K---------------------AKN------VGF--------------------G-----------------------------------------------------
--------------------------------------------------------------AAA-------------AA--------------------DD-----------K-A--
------GNL----------AKQI-----PICIGARLMLTCNLW----------Q----E--V-----G-----LCNGARGTVYDI
>Hel2_P_lil
-----TLNPEQRI-VYDTI------LGH----F-QCG-------------SE------EQ-----ILLHVDGGGGTGKSYLIKVLSSH---L-Q--R--FAGN---------------
----RP-SPIWRAAPTGVASNQI-------TG------TTLHSL-LRL---P--V------DR-A--------------FTE---LS------------------PADT---------
------N---A-------------LQKKL--RD--VRYLVIDEKSM--LGLRQLSW-VDKRLRQVRPS--------------R---------A-----------------------A-
--------------------------EFFGG-----ISIILVGDFFQLPPIA-----------------NK-----------------------------------------------
--------------------------------------------PLYFD-GPLKDLHEI----SGQ-TAYRA--------FNHTVFLKKAQR----Q-Q----------GDDQA----
---------GFRLALEELRGL---K-L-S------I--------------ES------------------------------WK--LL-SL--RV---------QAKL-----SQ---
-------------------------------------------------------------------REV-DSF------------------------------D-A-----------
A-------------------LRIYSKKARVNEYNYEHLV-RL----K---H--------------------------P-----------------------A-I-Q-V----------
----------M---------------------ARN------IGN--------------------G-----------------------------------------------------
--------------------------------------------------------------ADK-------------AT--------------------SE-----------Q-A--
------GNL----------AGQF-----PLCIGARLMLTQNIW----------H----P--T-----G-----LVNGAQGTVYDI
>Hel2_F_mon
-----QLERQQRR-LYDFV------VAD----Y-AGEL---------AGLPP------PQ-----FLLNLDGKAGTGKSFVIMLISAT---L-Q--Q--MATNAG-------------
----RQ-FPILRAAPTGVAAHGI-------SG------RTLHAL-LRL---P--I------KF-P---K---S------YEK---LS------------------QQNL---------
------Q---A-------------AQSTM--RE--IRYLIIDEKSM--IGLKMMSW-MDQRLREIY----------------P---------T----------------------RD-
--------------------------LPFGG-----INIIIAGDFCQLPPVA-----------------MK-----------------------------------------------
--------------------------------------------PLFFQ-GQLVDPTEV----AGR-TLYNL--------FDKTIELNVIKR----Q-D----------GQTTE----
-------AIAFREALNALRED---R-V-T------V--------------DD------------------------------WG--LL-TT--RV---------AGII----------
------------------------------------------------------------------PHEI-PTF------------------------------D-D-----------
A-------------------IHIYGKKQQVNEVNHARMR-DL----Q---Q--------------------------P-----------------------V-L-K-I----------
----------M---------------------ATH------E----------------------G-----------L--------------------------------------K--
--------------------------------------------------------------ANE-------------AS--------------------SD-----------A-A--
------GNL----------HAEL-----PLALGTRIMLTENIW----------V----E--R-----G-----LVNGALGTVRDI
>CAB1116976.1_Ectocarpus_sp._CCAP_1310/34
-----SLNERQRN-CYDVV------RDH----F-ENE-------------RE------EP-----LRMMVLGTAGTGKSYLVYALSRL---L--------------------------
-----G-GFLRRAAPTGMAAFLI-------AG------STLHSL-LRL---P--V------RQ-G---------------RN---LQ------------------GQSL---------
------K---A-------------LQNSL--TG--VKYIIIDELSM--VSQSQMAW-VDRRLRQGT----------------A---------V-----------------------D-
--------------------------KPFGG-----ISLIMTGDLGQLQPVG-----------------GT-----------------------------------------------
--------------------------------------------PLYKQ-NPAAALNV-----EGY-AAYSL--------FQDVFILDRVQR----Q-T-AAAA---NDDD-QR----
---------GFIELLPRARDG---Q-L-C------D--------------ED------------------------------WD--LL-LK--RQ---------PNRL----------
--------------------------------------------------------------------TA-AEK------------------AAF---------E-D-----------
A-------------------TRLFYSKKEVNKYNGKKLR-EL----D---N--------------------------P-----------------------V-A-R-V----------
----------S---------------------AVH------T----------------------G--------------------A-----------------------------N--
--------------------------------------------------------------ARR-------------AT--------------------AD-----------T-A--
------EGL----------ERDL-----YLAKGAKVMLSKNLY----------Q----Q--V-----G-----LVNGIRGEVVEL
>RZC87713.1_Papaver_somniferum
-----ALSRQQQV-ALNLV------LES----L-RSE---------------------ST-----IRLIISGGAGTGKSTLISAIVHS---T-R--E--LFGN---------------
-----E-KSVRIMAPTGVAVFNI-------GG------STIHHE-LAI---T--A------DK-N---L---S------YKK---LE------------------AERC---------
------R---R-------------MQVDF--KD--TKLIIIDEYSM--IGRKMLAN-IDLRLRDIF----------------S---------T-----------------------S-
--------------------------EPFGN-----ISIVLVGDMRQLPPVF-----------------DT-----------------------------------------------
--------------------------------------------PLYAE--GGGELQLT------GTLSYSV--------FKQCVRLEQVFR----Q-S-GV-------EE--S----
---------EYREALSRLSDG---N-S-T------L--------------ED------------------------------WK--LF-FT--RS---------YAPL-----SV---
-------------------------------------------------------------------QEK-DNF------------------------------K-N-----------
V-------------------VRLFPTKEDAANHNCQRLG-QL----R---C--------------------------P-----------------------V-A-R-I----------
----------P---------------------SKN------NCV-------------------TA--------------------N--------------------------------
----------------------------------------------------------------E-------------AN--------------------SD-----------E-A--
------KGL----------EDVL-----LLSKQSRVMLRKNYS----------T----Q--F-----G-----LVNGSIGTVKDI
>Blastocystis_sp
-----SLSAEQKS-VLETV------L--------------------------------KG-----YNVFFTGDAGTGKSHILRVMIEA---L-Q--E--QLG----------------
-----K-DKVFVTASTGIAACNI-------GG------ITIHSF-AGL---G--I------TN-M---D---V-------NQ---T--------------L-----RKV--R--Q---
-N----E---A-------------AVERW--KA--CQVLIIDEISM--LDGRLFDM-LEYVGRTVR----------------N---------D-----------------------S-
--------------------------TPFGG-----IQIVACGDFFQLPPVG-----------------LG---Q--H----------------------------------------
-------------------------GV-----I-----------FSFES------------------RWWNV----VI--E-KVVVLKTVFR----Q-K-----------D--M----
---------RLQRLLREVRYG---R-V-S------Q--------------QS------------------------------VH--VI-ES--MA---------SHDL-----
EQVIAR-----------------------------QNC-------------------------------DEESDE-EF---------------------H---------V-E------
-----S-------------------TKLFALNNDVNRYNQQKLD-AL----D---S--------------------------P-----------------------A-V-D-Y-----
---------------A---------------------SID------N----------------------G-----------V--------E---------------------------
E-S-------Y--------------------------------------------------------L-Y--------------Q--------------------LG-----------
K-S--------CQA----------PARL-----TLKLGAQVMLVKNLS----------V----S--D-----G-----LVNGCRGVVVSF
>Blastocystis_sp2
-----SLSEDQRK-VYTAA------V--------------------------------EG-----YSIFFTGDAGTGKSYVLRLIVSA---L-K--K--KYG----------------
-----A-NRVFVTASTGIAACNI-------GG------TTLHSF-ASI---G--L------GD-E---S---I-------TK---C--------------V-----HRV--L--Q---
-N----K---K-------------AKKRW--QD--CVVLIIDEISM--LDGCFFDK-LEAVSRRIR----------------G---------D-----------------------E-
--------------------------SCFGG-----IQIIACGDFFQLPPVG-----------------LG---K--N----------------------------------------
-------------------------KV-----I-----------YCFES------------------ECWNT----VI--Q-RTLIMTKVFR----Q-K-----------D--E----
---------EFQALLRDIRYG---K-V-S------Q--------------RS------------------------------RT--LL-QR--LE---------RNEL----------
```

```
----------------------------------------------------------------------N-TG----------------------K---------I-V-----------
P-------------------TKLFALNESVDQYNTSALA-QL----P---D---------------------------TH----------------------C-I-T-Y-----------
----------K---------------------AID------E-----------------------G--------------------Q----------------------------D-V--
-----Y---------------------------------------------------------L-K---------------Q-------------------LQ------------K-N--
------CQA----------PAVL-----PLKVGAQVMLLKNLS----------V----E--M-----G-----LVNGSRGVVDSF
>D_discoideum
-----LLTSEQQK-IVNLI------VD--------------------------------GG-----KNVFFTGSAGTGKSFVLKHLVSK---L-R--K--KYP-----------------
-------KSVYVTAATGIAAVNI-------GG------TTLHSF-AGI----K--L-------GV-A---P---A-------QR---L--------------A-----VEI--L--Q---
-S----K---K-------------LLQKW--LD--CSVLIIDEISM--IDAELFEK-LDTIGQMVR-----------------G---------N------------------------N-
--------------------------QPFGG-----IQLVLVGDFFQLPPVH------------------GN------------------------------------------------
--------------------------------------------YAFEC-------------------KAWKK----SI--D-ISVELTTVMR----Q-K------------E--T----
---------EFIDILNKIRVG---D-I-K------E--------------DM--------------------------------IN--RLVST--CN---------KPLD----------
----------------------------------------------------------------------I-SN----------------------G---------I-L-----------
P-------------------TRLYSTNASVDQENQSSLD-KL----L---G---------------------------E-----------------------P-F-S-F-----------
----------Q---------------------AVD------S-----------------------G--------------------N----------------------------K-E--
-----L---------------------------------------------------------I-E---------------L-------------------LD------------R-D--
------CPA----------MKNL-----TLKVGAQVVLLRKIE----------K----G--D-----G-----LVNGSRGVVVDF
>KAF2073656.1_Polysphondylium_violaceum
-----KLTKEQQK-IVNLI------VE--------------------------------GG-----KNVFFTGSAGTGKSFVLKHLVSK---L-R--K--KHP-----------------
-------KSVFVTAATGIAAVNI-------GG------TTLHSF-GGI----K--L-------GV-A---P---A-------QR---L--------------A-----VEI--L--Q---
-S----K---K-------------ALQKW--LD--CRVLIVDEVSM--IDSELFEK-LDTVAQIVR-----------------E---------N------------------------N-
--------------------------QPFGG-----IQLVLVGDFFQLPPVY------------------GN------------------------------------------------
--------------------------------------------YAFES-------------------KAWKK----SI--D-ICLELTTVMR----Q-R------------D--L----
---------EFIDVLNNLRVG---E-K-N------D--------------KI--------------------------------VN--FLLDR--CK---------RPLD----------
----------------------------------------------------------------------V-TN----------------------G---------V-L-----------
P-------------------TKLYSTNASVDEENSAALE-QL----A---S---------------------------E-----------------------P-H-S-F-----------
----------L---------------------AYD------T-----------------------G--------------------S----------------------------K-E--
-----L---------------------------------------------------------L-E---------------N-------------------LD------------R-D--
------CPA----------MQKL-----TLKVGAQVVLLRKLE----------K----H--D-----G-----LVNGSRGVVIDF
>KYQ93685.1_Tieghemostelium_lacteum
-----PLTSEQEK-IVNLI------VE--------------------------------GG-----KNVFFTGSAGTGKSFVLKHLVAR---L-R--E--KFP-----------------
-------KSVFVTAATGIAAVNI-------GG------TTLHSF-AGI----H--L-------GT-A---T---A-------EK---L--------------A-----ANI--I--K---
-K----K---K-------------YLQRW--RD--VKVLVIDEISM--IDSELFEK-LNTIGKIIR-----------------G---------N------------------------Q-
--------------------------LPFGG-----IQLVLVGDFFQLPPVL------------------GS------------------------------------------------
--------------------------------------------YTFES-------------------PQWES----CI--D-MCLELTTVMR----Q-K------------E--I----
---------EFINVLNSIRVG---R-V-H------D--------------GI--------------------------------VK--SL-QQ--CA---------RPLD----------
----------------------------------------------------------------------V-SN----------------------G---------V-L-----------
P-------------------TKLYTTNQSVDDENTLALG-AL----T---G---------------------------E-----------------------P-K-V-Y-----------
----------E---------------------SFD------S-----------------------G--------------------H----------------------------K-D--
-----M---------------------------------------------------------I-V---------------N-------------------LD------------N-D--
------CPA----------PKSL-----TLKVGAQVVLLRKLE----------K----N--D-----T-----LVNGSRGVVVDF
>XP_012754920.1_Acytostelium_subglobosum_LB1
-----TLTPEQER-VVNLI------V---------------------------------DG----KKNVFFTGSAGTGKSFVLKHIVAR---L-R--E--KHE-----------------
-------KAVYVTAATGIAAVNI-------GG------VTLHSF-AGI----K--M-------GH-G---T---P-------EQ---L--------------V-----SKI--L--K---
-S----R---I-------------YTKRW--TE--AKVLVIDEISM--VDAELFEK-LDVIARTLK-----------------V---------N------------------------D-
--------------------------KPFGG-----IQLVLCGDFFQLPPVV------------------GS------------------------------------------------
--------------------------------------------YAFEC-------------------EAWKR----CV--D-ECVQLTTVMR----Q-K------------E--G----
---------VFVKVLNNLRRG---Y-V-T------P--------------EA--------------------------------IK--VL-QD--CD---------RPLD----------
----------------------------------------------------------------------I-SN----------------------G---------V-L-----------
P-------------------TKLYSTNQHVDDENTKALE-AL----E---G---------------------------E-----------------------P-T-T-F-----------
----------T---------------------SID------S-----------------------G--------------------S----------------------------E-E--
-----L---------------------------------------------------------K-D---------------N-------------------IE------------R-D--
------CPA----------PQQL-----TLKVGAQVVLLRKLD--------G-A----K--S-----H-----LVNGSRGVVVDF
>A_subglobosum
-----TLSPEQRH-VVDLV------L---------------------------------GG-----SSIFFTGSAGTGKTYVLREIIQA---L-R--F--LHG-----------------
-------DCVHVTASTGIAACNV-------GG------TTLHSF-AAI----G--L-------GD-K---P---A-------KD---Y--------------I-----RSI--A--G---
-N----N---K-------------NLLRW--RQ--TKVLVIDEISM--ISAELLDK-LDQIGRALR-----------------S---------S------------------------P-
--------------------------RPFGG-----IQLVLVGDFCQLPPVS------------------KQ---T--------------------------------------------
-------------------------AA-----S-----------YCFRA-------------------VCWEM----MI--D-HSILLTKVYR----Q-K------------D--D----
---------KFVKVLNELRFG---V-I-S------D--------------VG--------------------------------LT--TL-NQ--CV---------SNNL-----D----
----------------------------------------------------------------------A-TD----------------------G---------I-I-----------
P-------------------TVLYPHRAKCDAENERKLQ-AL----A---G---------------------------E-----------------------A-M-V-F-----------
----------E---------------------AED------E-----------------------G--------------------P----------------------------D-P--
-----Y-----------------------------------------------------------R---------------D-------------------ML------------K-N--
------MQA----------QSTV-----TLKIGAQVILLKNLD----------F----E--E-----E-----LVNGSRGVVVAW
>H_album
-----LLSKEQRN-VVQLA------L---------------------------------DG-----NSIFFTGSAGTGKSFVLREIVNV---L-R--V--LHG-----------------
-------DNVHVTASTGIAACNI-------GG------TTLHSF-AGI----A--L-------GE-K---T---A-------LD---Y--------------I-----RSI--A--N---
-N----N---K-------------NLTRW--RQ--TKVLIIDEVSM--ISCELLDK-LDLIGQGLR-----------------K---------I------------------------P-
--------------------------KPFGG-----IQVILVGDFCQLPPVN------------------KN---R--D-NNA-------------------------------------
-------------------------AS-----S-----------FCFNA-------------------KCWKY----LI--D-HSILLTKVYR----Q-K------------D--N----
---------HFVNILNQLRFG---T-I-D------E--------------AG--------------------------------MT--TL-NK--CV---------NNII-----E----
----------------------------------------------------------------------S-ED----------------------G---------I-I-----------
P-------------------TILYPHRNKVELENERKLK-EL----K---S---------------------------D-----------------------E-M-I-F-----------
----------D---------------------AID------E-----------------------G--------------------P----------------------------D-Q--
-----Y-----------------------------------------------------------R---------------D-------------------ML------------K-N--
------MQA----------QTRL-----TLKIGSQVILLKNLD----------F----S--S-----E-----LVNGSRGVVVGF
>C_fasciculata
```

```
-----TLSEEQRF-VLDHV------L--------------------------------KG-----NNIFLTGGGGTGKSYLLRIMISC---L-R--K--KFK----------------
-----V-NELYATASTGVAAVNI-------GG------TTVHSF-GGI---G--L------GT-K---P---A-------ET---L--------------Y-----YQI--C--N---
-N----M---K-------------ALKRW--TS--CKCLVIDEISM--ISKTIFDL-LDYLAKRIR----------------S---------N-----------------------Q-
--------------------------EPFGG-----IQLIVVGDFSQLPPVV-----------------ND---R--K-ALQ------------------------------------
---PGEKKQ--------------YMHQ-----Q-----------FCFLS------------------PAFKQ----LF-TQ-NSFNLTQVYR----Q-S-----------D--T----
---------TFIDILNRIRFG---I-V-K------D--------------ED------------------------------IQ--LI-ND--RT---------SRPL-----S----
---------------------------------------------------------------------I-TD---------------------N---------I-I-----------
P-------------------TILYPKKVNVHEENIKQLA-LI----N---E--------------------------K-----------------------E-Y-T-F----------
----------K---------------------ADD------Y----------------------G--------------------D---------------------------P-Q--
----------------------------------------------------------------H--------------M--------------------GF-----------T-N--
------CQA----------PEIV-----KLKTGAQVLLMVNQD----------F----K--K-----K-----LVNGSRGVIVGW
>XP_014228054.1_Trichogramma_pretiosum
-----EMTAEQSQ-VLEYV------L--------------------------------NG-----KSIFFTGSAGTGKSFLLRKIIAA---L-P--P---------------------
-------DVTIATASTGVAACHI-------GG------ITLHQF-AGI---G--L------GT-G---T---M-------EK---C--------------K-----KMV-----A---
-K----S---A-------------AGTIW--RK--TKHLIIDEISM--VDGDYFDK-IEAIARFVR----------------N---------S-----------------------E-
--------------------------KPFGG-----IQLILCGDFFQLPPVS-----------------KR---D--E----------------------------------------
-------------------------QS-----K-----------FCFQS------------------KAWAS----CI--Q-MNFELKKVHR----Q-T-----------D--P----
---------QFISILNQLRMG---Q-V-T------D--------------ET------------------------------TK--IL-QE--TS---------RQTI----------
---------------------------------------------------------------------E-TK---------------------G---------I-L-----------
A-------------------TRLCSHVNEANEINETQLE-KL----S---G--------------------------V-----------------------S-K-T-Y----------
----------M---------------------AED------S----------------------D--------------------E-----------------------------S--
-----L--------------------------------------------------------T-R--------------S--------------------LD-----------Q-Q--
------LTV----------PNKL-----VLKVGAQVMLLKNIS----------L----S--A-----G-----LVNGARGVVVKF
>XP_017781591.1_Nicrophorus_vespilloides
-----PLTIEQRD-VLDAC------L--------------------------------SG-----QNLFFTGSAGTGKSYLLRKIIGA---L-P--P---------------------
-------DVTVATASTGVAACHI-------GG------TTLHQF-AGI---G--S------SD-G---S---L-------ER---A--------------K-----EVA-----N---
-R----P---P-------------TSSNW--RR--CKHLIIDEISM--IDGDYFEK-IEAVARHVR----------------K---------N-----------------------D-
--------------------------KPFGG-----IQLILCGDFLQLPPVV-----------------KT---K--E----------------------------------------
------------------------NKK-----R-----------FCFQT------------------KAWKE----CV--N-QTFELKQVHR----Q-S-----------D--C----
---------KFIDILNKLRIG---E-V-S------D--------------EV------------------------------VE--TL-AR--TS---------KQRI----------
---------------------------------------------------------------------E-KD---------------------G---------I-L-----------
A-------------------TRLCSHMADANMINESKIK-NL----P---G--------------------------E-----------------------A-K-L-Y----------
----------D---------------------AQD------S----------------------D--------------------N-----------------------------Y--
-----L--------------------------------------------------------T-K--------------Q--------------------LD-----------Q-Q--
------TPV----------PGKL-----QLKVDAQVMLLKNVN----------V----S--A-----G-----LVNGARGVVTGF
>NP_942102.1_Danio_rerio
-----KLSKEQTA-VLNAV------L--------------------------------SG-----KNVFFTGSAGTGKSFLLKRIVGS---L-P--P---------------------
-------KSTYATASTGVAACHI-------GG------TTLHSF-AGI---G--S------GS-A---P---L-------EQ---C--------------I-----ELA-----Q---
-R----P---G-------------VLRHW--TS--CKHLIIDEISM--VEAEFFDK-LEAIARSIR----------------R---------S-----------------------T-
--------------------------EPFGG-----IQLIVCGDFLQLPPVT-----------------KG---K--E----------------------------------------
-------------------------KA-----N-----------FCFQS------------------RSWRK----CI--H-MNMELMEVRR----Q-T-----------D--K----
---------TFISLLQAVRVG---R-V-T------E--------------EV------------------------------TA--QL-LK--SA---------NHCI----------
---------------------------------------------------------------------E-RD---------------------G---------I-L-----------
A-------------------TRLCTHKDDVELTNENKLK-QL----P---G--------------------------V-----------------------V-R-M-Y----------
----------E---------------------AVD------S----------------------D--------------------P-----------------------------M--
-----L--------------------------------------------------------V-Q--------------T--------------------ID-----------A-Q--
------SPV----------SRLL-----QLKVGAQVMLTKNLD----------V----Q--R-----G-----LVNGARGVVVDF
>H_sapiens
-----QLSEEQAA-VLRAV------L--------------------------------KG-----QSIFFTGSAGTGKSYLLKRILGS---L-P--P---------------------
-------TGTVATASTGVAACHI-------GG------TTLHAF-AGI---G--S------GQ-A---P---L-------AQ---C--------------V-----ALA-----Q---
-R----P---G-------------VRQGW--LN--CQRLVIDEISM--VEADLFDK-LEAVARAVR----------------Q---------Q-----------------------N-
--------------------------KPFGG-----IQLIICGDFLQLPPVT-----------------KG---S--Q----------------------------------------
-------------------------PP-----R-----------FCFQS------------------KSWKR----CV--P-VTLELTKVWR----Q-A-----------D--Q----
---------TFISLLQAVRLG---R-C-S------D--------------EV------------------------------TR--QL-QA--TA---------SHKV----------
---------------------------------------------------------------------G-RD---------------------G---------I-V-----------
A-------------------TRLCTHQDDVALTNERRLQ-EL----P---G--------------------------K-----------------------V-H-R-F----------
----------E---------------------AMD------S----------------------N--------------------P-----------------------------E--
-----L--------------------------------------------------------A-S--------------T--------------------LD-----------A-Q--
------CPV----------SQLL-----QLKLGAQVMLVKNLS----------V----S--R-----G-----LVNGARGVVVGF
>XP_034314500.1_Crassostrea_gigas
-----KLSKEQST-ILDAV------L--------------------------------KG-----KNVFFTGSAGTGKSFLMRRIIGS---L-P--P---------------------
-------QHTYATASTGVAACHI-------GG------TTLHAF-AGI---G--S------GS-A---P---L-------EQ---C--------------V-----QLA-----S---
-R----P---Q-------------IAQQW--RK--CHHLVIDEISM--VSGAFFDK-LETVARVVR----------------K---------N-----------------------D-
--------------------------NPFGG-----IQLIICGDFLQLPPVT-----------------KG---T--D----------------------------------------
-------------------------KK-----T-----------FCFQA------------------KSWSR----CV--Q-VNMELKEVRR----Q-N-----------D--L----
---------SFINILQNIRLG---K-C-S------E--------------ET------------------------------HQ--IL-RQ--TV---------HHEI----------
---------------------------------------------------------------------Q-KN---------------------G---------I-L-----------
A-------------------TRLCTHKEDVNKINQYHLG-KL----Q---G--------------------------E-----------------------E-R-T-F----------
----------V---------------------AVD------G----------------------E--------------------E-----------------------------A--
-----Y--------------------------------------------------------K-D--------------Q--------------------LE-----------V-L--
------SPV----------PKKV-----VLKVGAQVMLAKNLD----------V----Q--R-----G-----LVNGARGVVVGF
>XP_028415325.1_Dendronephthya_gigantea
-----KLNKSQLR-VLNAV------K--------------------------------CG-----QSVFITGSGGTGKSFLLRKIIGL---L-P--P---------------------
-------HNTFVTASTGVAACQI-------GG------MTLHSF-SGI---G--C------GK-G---N---L-------EN---C--------------I-----AMA-----S---
-N----R---I-------------HLQQW--KN--CKHLIIDEISM--IDSELFDK-IEAVARALR----------------K---------N-----------------------D-
--------------------------RPFGG-----IQLIVCGDFLQLPPVI-----------------KP---G--E----------------------------------------
-------------------------KK-----K-----------FCFQA------------------ESWST----CI--H-KTIELIEVKR----Q-S-----------D--P----
---------LFINILNNIRVG---R-C-P------D--------------EV------------------------------VE--RL-SR--SK---------ENKI----------
```

```
---------------------------------------------------------------------D-SE---------------------G---------I-L-----------
A-------------------TRLCTHKENVDQINKVQLQ-SL----P---G--------------------------K-----------------------A-K-S-F----------
----------Q---------------------AVD------S----------------------D--------------------N-----------------------------N--
-----F--------------------------------------------------------S-K--------------T--------------------LD-----------S-C--
------CPA----------KAKL-----ELKEGAQVLLTKNLD----------V----G--Q-----G-----LVNGARGVVKSF
>XP_031556309.1_Actinia_tenebrosa
-----NLTSEQSE-VIKAV------R--------------------------------AS-----RNVFFTGSAGTGKTFLLRKLLGI---L-P--P---------------------
-------EGTFVTASTGAAACHI-------GG------TTLHAF-AGI---G--S------GS-A---T---I-------EQ---C--------------I-----DLA-----S---
-R----P---E-------------RSRQW--KN--CKRLIIDEISM--IDGDLFDK-LEAVARSVR----------------N---------N-----------------------D-
--------------------------RPFGG-----IQLILSGDFLQLPPVW-----------------KK---E--G----------------------------------------
------------------------NNK-----K-----------FCFQA------------------ESWQD----CI--S-NTIELTSVFR----Q-K-----------D--P----
---------MFVSILQNIRVG---S-C-P------E--------------KL------------------------------VA--KL-VE--TR---------NHTI----------
---------------------------------------------------------------------E-KD---------------------G---------I-L-----------
A-------------------XKLCTHKENVDQINEIHMS-KI----S---G--------------------------K-----------------------V-H-T-F----------
----------A---------------------ACD------S----------------------D--------------------P-----------------------------G--
-----Q--------------------------------------------------------A-H--------------V--------------------LS-----------K-Q--
------LAA----------PDKI-----DLKVGAQVMLVKNLN----------V----S--E-----G-----LVNGARGVVKGF
>XP_003388034.1_Amphimedon_queenslandica
-----SLSEDQLQ-VINVI------K--------------------------------NG-----DSVFITGSAGTGKSYLLQRIIGM---L-P--P---------------------
-------DTTYCTASTGAAACII-------GG------TTLHSF-AGI---N--T------DA-A---P---L-------KQ---C--------------V-----SMA-----M---
-R----E---H-------------KAVHW--KR--CKVLLIDEISM--VDGEFFDK-LEAVARAVR----------------K---------S-----------------------K-
--------------------------KPFGG-----IQLVLCGDFLQLPPVC-----------------KD---G--K----------------------------------------
-------------------------KR-----L-----------FCFQA------------------ESWRK----CV--N-RTIELNDVYR----Q-K-----------D--R----
---------EFIAILQNIRIG---R-C-P------P--------------AI------------------------------TK--LL-KN--TE---------NQLI----------
---------------------------------------------------------------------E-KG---------------------G---------I-R-----------
A-------------------TKLYTHTNEVESTNQTELN-AL----A---G--------------------------E-----------------------G-R-R-F----------
----------D---------------------ATD------N----------------------Q--------------------P-----------------------------N--
-----C--------------------------------------------------------M-Q--------------Q--------------------LN-----------A-L--
------CLV----------PHTL-----VLKIGAQVMLAKNID----------V----S--R-----S-----LVNGARGIVTSF
>NP_001293174.1_Caenorhabditis_elegans
-----QLSDEQKS-VVRCV------IN-------------------------------SR-----TSVFFTGSAGTGKSVILRRIIEM---L-P--A---------------------
-------GNTYITAATGVAASQI-------GG------ITLHAF-CGF---R--Y------EN-S---T---P-------EQ---C--------------L-----KQV--L--R---
-Q----N---H-------------MVRQW--KQ--CSHLIIDEISM--IDRDFFEA-LEYVARTVR----------------N---------N-----------------------D-
--------------------------KPFGG-----IQLIITGDFFQLPPVS-----------------KD-----------------------------------------------
-------------------------EP-----V-----------FCFES------------------EAWSR----CI--Q-KTIVLKNVKR----Q-N-----------D--N----
---------VFVKILNNVRVG---K-C-D------F--------------KS------------------------------AD--IL-KE--SS---------KNQF----------
-----------------------------------------------------------------------PS---------------------S---------V-I-----------
P-------------------TKLCTHSDDADRINSSSIE-TT----Q---G--------------------------D-----------------------A-K-T-F----------
----------H---------------------AYD------D----------------------E--------------------S--------------------------------
-----F--------------------------------------------------------D-T--------------H--------------------AK-----------A-R--
------TLA----------QKKL-----VLKVGAQVMLIKNID----------V----I--K-----G-----LCNGSRGFVEKF
>XP_004991536.1_Salpingoeca_rosetta
-----SLTPEQKD-VLMAV------L--------------------------------SG-----RNVFFTGSAGTGKSYLIGKIIEA---L-P--K---------------------
-------ATTVVTASTGVAACAI-------GG------TTLHAF-AGV---Q--A------GS-S---R---L---------------------------------------------
------P---V-------------NTSAW--TT--AKVLLIDEVSM--IDAPYFDQ-LEQTARRVR----------------R---------C-----------------------N-
--------------------------KPFGG-----LQLVLVGDFLQLPPVT-----------------KR---G--E----------------------------------------
-------------------------ET-----Q-----------FCFQA------------------KSWDA----CV--H-ECFHLSQVHR----Q-R-----------D--R----
---------TFVDILHRCRLG---Q-C-T------P--------------SD------------------------------IT--YI-QR--SA---------THRL----------
---------------------------------------------------------------------D-SS---------------------H---------I-R-----------
A-------------------TRLCTHVKEAKQINEQQLS-KL----S---G--------------------------S-----------------------S-K-L-F----------
----------T---------------------RSD------A----------------------S--------------------P-----------------------------D--
-----V--------------------------------------------------------S-R--------------S---------------------S-----------L-A--
------SRV----------EKVL-----ELKVGAQVMLSANVN----------V----S--A-----G-----LANGSRGVVVKF
>M_conductrix
-----DLSDEQQR-ALQLV------Q--------------------------------SG-----RSIFFTGCAGTGKSLLLRHILRC---L-P--R---------------------
-------NTTFVTGTTGLAACHL-------GG------TTINSY-AGI---G--R------GE-G---S---L-------ES---L--------------V-----RMA-----G---
-R----G---E-------------SLQRW--RA--TTHLIVDEVSM--MDGRLFDT-LEAVARKVR----------------G---------S-----------------------A-
--------------------------APFGG-----IQLILSGDFHQLPPVA-----------------KG---R--E-GAA------------------------------------
-------------------------QR-----K-----------FCFEA------------------ESWAR----CI--P-ESCFLSKVFR----Q-S-----------D--N----
---------EFVDLLGKIRSG---S-C-P------Q--------------DK------------------------------VS--QLLKT--CA---------RPLP----------
---------------------------------------------------------------------T-DD---------------------G---------I-L-----------
P-------------------TKLFTHREDVDLINAQQLK-AL----P---S--------------------------E-----------------------P-H-K-F----------
----------V---------------------AQD------V----------------------G--------------------S---------------------------G----
----------------------------------------------------------------E--------------V--------------------LA-----------A-A--
------CPA----------RRTL-----ELKVGAQVTLIKNIS----------Q----R--Q-----G-----LVNGARGVVEKF
>Helicosporidium_sp
-----PLSTEQRR-ALEAV------A--------------------------------SG-----RSLFFTGCAGTGKSHLLRAVLDS---L-P--A---------------------
-------HGTHVTGTTGLAASAL-------GG------CTLASW-AGT---G--R------LD-H---G---A-----SFAE---L--------------L-----AAA-----S---
-R----G---E-------------AARRW--LA--VRTLVVDEVSM--LDGRWFDA-LERLAREIR----------------R---------D-----------------------S-
--------------------------RPWGG-----VQLVLSGDFHQLPPVS-----------------RD---G-------------------------------------------
-------------------------SR-----V-----------YCFEA------------------STWGR----VI--K-EQLTMTQVFR----QGE-----------D--L----
---------DFVHLLADVRRG---V-C-T------G--------------EG------------------------------VR--AL-RL--RC---------
RSLENHGPGEEEERKQDQ--AGVERKQDQAGVERKEDRAGATKTEDQ-IDVIFKNDPAQP-------------------PFSLA-SA---------------------A---------
I-V-----------S-------------------TKLMTHRQQVAEDNARQLA-AL----P---F--------------------------P-----------------------S-
R-V-F--------------------Q---------------------AED------E----------------------G--------------------D-----------------
-----------------------------------------------------------------------------V-S--------------L--------------------VR-
----------G-A--------CPA----------ESRL-----ELKLGAQVILVRTVC----------A----A--R-----G-----LVNGARGVVVGF
>XP_023909855.1_Quercus_suber
```

```
-----FLSEEQQH-VLDLV------LE-------------------------------KN-----SSVFFTGSAGTGKSVLMREIIAA---L-R--K--KYQ----------------
---REP-DRVAVTASTGLAACNV-------GG------VTLHSF-AGI---G--L------GK-E---D---V-------PE---L--------------V-----RKI--K--R---
-N----Q---K-------------SKQRW--MR--TKVLVVDEVSM--VDGELFDK-LEAIARQLR----------------N---------N-----------------------G-
--------------------------RPFGG-----IQLVVTGDFFQLPPVP-----------------DK---G-------------------------------------------
------------------------KVA-----K-----------FAFDA------------------ATWTT----TI--E-HTIGLHHVYR----Q-K-----------D--P----
---------IFAGMLNEMREG---R-L-S------E--------------SS------------------------------IK--AF-RC--LS---------RKPE----------
---------------------------------------------------------------------F-ED---------------------D---------M-D-----------
A-------------------TELFPTRNEVDRANNERLW-KL----Q---G--------------------------V-----------------------D-V-V-F----------
----------E---------------------ARD------G----------------------G--------------------S-------VVE-----------------K-D--
-----R--------------------------------------------------------R-D--------------K--------------------LL-----------S-N--
------CMA----------PERI-----VLKKGAQVMLIKNVD------------------D-----S-----LVNGSPGRVLGF
```

```
>P_grisea1
-----TLSNEQRH-VKDLV------CS-------------------------------RS-----QSVFFTGPAGTGKSVLMRAIIED---L-K--K--KWK----------------
---KDP-DRLAVTASTGLAACNI-------GG------MTLHSF-AGI---G--L------GK-E---D---V-------TT---L--------------V-----KKI--R--R---
-N----P---K-------------AKNRW--LR--TKVLIIDEISM--VDGDLFDK-LSQIGRIIR----------------N---------H-----------------------G-
--------------------------KAWGG-----IQLVITGDFFQLPPVP-----------------DG---S--D-K--------------------------------------
------------------------RDI-----K-----------FAFEA------------------ATWNT----SI--D-HTIGLTEVFR----Q-K-----------D--P----
---------AFANMLNEMRLG---K-I-S------E--------------KT------------------------------VA--NF-KS--LE---------RELR----------
---------------------------------------------------------------------F-DD---------------------G---------L-E-----------
V-------------------TELFPTRSEVERSNNLRLA-AL----K---S--------------------------K-----------------------T-Y-R-Y----------
----------D---------------------AQD------S----------------------G------------------------------D-----------------P-N--
-----F--------------------------------------------------------R-D--------------K--------------------LL-----------Q-N--
------MMA----------PQKL-----ELRKGAQVMLIKNMD------------------E-----T-----LVNGSLGTVVGF
```

```
>XP_009351018.1_Pyrus_x_bretschneideri
-----LLSHEQRH-ILQLV------E--------------------------------EG-----HSIFYTGSAGTGKSVLLREIIKT---L-R--R--KYS----------------
---RSL-DAIAVTASTGIAACNI-------GG------VTIHSF-AGI---G--L------GR-E---T---A-------EQ---L--------------A-----IKV--H--K---
-N----K---K-------------ATTRW--LR--TQVLIIDEVSM--VEGDLFDK-LARIGSLIR----------------K---------K-----------------------V-
--------------------------EPFGG-----IQVIVTGDFFQLPPVA-----------------RD---T-------------------------------------------
-------------------------AV-----K-----------FAFEG------------------EMWSQ----TI--K-KTFNLTKVFR----Q-K-----------D--P----
---------EFVDILNEMRFG---R-L-T------Q--------------KS------------------------------ID--KF-KS--LS---------REII----------
---------------------------------------------------------------------Y-ED---------------------G---------L-G-----------
A-------------------TELFPRREDVERSNTVRMS-GI----E---G--------------------------T-----------------------V-H-L-F----------
----------Q---------------------AVD------G----------------------G--------------------M-------ITD-----------------K-E--
-----Q--------------------------------------------------------R-N--------------K--------------------LL-----------S-N--
------FMA----------PETL-----KLKIGAQVMLIKNLD------------------E-----T-----LVNGSIGMVVAF
```

```
>S_cerevisiae
-----CLSKEQES-IIKLA------E--------------------------------NG-----HNIFYTGSAGTGKSILLREMIKV---L-K--G--IYG----------------
-----R-ENVAVTASTGLAACNI-------GG------ITIHSF-AGI---G--L------GK-G---D---A-------DK---L--------------Y-----KKV--R--R---
-S----R---K-------------HLRRW--EN--IGALVVDEISM--LDAELLDK-LDFIARKIR----------------K---------N-----------------------H-
--------------------------QPFGG-----IQLIFCGDFFQLPPVS-----------------KD---P--N-R--------------------------------------
-------------------------PT-----K-----------FAFES------------------KAWKE----GV--K-MTIMLQKVFR----Q-R----------GD--V----
---------KFIDMLNRMRLG---N-I-D------D--------------ET------------------------------ER--EF-KK--LS---------RPLP----------
-----------------------------------------------------------------------DD---------------------E---------I-I-----------
P-------------------AELYSTRMEVERANNSRLS-KL----P---G--------------------------Q-----------------------V-H-I-F----------
----------N---------------------AID------G----------------------G--------------------A-------LED-----------------E-E--
-----L--------------------------------------------------------K-E--------------R--------------------LL-----------Q-N--
------FLA----------PKEL-----HLKVGAQVMMVKNLD------------------A-----T-----LVNGSLGKVIEF
```

```
>T_phaffii
-----TLSEEQKT-IIKLA------K--------------------------------DG-----HNIFYTGSAGTGKSVLLRELIKV---L-K--S--QHG----------------
-----S-DSVAVTASTGLAACNI-------GG------TTVHSF-AGI---G--L------GK-E---D---A-------ER---L--------------V-----SKV--Y--K---
-S----I---R-------------HRERW--KN--IKILVIDEISM--IDSSLLDK-LDYIAKKLR----------------K---------N-----------------------N-
--------------------------FPFGG-----IQLIFCGDFFQLPPVK-----------------KT---N--D-P--------------------------------------
-------------------------TV-----K-----------KAFES------------------DLWNN----AF--N-ITVKLENVFR----Q-K----------GD--L----
---------EFISMLEKARLG---K-I-D------D--------------ET------------------------------EK--QF-KQ--LD---------RMLD----------
-----------------------------------------------------------------------ND---------------------D---------I-A-----------
P-------------------AQLFPTRKEVEIANISQLR-IL----K---G--------------------------D-----------------------I-Y-A-Y----------
----------T---------------------SID------G----------------------G--------------------S-------IKD-----------------P-K--
-----M--------------------------------------------------------R-Q--------------N--------------------LL-----------E-N--
------FMA----------PKVL-----PLKVGAQVMMIKNVD------------------S-----T-----LVNGSLGKIVAF
```

```
>C_viswanathii
-----ILSKEQEY-ILKRV------M--------------------------------HG-----VSLFYTGSAGTGKSVLLRSIIKS---L-R--E--KYD----------------
-------RGIAVTASTGLAACNI-------GG------ITLHSF-AGI---G--L------GQ-G---T---V-------ES---L--------------L-----RKV--R--R---
-N----R---T-------------ALRRW--QE--TRVLIIDEISM--VDGNLLDK-LNELAKRIR----------------H---------N-----------------------T-
--------------------------SPFGG-----IQLVACGDFYQLPPVV-----------------KN---M--D-P--------------------------------------
------NEK--------------KVEP-----Y-----------FSFEC------------------KAWEE----AI--K-QTLTLKEIFR----Q-K----------GD--Q----
---------PFIDMLNEIRDG---R-I-S------L--------------GT------------------------------IN--KF-RS--LE---------RRLK----------
---------------------------------------------------------------------C-PE---------------------G---------F-V-----------
P-------------------SELYATRNEVERANNRKLN-SM----E---G--------------------------E-----------------------I-V-T-Y----------
----------T---------------------ARD------G----------------------G--------------------T-------LEK-------------------K--
-----R--------------------------------------------------------I-E--------------M--------------------LV-----------S-N--
------FLA----------PKKL-----QLKIGAQVMCIKNYD------------------E-----R-----LVNGSLGKVVAF
```

```
>XP_002163633.2_Hydra_vulgaris
-----CLSAEQKK-VLDIV------K--------------------------------SG-----RNVFITGSAGVGKSFLLNELIKS---Q-T--K---------------------
-------KGVYVTASTGVAACNI-------NG------TTLHSF-AGI---G--L------GN-K---P---A-------SI---L--------------A-----FDI--L--K---
-K----PYKVE-------------AKKRW--LG--CRILVIDEISM--IDAGLFST-VEEVARIVR----------------N---------N-----------------------N-
--------------------------SPFGG-----IQVILCGDFLQLPPVN-----------------VK-----------------------------------------------
--------------------------------K-----------FAFET------------------QAWRD----VV--H-ETVVLKKVFR----Q-K-----------L--V----
---------GFVSLLNRLRIG---Y-L-T------P--------------LD------------------------------IE--VL-KH--CK---------GTAF----------
```

```
---------------------------------------------------------------------P-DD---------------------G---------I-K-----------
A-------------------TCLFPHKASCDKLNQAELS-KL----P---G--------------------------K-----------------------M-F-T-F----------
----------E---------------------AVD------W----------------------F--------------------K---------------------------N-SM-
-----A--------------------------------------------------------Q-E--------------Q--------------------LN-----------K-T--
------SRY----------FKVL-----NLKVGAQVMLLNNLS----------V----S--N-----G-----LVNGARGVVTKF

>KAF4753487.1_Perkinsus_olseni

-----RMTSEQKK-VVEAV------L--------------------------------GG-----KSVFFTGGAGTGKSFVLHRLIRL---L-K--P---------------------
-------EHTAVTSSTGLAASHL-------GG------QTIHSF-AGI---G--S------GG-R---D---A-------AA---L--------------A-----QKI--K--R---
-S----P---E-------------LLGRW--KR--VKTLIMDEISM--LDGRLFDK-LEQIARLVR----------------Q---------D-----------------------S-
--------------------------RPFGG-----VQLVLTGDFLQLPPVS-----------------QT---L--P-NGK------------------------------------
-------KE----------------EA-----S-----------FCFEA------------------KSWRK----CI--R-KTMVLKEIKR----Q-E----------GD--A----
---------TFTTMLNEIRRG---I-C-S------Q--------------ET------------------------------QD--VL-SL--VA---------KRKH-----S----
--------------------------------------------------------------------TL-TG---------------------G---------V-V-----------
A-------------------SQLLPTRREVDAINERELA-RL----S---T--------------------------P-----------------------P-T-T-F----------
----------T---------------------AVD------T----------------------V--------------------Y---------------------------D-S--
-----S--------------------------------------------------------S-L--------------N--------------------LD-----------V-M--
------CSA----------RPKV-----VLKVGAQVMLTKTFS----------P----Q--K-----R-----LVNGSRGIIVRF

>C_roenbergensis

-----PLTSEQRQ-VLRAI------G--------------------------------EG-----HSVFFTGAAGCGKSVLLRRIIAS---L-P--A---------------------
-------ASTAVTAPTGVAACNV-------GG------TTLHAF-SGA---G--T-----RPD-A---S---A-------SD---V--------------A-----ALV--R--R---
-S----P---E-------------TLARW--RR--TRVLIVDEVSM--LDGAALDM-LEEVARRVR----------------S---------D-----------------------P-
--------------------------RPFGG-----LQLVLAGDFLQLPPVS-----------------KG---G--A-A--------------------------------------
-------------------------RK-----P-----------YAFEA------------------ACWGK----CV--S-VEVELTRVFR----Q-A-----------D--R----
---------DFVDVLNAIRWG---V-V-T------P--------------AA------------------------------RA--AL-DA--RW---------GADV-----A----
-------------------------------------------------------------------IAG-AD---------------------GGP------AI-R-----------
P-------------------TLLFTHRADVDAVNEKELA-RL----R---G--------------------------D-----------------------E-V-V-L----------
----------R---------------------GDD------T----------------------A--------------------H---------------------------S-PG-
-----A--------------------------------------------------------R-R--------------A--------------------LE-----------S-A--
------CPA----------PASL-----RLRVGAQVMLVRNLD----------V----G--A-----G-----LVNGARGVVLGF

>KAG5544865.1_Rhododendron_griersonianum

-----KGTDEQSR-VLDAI------S--------------------------------SG-----KSVFITGSAGTGKTLFLQHIIKR---L-K--K--LHH----------------
-----P-SRVFVTASTGLAACAI-------KG------RTLHSF-AGI---G--L------GE-D---D---R-------QT---L--------------L-----LKV--I--S---
-N----R---R-------------AYRRW--TK--VGALVIDESSM--IDGEILDT-LEFIARTVR----------------G---------GEE-----------------RDSEN-
--------------------------KVWGG-----IQLVVSGDFFQLPPIV-----------------KR---E--N----------------------------------------
-------------------------RK-----E-----------FAFEA------------------DCWGS----SF--D-MQVELTRVFR----Q-S-----------E--A----
---------NLVKLLQNVRRG---E-V-D------R--------------ED------------------------------LD--LL-KK--CC---------TEA-----------
-----------------------------------------------------------------------EP---------------------D---------S-S-----------
A-------------------VQLYPRNQDVNRVNKKKME-DL----K---K--------------------------P-----------------------T-Y-I-Y----------
----------H---------------------AHD------S----------------------G--------------------E---------------------------D-P--
-----W----------------------------------------------------------L--------------G--------------------QL-----------N-Q--
------GIA----------PDEL-----PLCEGARVMLCKNLS------------------R-----T-----LVNGATGTVTKF

>XP_020415763.1_Prunus_persica

-----QWTDQQKQ-VMSAI------S--------------------------------EG-----KSVFITGSAGTGKTILVKHIIKQ---L-K--K--RHG----------------
-----P-SKVFVTAPTGVAACAI-------SG------QTLHSF-AGI---G--C------AM-A---D---R-------DT---L--------------L-----HRI--S--K---
-N----D---K-------------AYKRW--RK--AEALVLDESSM--VDAELFES-LDFIARAIK----------------Q---------V-----------------------D-
--------------------------EVWGG-----IQLVVSGDFFQLPPVK-----------------PQ---Q--N----------------------------------------
-------SG----------------GK-----E-----------FAFEA------------------ECWDS----SF--D-LQVNLTKVFR----Q-S-----------D--P----
---------QLIKLLQGIRRG---E-S-D------P--------------ED------------------------------LK--LL-EQ--SC---------SKA-----------
-----------------------------------------------------------------------EP---------------------D---------P-T-----------
V-------------------VQLYPRNEDVNRVNSSRLA-SL----G---N--------------------------E-----------------------L-V-V-Y----------
----------T---------------------AVD------S----------------------G--------------------E---------------------------D-S--
-----L----------------------------------------------------------K--------------R--------------------QL-----------E-Q--
------GIA----------PKEI-----ALCEDARVMLVKNLN----------T----W--R-----G-----LVNGATGTVTGF

>XP_016507676.1_Nicotiana_tabacum

-----KLTDQQNQ-ILEAI------S--------------------------------NG-----NSVFITGSAGTGKTYLLQDIITK---L-R--K--IHG----------------
-----K-SRVFVTASTGVAACSL-------NG------QTLHSF-AGI---G--L------GD-A---S---A-------VD---L--------------L-----SRV--T--L---
-D----K---R-------------AYRRW--NK--VRALVIDEISM--ISGEVFDN-LEFIARSIR----------------S---------DEV------------------GCED-
--------------------------KSWGG-----IQLVVSGDFFQLPPVI-----------------NK---K-------------------------------------------
-------GQ----------------NK-----E-----------FAFEA------------------ECWNA----SF--D-MQIELKTIFR----Q-S-----------D--A----
---------QLIKLLQGIRKG---K-Y-D------S--------------ED------------------------------LQ--LL-DQ--CC---------SEV-----------
-----------------------------------------------------------------------EP---------------------D---------A-S-----------
A-------------------VQLYPRIEDVSRVNADRLD-RL----D---E--------------------------V-----------------------L-Y-H-Y----------
----------Q---------------------ALD------S----------------------G--------------------K---------------------------D-P--
-----W----------------------------------------------------------K--------------K--------------------QL-----------K-N--
------GIA----------PELL-----KLCVGARVLLTKNID----------V----I--G-----G-----LVNGATGTILDF

>XP_021598660.1_Manihot_esculenta

-----NWTKEQND-VLNHV------R--------------------------------GG-----LSVFITGSAGTGKSVLLKTIINV---L-K--K--VHG----------------
-----S-SGVFVTASTGVAACAL-------NG------RTLHSF-AGF---G--I------RN-D---E---Y-------GT---L--------------L-----DRV--I--M---
-S----S---C-------------ACERW--RQ--VKALVIDEISV--ISANMFDN-LESIAREIR----------------G---------S-----------------------K-
--------------------------EIWGG-----IQLIVSGDFFQLSPVP-----------------DK---C--N----------------------------------------
-------SS----------------GK-----E-----------FAFEA------------------NCWDA----SF--D-MLVELTKVFR----Q-S-----------D--A----
---------GQIELLQRTRKG---I-I-Y------P--------------ED------------------------------MQ--IL-EQ--CC---------SSN-----------
-----------------------------------------------------------------------EP---------------------D---------S-S-----------
V-------------------VSFYPRNEDVNKVNEERIK-SL----G---E--------------------------K-----------------------V-V-V-Y----------
----------K---------------------AAD------G----------------------G--------------------V---------------------------D-N--
-----Q----------------------------------------------------------R--------------E--------------------EL-----------K-Q--
------GIA----------PDQL-----ELCKGARVMLIKNLN----------V----R--R-----N-----LCNGATGTVTGF

>XP_019426349.1_Lupinus_angustifolius
```

```
-----QWTEEQKS-VLSSV------S--------------------------------QG-----KSVFITGAAGTGKTKLVTEIVKL---L-N--K--LHT----------------
-----P-SKVFVTASTGVAAFSI-------KG------QTLHSF-AGI---R--Y------HT-Y---D---P-------KI---L--------------Y-----DSI--K--S---
-C----K---R-------------ACWRW--QE--VKALVIDEISM--VDARLFDN-LERVARELR----------------G---------V-----------------------G-
--------------------------EPWGG-----IQLVVVGDFCQLPPIP-----------------DD---H--S----------------------------------------
------------------------LGV-----K-----------YAFEA------------------DCWNE----SF--D-FMIELTKILR----Q-S-----------D--P----
---------RFIELLQGIRIG---K-S-N------P--------------ED------------------------------LS--FL-KS--YC---------SKT-----------
-----------------------------------------------------------------------KS---------------------D---------L-S-----------
A-------------------VQLFPRKQNVTKVNEERLK-SL----Q---K--------------------------S-----------------------V-V-V-Y----------
----------K---------------------AVD------D----------------------G--------------------A---------------------------K-A--
-----W----------------------------------------------------------M--------------S--------------------QL-----------N-H--
------GIA----------PDEV-----SICVGARVMLIKNLS----------T----W--K-----G-----LVNGATGTVVEL
>XP_029145904.1_Arachis_hypogaea
-----QWTEEQKS-VLSAI------E--------------------------------QG-----KSVFITGSAGTGKTMLVIEVIKR---L-K--K--MHT----------------
-----P-SKVFITASTGVAAVAL-------KG------QTLHSF-GGIR--G--P------FY-H---D---P-------KK---L--------------F-----ESI--L--A---
-D----N---R-------------AVRRW--QK--ANALVVDECSM--VDGELFDG-LEYVARKVR----------------G---------V-----------------------D-
--------------------------EMWGG-----IQMVVVGDFCQLPPIP-----------------ND---S--S----------------------------------------
------------------------KPV-----K-----------YAFEA------------------RCWDE----SF--H-LQKELTKVFR----Q-S-----------D--P----
---------QFIELLQRMRKG---E-I-D------S--------------LD------------------------------LS--LL-EK--CY---------SERV----------
---------------------------------------------------------------------C------------------------D---------S-S-----------
V-------------------VKLFPLKKKVMEVNEKMLK-SL----Q---K--------------------------D-----------------------V-T-V-Y----------
----------P---------------------AVD------T----------------------G--------------------K---------------------------D-T--
-----W----------------------------------------------------------K--------------K--------------------LL-----------N-Q--
------GIA----------PDQL-----ELCEGSRVMLIKNLD----------V----R--K-----G-----LVNGATGTVVGF
>XP_030518540.1_Rhodamnia_argentea
-----EWTEEQTR-IISAV------S--------------------------------GG-----RSVFIAGSAGTGKTALLKHIIKL---L-K--D--SLG----------------
-----R-STVFVTASTGVAACAL-------RG------QTLHSF-AGI---K--N------PG-R---E---A-------SA---L---------------------DI--Y--M---
-D----K---K-------------ACKRW--RK--VRALFIDEISM--VDGELFDN-LECIARELR----------------E---------S-----------------------G-
--------------------------ETWGG-----IQLIATGDFLQLPPIP-----------------RK---G--N----------------------------------------
------CLS----------------SK-----Q-----------FAFEA------------------DCWQS----SF--D-LQIELTKVFR----Q-S-----------D--E----
---------RLVKVLQGIRKG---E-I-S------P--------------DD------------------------------WE--FL-EQ--SC---------ATD-----------
-----------------------------------------------------------------------EP---------------------D---------P-S-----------
V-------------------VRLYPRNEDVNEVNNYKIE-EL----A---A--------------------------E-----------------------G-Y-V-F----------
----------T---------------------AAD------S----------------------G--------------------S---------------------------D-P--
-----W----------------------------------------------------------K--------------R--------------------QL-----------K-R--
------GMA----------PDEI-----FLCKGARVMLIKNKN----------T----S--R-----G-----LVNGAVGTVVGF
>XP_026396572.1_Papaver_somniferum
-----KLTKQQKQ-VLEEV------S--------------------------------KG-----KSVFITGSGGTGKTFLLKQIVNL---L-K--Q-EVHK----------------
-----P-DEVFVTASTGVAACAL-------NG------QTLHSF-AGI---G--L------GE-D---D---E-------DE---L--------------L-----GRV--C--K---
-N----K---L-------------ASQRW--KQ--VKALVIDEISM--ISGELFDK-IEYIAQMCKPK--------------R---------R-----------------------G-
--------------------------EIWGG-----IQLIVSGDFFQLPPII-----------------KY---S--N----------------------------------------
-------GE----------------VK-----E-----------FAFEA------------------ECWNE----SF--D-LQIELTRVFR----Q-S-----------D--S----
---------QFIELLQRIRKG---Y-R-D------A--------------NM------------------------------LK--LL-DK--CC---------LNEL----------
---------------------------------------------------------------------V-NV---------------------P---------S-D-----------
V-------------------PRLFPRNEDVKRLNNERLK-NL----G---Q--------------------------E-----------------------I-V-S-Y----------
----------R---------------------AVD------R----------------------G--------------------V---------------------------N-P--
-----W----------------------------------------------------------R--------------N--------------------QL-----------Q-Q--
------GIA----------PDVL-----EICLGARVMLIKNKD----------V----E--A-----G-----LVNGAVGTVIGF
>KAF6167112.1_Kingdonia_uniflora
-----TLSKQQQE-VLDAI------S--------------------------------KR-----KSIFITGSAGTGKTHLLLQIIKT---L-K--T--IYK----------------
-----P-REVFVTASTGIAAFAI-------NG------QTIHSF-AGV---G--F------SD-A---D---T-------NV---L--------------L-----NRV--V--K---
-N----K---F-------------ATNRW--RN--VKALVIDEISM--INGHLFDD-LEYIAREVRPVLS------------G---------E-----------------------V-
--------------------------ESWGG-----IQLIVCGDFFQLPPVN-----------------KG---E--H----------------------------------------
------------------------IVK-----E-----------FAFEA------------------NCWKS----SF--D-LLVELTRVYR----Q-S-----------D--P----
---------RLLVLLQGIRRG---Y-T-N------T--------------HH------------------------------LE--IL-KQ--CC---------KRPI----------
-----------------------------------------------------------------------ET---------------------T---------V-V-----------
V-------------------PRLYPMNDDVKRVNDANLG-LL-RRSG---K--------------------------E-----------------------I-F-T-Y----------
----------R---------------------AND------K----------------------G--------------------E---------------------------C-P--
-----W----------------------------------------------------------K--------------D--------------------QL-----------K-S--
------GIA----------PDTL-----ELCIGARVMLIKNKD----------F----H--S-----G-----LVNGATGTVINF
>KAF5202456.1_Thalictrum_thalictroides
-----NLSNQQQS-ILKAI------T--------------------------------ET-----QSVFISGPAGTGKSYVVSLATEL---L-R--R-KIYQ----------------
-----P-YEVFVTASTGVSACAL-------NG------QTLHSF-AGI---G--L------GE-G---E---K-------EV---L--------------L-----KKV--L--K---
-N----G---K-------------ACSRW--RT--AKALVIDEISM--IECDLFEK-IEYIARNIR----------------G---------A--------------------AHRN-
--------------------------KPWGG-----IQLIVSGDFFQLPPIM-----------------KE---Q--E----------------------------------------
-------HL----------------GK-----E-----------FAFEA------------------TCWEA----SF--D-LQVELTQIFR----Q-T-----------D--L----
---------DFINLLQRVRRG---Q-K-D------E--------------HH------------------------------LE--LL-HH--CC---------NVLT----------
-----------------------------------------------------------------------DS---------------------S---------E-S-----------
V-------------------PSLFPRNKDVNRVNEGRLR-RL----G---N--------------------------E-----------------------T-F-K-Y----------
----------T---------------------ARD------S----------------------G--------------------K---------------------------Q-P--
-----W----------------------------------------------------------K--------------G--------------------QL-----------K-L--
------GIA----------PDEL-----EICIDARVMLIKNKD----------L----R--A-----G-----LVNGATGTVVDF
>XP_010275116.1_Nelumbo_nucifera
-----PWTDQQLE-VLKAV------A--------------------------------EG-----QSVFITGSAGTGKTILLRRVVEV---L-K--Q--IHN----------------
-----P-KHVFVTASTGVAACAL-------NG------HTLHSF-AGI---G--R------RT-Q---D---R-------EA---M--------------L-----FNA--T--S---
-N----K---G-------------AFYRW--KR--AKALVIDEISM--VDADLFDT-LGYISGEIRYE--------------K---------S-----------------------S-
--------------------------EIWSG-----IQLIVSGDFFQLPPVW-----------------NRL-SS--S----------------------------------------
-------ES----------------GK-----E-----------FAFEA------------------DWWND----SF--D-QQIELTQVFR----Q-S-----------D--L----
---------KLIELLQGIRRG---E-T-D------P--------------EM------------------------------LR--LL-YS--RT---------VTSE-----P----
```

```
----------------------------------------------------------------------------------------------D---------S-K-----------
V-------------------IRLFPRKDDVNRVNQERLR-SL----G---R--------------------------E-----------------------T-I-T-Y----------
----------T---------------------ALD------V----------------------G--------------------Q---------------------------E-P--
-----W----------------------------------------------------------K--------------S--------------------EL-----------K-L--
------GIA----------PDEV-----ELCVGARVMLTKNIA----------L----S--D-----G-----LVNGATGTITGF
>MQL92731.1_Colocasia_esculenta
-----LLTPQQEA-VLRAV------Q--------------------------------QG-----CSIFITGSAGTGKSFLLGHIIAA---L-R--R--IHQ----------------
-----P-DAVFVTASTGIAACAL-------GG------QTLHSF-AGI---G--L------GR-G---D---R-------DT---L--------------L-----RRA--A--T---
-S----H---G-------------AAKRW--RR--AAALVIDEISM--VDGGVFDA-LDYIARALR----------------W----------Q----------------------H-
--------------------------RRWGG-----LQLVVSGDFFQLPPIK-----------------AP---D--P----------------------------------------
-------------------------TK-----E-----------FAFEA------------------DCWDS----SF--D-LQVELTHVFR----Q-S-----------D--S----
---------RLIDLLQGIRRG---E-P-I------P--------------QQHL----------------------------LP--LL-EP--CS---------KGGE----------
-------------------------------------------------------------------CDR-DD---------------------E---------E-N-----------
V-------------------TRLFPRNDDVRRVNEERLR-SL----G---R--------------------------E-----------------------V-I-T-F----------
----------V---------------------AAD------T----------------------G--------------------S---------------------------E-P--
-----W----------------------------------------------------------R--------------S--------------------QL-----------R-Q--
------GIA----------PEVL-----ELCVGARVMLIKNTD----------P----A--A-----G-----LVNGSTGVVTGF
>RWR91934.1_Cinnamomum_micranthum_f._kanehirae
-----VLTEKQKE-VLKAV------V--------------------------------DG-----RSVFITGSAGTGKSFLLHHIIHL---L-R--L--LHS----------------
-----P-RNVFVTASTGVAACAL-------NG------LTLHSF-AGV---G--L------AN-D--PS---P-------DL---L--------------L-----HKV--R--R---
-N----I---P-------------AFKRW--RF--AKALVVDEISM--IDGQLFDR-LEFIARSLR----------------P----------G----------------------R-
--------------------------KVWGG-----IQLIVAGDFFQLPPVN-----------------SP---D--P----------------------------------------
-------------------------NR-----E-----------FAFEA------------------DCWSN----SF--H-LLVELTHVFR----Q-S-----------D--A----
---------RLVELLQAIRKG---R-S-D------H--------------ID--W---------------------------FH--FL-NS--CF---------VEPH----------
--------------------------------------------------------------------FG-ER---------------------D---------N-T-----------
V-------------------TRLYPRNEDVRRVNEEKLR-SL----G---G--------------------------E-----------------------V-I-T-Y----------
----------I---------------------AQD------E----------------------G--------------------G---------------------------E-S--
-----G----------------------------------------------------------K--------------K--------------------QL-----------K-Q--
------GIA----------PQEL-----ELSLGARVMLIKNLD----------P----K--N-----G-----LVNGATGTVTGF
>XP_021855182.1_Spinacia_oleracea
-----QWTDQQLQ-VFEAI------E--------------------------------RR-----QSVFVTGSAGTGKTMLVQELIKL---L-R--K--IYG----------------
-----K-RNVSVTAPTGVVACAL-------GG------QTLHSF-AGV---G--L------AE-A---D---A-------ET---L--------------L-----SRV--L--D---
-N----R---T-------------VIKRW--KT--IKALVIDEISM--VEGELFDK-LEYIARTIR----------------E----------I----------------------D-
--------------------------EPWGG-----IQLVVSGDFLQLPPVN-----------------VG---K--S-S--------------------------------------
-------DN----------------RK-----E-----------FAFEA------------------DSWDS----SF--Q-LEVGLKTVFR----Q-S-----------D--P----
---------ELIKLLQGIRTG---E-L-D------A--------------EG------------------------------LE--LL-QQ--RR---------CFEE-----P----
----------------------------------------------------------------------------------------------D---------E-T-----------
V-------------------VRLFPRIADVNRVNDMRLK-GL----G---E--------------------------E-----------------------T-I-V-Y----------
----------E---------------------AFD------K----------------------G--------------------D---------------------------K-P--
-----W----------------------------------------------------------I--------------D--------------------QL-----------N-R--
------GMA----------PTKL-----QLCVGARVMLLQNLN----------V----K--G-----R-----LVNGATGTIIGF
>M_polymorpha
-----KLSKQQLK-VLKAI------S--------------------------------IG-----DSVFLTGSAGTGKSFVLEFAIRV---L-K--A--KYG----------------
-----A-SSVYVTASTGLAACAL-------GG------TTVHSF-AGV---G--L------GT-G---N---K-------ES---L--------------V-----DKV--K--S---
-R----R---E-------------SRTRW--QS--AKALVVDEISM--IDGEFFDK-LDYVGRIVR----------------K----------D----------------------S-
--------------------------RPFGG-----IQLVVTGDFYQLPPVN-----------------PE---N--P----------------------------------------
-------------------------VK-----Y-----------FAFEA------------------ECWNR----CF--H-LQVELLHVFR----Q-A-----------D--E----
---------EFVALLNEIRRG---G-C-S------S--------------EH------------------------------EE--KL-RK--CS---------GPVD----------
---------------------------------------------------------------------Q-SS---------------------G---------I-A-----------
L-------------------TRLYPRKVDVSRENEQNLR-AL----N---Q--------------------------P-----------------------T-V-M-F----------
----------I---------------------AKD------E----------------------A--------------------R---------------------------T-E--
-----F--------------------------------------------------------A-K--------------R--------------------QL-----------D-N--
------VRV----------EAIV-----ALSVGAQVMLAKNLE----------T----S--V-----G-----LVNGARGVVVGF
>KAG6555887.1_Marchantia_paleacea
-----KLSKQQLK-VLKAI------S--------------------------------IG-----DSVFLTGSAGTGKSFVLEFAIRV---L-K--A--KYG----------------
-----A-SSVFVTASTGLAACAL-------GG------TTVHSF-AGV---G--L------GT-G---N---K-------ES---L--------------V-----DKV--K--S---
-R----R---E-------------SRTRW--QS--AKALVVDEISM--IDGEFFDK-LDYVGRIVR----------------K----------D----------------------S-
--------------------------RPFGG-----IQLVVTGDFYQLPPVN-----------------PE---N--P----------------------------------------
-------------------------VK-----Y-----------FAFEA------------------ECWNR----CF--H-LQVELLHVFR----Q-A-----------D--E----
---------EFVGLLNEIRRG---G-C-S------S--------------EH------------------------------EE--KL-RK--CC---------GPVD----------
---------------------------------------------------------------------Q-SS---------------------G---------I-A-----------
L-------------------TRLYPRKVDVSRENEQNLR-AL----N---Q--------------------------P-----------------------T-V-M-F----------
----------I---------------------AKD------E----------------------A--------------------R---------------------------T-E--
-----F--------------------------------------------------------A-K--------------R--------------------QL-----------D-N--
------VRV----------EAIV-----ALSVGAQVMLAKNLE----------T----S--V-----G-----LVNGARGVVVGF
>KAG0621209.1_Ceratodon_purpureus
-----KPSPQQME-VLKAI------A--------------------------------QR-----KSVFVTGSAGTGKSFIVEDALQI---L-R--G--MYG----------------
-----D-DKVFVTASTGLAACAV-------GG------TTLHSF-AGV---G--I------GV-N--ET---K-------EQ---L--------------A-----DKV--L--K---
-K----R---E-------------VRARW--AK--AKALIIDEISM--IDGELFDK-LEYIARRVK----------------G----------R------------------AKGPD-
--------------------------EVWGG-----LQLIVTGDFFQLEPVK-----------------PS---N--P----------------------------------------
-------------------------QK-----Y-----------FAFQA------------------DCWDE----SF--D-VQVELSHVFR----Q-S-----------D--M----
---------KFVNMLNEIRRG---V-C-S------P--------------ST------------------------------LH--RL-RQ--CQ---------GPSE-----G----
---------------------------------------------------------------------A-SN---------------------G---------I-E-----------
M-------------------TRLYPHQMDVRRENDQNLR-SI----G---G--------------------------D-----------------------M-I-V-Y----------
----------K---------------------AKD------E----------------------A--------------------H---------------------------N-E--
-----F--------------------------------------------------------G-L--------------R--------------------QL-----------E-N--
------VRA----------AAVQ-----PLCVGAQVILLKNLE----------T----G--V-----G-----LVNGARGVVVRF
>XP_024357988.1_Physcomitrium_patens
```

```
-----VPSKQQME-VLKAI------T--------------------------------QQ-----KSVFITGSAGTGKSFIIEDALRV---L-R--Q--MYG----------------
-----E-DAVFVTASTGLAACAL-------GG------ITLHSF-AGV---G--I------GS-D-TET---K-------EQ---L--------------L-----TKV--R--K---
-R----R---D-------------VKARW--TK--AQALVIDEISM--IDGEFFDN-LEYIASKIK----------------G---------G-----------------------S-
--------------------------EPWGG-----LQLIVTGDFYQLEPVK-----------------PS---N--P----------------------------------------
-------------------------LK-----Y-----------FAFQA------------------ECWNR----SF--D-IQVELTHVFR----Q-L-----------D--M----
---------EFVNMLNEIRRG---V-C-S------P--------------ST------------------------------LH--RL-RQ--CQ---------GPPD-----R----
---------------------------------------------------------------------A-DN---------------------G---------I-E-----------
M-------------------TRLYPHQMDVRRENDQNLR-CL----G---G--------------------------D-----------------------M-I-I-Y----------
----------R---------------------AKD------D----------------------A--------------------S---------------------------T-S--
-----F--------------------------------------------------------A-Q--------------R--------------------QL-----------D-N--
------VRA----------AAVQ-----PLCVGAQVMLLKNLE----------T----A--A-----G-----LVNGSRGVVVRF
```

```
>XP_002987435.1_Selaginella_moellendorffii
-----TPSPEQLR-VLEAV------C--------------------------------NR-----QSVFVTGSAGTGKSYILERAIQV---L-R--T--VYH----------------
-----P-SAVYVTASTGIAACAI-------GG------TTFHAF-AGV---G--I------GL-S---K---K-------EQ---L--------------V-----DMV--M--R---
-S----K---E-------------KKQRW--LN--AAALVIDEISM--IDAELFDK-VDFVGRAVR----------------R---------S-----------------------K-
--------------------------ERFGG-----LQLIVTGDFFQLPPVQ-----------------KP---G--E----------------------------------------
-------------------------TK-----S-----------FVFNA------------------KCWKE----CF--D-LQMELTQVFR----Q-S-----------D--R----
---------EFVGMLNEIRRG---E-C-S------F--------------AT------------------------------ET--RL-KS--CT---------SIST----------
-----------------------------------------------------------------------AP---------------------G---------I-E-----------
P-------------------TRLYPRRADVDRENEQKLR-SL----NPS-S--------------------------K-----------------------S-V-T-F----------
----------S---------------------AKD------S----------------------G--------------------R--------------------------------
----------------------------------------------------------------T--------------Q--------------------ML-----------N-G--
------SRA----------EAEI-----TLAIGAQVMLIKNLG----------T----E--Q-----G-----LVNGARGIVVGF
```

```
>XP_024538624.1_Selaginella_moellendorffii
-----TMSLEQLR-VLEAV------A--------------------------------AK-----KSVFVTGSAGTGKSFILEYAIKV---L-R--E--LHG----------------
-----E-FAVFVTASTGIAACSI-------GG------TTLHSF-AGV---G--L------GQ-L---D---E-------RR---L--------------A-----AAV--M--A---
-S----K---E-------------SRSRW--TT--AKALVIDEISM--IDAELLDK-IDYVGRAVR----------------N---------R-----------------------P-
--------------------------ERFGG-----LQLLVTGDFFQLPPVQ-----------------KA---G--E----------------------------------------
-------------------------TK-----N-----------FAFQA------------------RCWRE----CF--D-LQMELTYVFR----Q-S-----------D--R----
---------NFVAILDEIRRG---R-C-S------P--------------ST------------------------------IE--SL-KA--CS---------VVSA-----A----
---------------------------------------------------------------------S-SS---------------------S---------P-P-----------
P-------------------TRLFPHLQSVDRVNKEKLA-AL----G---G--------------------------E-----------------------T-V-T-Y----------
----------I---------------------ARD------V----------------------G--------------------K--------------------------------
----------------------------------------------------------------I--------------H--------------------LL-----------S-G--
------CRA----------ESQI-----TLAVGAEVMLVKNID----------T----L--G-----G-----LVNGTRGVLVDF
```

```
>E_dispar
-----HLSSDQEL-VLKAA------L--------------------------------EG-----KSFFFTGAAGCGKSYVLSAIVEK---L-K--H---------------------
-----D-KEVYVTASTGIAACNV-------NG------MTIHSF-SGI---G--K------GE-G---S---S-------SE---L--------------W-----DKV--K--Q---
-D----K---K-------------ALKKW--NK--VEVLIIDEISM--IDGDLFDK-LEFVARKAR----------------N---------N-----------------------N-
--------------------------LAFGG-----IQMIICGDFCQLPPIS-----------------RN---G-------------------------------------------
-------------------------TT-----K-----------FAFES------------------NCWNR----VI--P-YCYYLTTVHR----Q-N-----------D--Q----
---------KFITLLNGIRIG---E-I-S------D--------------EM------------------------------VN--CL-KG--CC---------DKEC----------
----------------------------------------------------------------------------------------------K---------Q-Q-----------
C-------------------THLLSYIKEVDDVNTKELQ-KL----Q---G--------------------------N-----------------------E-V-V-Y----------
----------H---------------------SVD------T----------------------G--------------------N---------------------------S-I--
-----Y----------------------------------------------------------------------------------------------L------------T-S--
------MKI----------TDEL-----HLKAGAFVMINKNID----------V----E--R-----G-----LVNGSVGIVIGF
```

```
>XP_004258641.1_Entamoeba_invadens_IP1
-----TLSEDQKT-IVDSA------M--------------------------------RG-----ESFFFTGAAGTGKSHVLRVIVAA---L-R--R--N------------------
-----G-KNVFVTASTGVAACNI-------SG------MTVHSF-FGI---G--I------GS-G---T---V-------EE---L--------------L-----NKV--K--K---
-D----S---I-------------AKARI--RS--ADVLVIDEISM--IDDRLFDK-IETISRVIC----------------D---------S-----------------------P-
--------------------------KPFGG-----IQVILCGDFFQLPPVS-----------------SD---G-------------------------------------------
-------------------------LK-----R-----------FAFEG------------------EQWNK----VV--R-RMYNLSVVHR----Q-K-----------D--K----
---------EFIYVLNKIRYG---T-V-D------E--------------WC------------------------------LN--KL-RE--RI---------DQE-----------
-----------------------------------------------------------------------KK---------------------G---------E-T-----------
Y-------------------TILFSKLNDVDETNSWKLK-EL----H---N--------------------------E-----------------------S-K-L-F----------
----------K---------------------AKD------S----------------------G--------------------N--------------------------------
----------------------------------------------------------------T--------------Q--------------------LF-----------K-T--
------SKV----------PTTL-----ELKIGAFVMVTKNIS----------I----E--K-----N-----LANGSLGVVIGF
```

```
>A_castellanii
-----CLSEQQQR-ALDLA------E--------------------------------RG-----YSMFLTGSAGTGKSFLLRQMIER---L-R--L--KHG----------------
-----P-EAVAVTASTGVAAINI-------DG------MTLHKW-AGV---G--L------GN-E---G---I-------KV---M--------------L-----GRA--F--G---
------------------------KRKEY--KQ--TRVLIIDEISMPQIKSDLFDQ-LEYIARRVR----------------N--WKKVPA-K-----------------------E-
--------------------------KPFGG-----IQLICCGDFFQLPPVL-----------------DK---T--N-KRM------------------------------------
-----------------------SMSQ-----A-----------FAFNA------------------ESWQS----CI--D-VVVQLSKVFR----Q-K-----------D--E----
---------RFQGILNEIRQG---A-C-S------D--------------ES------------------------------RR--IL-NE--CV---------GRRF-----E----
---------------------------------------------------------------------D-DL---------------------D---------I-L-----------
P-------------------TKLHPTNQQVDAINQTHMD-EL----D---G--------------------------D-----------------------S-C-Q-Y----------
----------L---------------------AKD------KLPAA-----------------TG--------------------P---------------------------R-KI-
-----L--------------------------------------------------------T-Q--------------W--------------------LH-----------Q-S--
------CSA----------LEAL-----ELKEAAQVMLIRNLT------------------S-----K-----LVNGSRGVVLGW
```

```
>N_gaditana
-----ELSEEQKG-ILRSV------M--------------------------------EG-----HNVYYSGRAGSGKTHLLRAIIDR---A-P--A---------------------
-------GKTFVTASTGIAAVNV-------GG------TTLHSF-AGI---G--L------GD-D---P---L-------EV---L--------------K-----ERA--G--K---
-N----R---T-------------AAANW--AA--VEVLIVDEVSM--LHGSLLSK-LNEIAKHVK----------------N---------Q----------------------PH-
--------------------------RPFGG-----VQLIFTGDFFQLPPVS-----------------RG---R--R-AG-------------------------------------
-------------------------DH-----D-----------YAFLH------------------PVWKE----LFGPE-SCYELTRVFR----Q-A-----------E--K----
---------PLVALLNDVRYG---R-A-S------A--------------ES------------------------------IA--LL-QE--LS---------RDLK----------
```

```
---------------------------------------------------------------------P-PP---------------------G---------I-E-----------
P-------------------TLLFATNNRVDEMNQQKLG-LL----A---G-------------------------AE-----------------------D-H-V-F----------
----------Q---------------------ATD------S----------------------G--------------------V---------------------------E-P--
-----F--------------------------------------------------------L-G--------------Q---------------------LR----------K-N--
------CLA----------ESRL-----RLRTGAQVMLLKNVN----------A----N--L-----G-----LVNGAKGRVTSF

>P_multistriata

-----SLTEEQRK-AAEWI------FGN----A-GEDHD-----------EE----DSAP-----RNVFVTGSAGTGKSHLLKYIVHA---L-Q--S--RES----------------
-DGTGE-ARVGVCAPTGVAAVIV-------GG------STLHSF-FGI---G--L------GK-G---S---P-------SS---I--------------L-----QKV--R--K---
-N----S---S-------------AMDRI--DD--TDVLIIDECSM--MSSELLET-LDMVSRRIR----------------N---------G-------------------GVFRD-
--------------------------EPFGG-----MQVIAFGDFFQLPPVY-----------------RN---D--G-TKD------------------------------------
-----------------------WTWR-----P-----------FCFES------------------PVWED----LGLSE-NVVELREVQR----Q-E-----------H--G----
---------DFVDLLNKVRIG---K-V-T------E--------------QD------------------------------IR--EL-NRQ-CL-IGP-----NNPI----------
---------------------------------------------------------------------P-TD---------------------G---------I-L-----------
P-------------------TRLYVLNKDVDSENINRLA-EL----K---G--------------------------R-----------------------E-I-V-C----------
----------K---------------------ASD------NWR-----QSMP----------LG--------------------T---------------------------P-A--
-----A--------------------------------------------------------T-K--------------K----------K---------MK-----------E-S--
----IAMEI----------PDEV-----RLKIGAQVMLTRNKD----------L----Q--R-----N-----LVNGSRGVVERI

>CAZ69470.1_Emiliania_huxleyi_virus_99B1

-----SLTVCQQD-VLNKT------L--------------------------------NG-----KNVLISGSAGTGKSFLTRHIIHQ---L-K--L--QKG----------------
-----P-NNVGVVSPTGIAAANI-------NG------TTIHAW-GGI---G--I------GD-G---D---A-------TA---L--------------I-----KKA--R--G---
-N----R---L-------------AFTRW--KT--ARVLIIDEVSM--LDGELFNK-LEKIAQSIR----------------S---------N-----------------------S-
--------------------------RPFGG-----IQLILVGDFYQLPPVT-----------------VT-----------------------------------------------
-------------------------DA-----G-----------FCFES------------------DAWNA----AN--I-EKCELTEVIR----Q-Q----------ND--T----
---------EFISILNSIRIG---Q-C-A------T--------------ET------------------------------EN--AL-AK--CH---------VSVK-----PP---
---------------------------------------------------------------------P-SD---------------------G---------I-V-----------
P-------------------TKLYCINRDVDRENELFLE-RL----P---G--------------------------E-----------------------R-V-L-F----------
----------K---------------------AID------VFN-----ASTP----------AG--------------------A---------------------------ETK--
-----L--------------------------------------------------------V-D-----------------------------------ML-------------N--
------KKT----------HQVL-----PLKIGAQVMITKNMA---------------D--F-----S-----LVNGSRGIVTDF

>CAE7678393.1_Symbiodinium_microadriaticum

-----QLTAEQRA-AASRA------L--------------------------------AG-----ENLFLTGPAGTGKSFLLRFLVQE---F-S--H--RH-----------------
-----P-GQVAVTASTGIAAAHL-------GG------QTIHSF-AGV---G--V------GT-A---P---L-------AK---T--------------L-----QQV--Q--R---
-S----S---A-------------AVQRW--KS--TKVLVIDEISM--IDGELLEL-LCGVARAVR----------------K---------Q-----------------------S-
--------------------------APFGG-----LQVLFCGDFLQLPPVQ-----------------ER---G--K-L--------------------------------------
-------------------------AR-----K-----------FCFAS------------------SAWKK----AGLHE-GTVLLCQTVR----QAS-----------D--V----
---------QFGKVLNELRIG---H-V-S------D--------------EA------------------------------RE--ML-AK--CH---------VGVK-----S----
--------------------------------------------------------------------QP-KD---------------------G---------I-L-----------
P-------------------TKLYCLNKNVDAENAARLR-QL----P---G--------------------------A-----------------------P-Q-I-L----------
----------R---------------------AHD------LCP-------------------RN--------------------T---------------------------T-P--
-----A----------------------------------------------------------Q--------------MA-------------------LL-------------D--
------KKV----------PAEL-----HLKVGAQVLHLKNEP---------------N--L-----G-----LVNGSRGIVEAF

>RHW71036.1_Trypanosoma_brucei_equiperdum

-----SLSPEQQR-ALRLA------L--------------------------------KG-----RNLFITGGAGSGKSLLIREIVYQ---L-R--H--N------------------
----KR-RCVYVTATTGVAALNV-------RG------STVNSF-AGV---K--F------GD-G---D---A-------RQ---L--------------L-----KWV--R--R---
-S----R---R-------------AAGRW--RY--CQTLIIDEISM--MDPLLLDK-LDVIARAIR----------------R---------R-----------------------N-
--------------------------EPFGG-----IQVILCGDFLQLPPIP-----------------PR---N--K-PQQ----------KTEENAEAQEGG--------------
---DPTDGT------------PAPSKL-----Q-----------YCFET------------------STWTS----LN--L-ITVILHKKFR----Q-H-----------D--D----
--------LAFQQVLDELRVG---S-L-S------P--------------ES------------------------------YE--LL-LS--RT---------VASK-----
SSAKSRKKK-------------------------------------------------------------DEDAG-ND---------------------G---------V-L------
-----P---LTDAETTPAAAEKDRHVRLCATNKEVEMRNAKYFA-AL----EPK-G-LPIYPSPNDGSSQQTGSTNGANSVTEED-----------------TMRPL-Q-V-Y-----
---------------R---------------------AYD------AYSTH---ETEPET---TEETTTG--------------------T---------------------------
Q-P------------------------------------------------------------------S--------------Q-------PW-----------VR----------
FE-D--------STL----------PTDL-----ALKVGTRVMVLQNIS----------L----R--L-----G-----LVNGSVGEVVGF

>XP_029239885.1_Trypanosoma_rangeli

-----NLSTEQQR-VFDLV------VK-------------------------------YG-----RSVFLTGGAGTGKSHLLRAIIET---L-P--R---------------------
-------ASTFVTATTGIAALNL-------GG------STLHSF-AGC---G--I------VD-AQTHV---A-------ED---V--------------C-----RTV--R--G---
-K----A---K-------------AKRNW--RF--CKVLVVDEVSM--MDAWFFDV-LEYVARKIR----------------G---------S-----------------------N-
--------------------------SPFGG-----IQLVLAGDFLQLPPVV-----------------KQ---R--G-Q--------------------------------------
-------------------------DP-----R-----------FCFES------------------ETWCR----VN--P-RVCILSQRFR----Q-S-----------D--E----
---------MFFGMLNEIRCG---V-L-T------A--------------PS------------------------------LA--LL-AS--LS---------STTT-----
VRFVQEQA---------------------------------------------------------------KTVKTEN-----------GDAGALETPSQ---------L-L------
-----PTGDVVDSRGRTREERHDGFSILRARRVEVDDVNFKRFN-EL----T---T--------------------------E-----------------------I-F-S-Y-----
---------------R---------------------GFH------W----------------------G--------------------E---------------------------
G-K-------Y-----------------------------------------------------------------------------------------------------------
--P--------SDL----------PAVV-----SVRVGCRVMLLKNLD----------V----S--V-----G-----LVNGSVGTVENF

>T_grayi

-----GLSAEQQT-VFDLV------VR-------------------------------RG-----RSVFLTGGAGTGKSHLLRAIIEA---L-P--R---------------------
-------QTTFVTATTGIAALNL-------GG------RTLHSF-AGC---G--I------VD-RHRHT---P-------QD---V--------------F-----NTV--R--G---
-K----K---T-------------AKKNW--RT--CRVLVVDEVSM--MDGWFLSV-LEYVARMIR----------------G---------S-----------------------F-
--------------------------APFGG-----IQVVFAGDFLQLPPVS-----------------KA---N--R-QGG------------------------------------
-----------------------RQEA-----T-----------LAFEA------------------AAWRR----IN--P-RVCVLSQRFR----Q-K-----------D--E----
---------VFFGILNEMRSG---A-L-T------A--------------TS------------------------------IA--ML-TS--LS---------SATT-----
VALLRDDEVTHGN--------------------------------------------------------DDKTVALE----------------------G---------V-E------
---ALSSDIVVDSRGRTKQERYEGFTILRALHKEVEAVNIECFN-KL----T---T--------------------------E-----------------------I-V-S-Y-----
---------------K---------------------GYH------T----------------------G--------------------E---------------------------
G-R-------F-----------------------------------------------------------------------------------------------------------
--P--------AEL----------PAVV-----SVRAGCRVMLLKNLD----------V----S--Q-----G-----LVNGSVGTLVRF

>L_braziliensis
```

```
-----ALSSEQRY-AFHVT------VK------------------------------EH-----HSAFITGGAGTGKSHLLRTIIRA---L-P--A---------------------
-------SSTFITATTGIAALNL-------SG------STLHSF-AGC---G--I-----PN-R-SST---R-------DS---L--------------L-----SSV--L--S---
-K----Q---R-------------CVRSW--RI--CRVLIIDEVSM--LEPSFFGL-IDYIARHVR---------------N---------R----------------------PH-
--------------------------EPFGG-----IQLILSGDFLQLPPVS----------------RE---R--R-DS-------------------------------------
-------------------------SP-----Q-----------FCFET-----------------ESWWK----VN--P-TVCLLSTPFR----Q-R-----------N--L----
---------RFFSILNEMRFG---E-L-Q------P-------------DS------------------------------VE--LL-YS--MD---------TTER-----
VHFVQRTDVTASL------------------KVGSGDGPSVVRVGHKRGADVSAEAVSRGAATCK----------TE----VSS--T-----------S---------MQQ------
-----TRLELVNGAGRAIDAPFDGYTILRATRAEVDAENQKYYH-QL----T---T-------------------------E-----------------------E-F-V-Y-----
---------------R---------------------GFH------T---------------------G--------------------R---------------------------
G-A-------F----------------------------------------------------------------------------------------------------------
P-D--------GAL----------AKVV-----QLRKGCRVMLIKNFD----------S----R--L----G-----LVNGSTGTVTDF
>L_seymouri
-----ALSDEQRY-AYRLA------VH------------------------------EH-----RNVFITGGAGTGKSHLLRAIIKD---M-P--C---------------------
-------STTFVTATTGIAALNL-------SG------TTLHSF-VGC---C--V-----PD-K-RAK---P-------SK---L--------------L-----STV--A--S---
-N----A---R-------------CLRNW--RL--CRALVIDEVSM--LEASFFDL-VDYIARHVR---------------N---------R----------------------PR-
--------------------------EPFGG-----IQLILSGDFLQLPPVV----------------KE---R--R-DG-------------------------------------
-------------------------SA-----P-----------FCFET-----------------KTWIR----VN--P-RICLLSAPFR----Q-R-----------D--L----
---------RFFEILNEMRFG---D-L-Q------P-------------DS------------------------------VA--LL-RS--IT---------TTNS-----
VHFARRLA-------------------------DWENQTVKVGLKRERGPPTDSPAGSSACTT--------SVV-DDTTNRCSS--TFYDVSARP---G---------ERQ------
-----ARLELVDGAGRAVDAPFDGYTILRSTRAEVDAQNEWHFR-RL----D---T-------------------------E-----------------------I-F-T-Y-----
---------------V---------------------GAH------S---------------------G--------------------L---------------------------
G-A-------F----------------------------------------------------------------------------------------------------------
P-A--------NNL----------SEVV-----RLRKGCRVMVIKNFD----------A----Q--T----K-----LVNGSTGTVTGF
>B_saltans
-----TLSAEQQF-ILDLV------VK------------------------------HQ-----RSVFLTGGGGTGKSFLLREIIDQ---L-D--K---------------------
-------RTTFVTAPTGIAALNV-------GG------VTLHSF-AGI---G--I-----GE-G---S---R-------DD---L--------------L-----GRV--R--G---
-N----K---A-------------AKLQW--LG--CRVLIIDEVSM--VPKKLLDD-LEFIARKIR---------------G---------R-----------------------N-
--------------------------EPFGG-----IQLVLCGDFLQLPPVN----------------RR---S--G-RQS------------------------------------
-----VQAN----------------EC-----D-----------FCFAS-----------------AAWDR----IN--P-RVFFLRTLFR----Q-H----------TD--S----
---------LFATILNELRLG---E-L-S------H-------------DS------------------------------IH--TM-MS--IS---------HSTR-----
AAFVD-------------------------------------------------------------------TT-AN---------------------G---------E-I------
--------VVTDDVGAQGEDRRGGRTVLRSTNNVVKNINSDCFD-EL----N---T-------------------------D-----------------------V-Q-S-Y-----
---------------T---------------------AVT------G---------------------G--------------------P---------------------------
Q-P--------------------------------------------------------------------------------H--------------------LL-----------
D-Q--------CPA----------ESEV-----SLRVGARVMLLKNLD----------Q----R--A----G-----LVNGSIGVVTTF
>P_fungivorum
-----PMSDEQAE-IYAAV------M-------------------------------SG-----NNLFFTGSAGTGKSFLLKKIWAG---L-D--K--L------------------
-----G-KKVAMTAPTGIAAVNV-------GG------ITLHKW-SGV---G--V-----ST-A---I---T-------RE---L--------------R-----EEE--M--R---
-N----R---A---------W-G-NQATW--KD--TEVLIVDEVSM--VSGELFDL-LEDVARGIL---------------D---------N-----------------------D-
--------------------------RPFGG-----MQVICCGDFLQLPPVP----------------DR---G--E----------------------------------------
-------------------------TV-----S-----------FCFES-----------------ESWKR----VI--G-LSRELKTIHR----Q-A-----------D--P----
---------IFANMLNDIRLG---N-V-S------P-------------ST------------------------------SD--IL-MD--LQ---------RKHV-----
QRKSQRE---------------------------------------------------------------AEKD-SG---------------------A---------I-L------
-----P-------------------TTLYSKNVDVDRVNNEQLS-DL----P---G-------------------------A-----------------------T-V-T-F-----
---------------R---------------------SED------NAYPF----------------DGVFD-----------------SRTL------------------------
E-N---------------------------------------------------------------L-K--------------Q--------------------LL-----------
N-P--------LTI----------NEEI-----HLKVGAQVMLLSNIS------------------D----E-----HINGSRGIITSF
>T_socialis
-----ALDPAQQL-VVDKV------M-------------------------------RG-----ESVFFTGSAGTGKTFLLNTILNM---L-K--E--KWG----------------
--ASFG-DHVAVAAMTGIAATHI-------EG------TTLNAA-IGI---G--A-----PS-R---Y---R-------DF------------------------KTM-----H---
-R----P---D-------------VRARI--KG--WDVLVIDECSM--MSAEMFEI-IEHMFRVIR---------------R---------S-----------------------Q-
--------------------------RPAGG-----LQLILCGDFFQLPPVC----------------KV---S--M-ADA------------------------------------
---PPPMDV--------------YSNY-----G-----------YAFQA-----------------PAWQR----VFTIE-NRVILTQIFR----Q-S-----------D--A----
---------TFAGMLNCIRVG---E-G-S------R-------------QV------------------------------TA--RLVSE--CG---------RAIS----------
--------------------------------------------------------------------C-AE---------------------G---------I-K-----------
P-------------------TQIFARNADVDRINMAELV-AL----P---G-------------------------Q-----------------------S-V-L-C----------
----------A---------------------SID------EVIFTT---------------EA-------------------DA---------------------------A-T--
-----R-------------------------------------------------------K-R--------------D--------------------FF-----------K-D--
------CIA----------AQQL-----SLKEGAQVMLLKNLD---------P----A--G-----G-----LVNGSRGVVTGF
>T_socialis2
-----MLDAIQQE-VTDKV------L-------------------------------RG-----ESVFFTGSAGTGKTFLLNTILQC---L-K--E--KWG----------------
--DLYG-ERVAVTAMTGIAATHI-------EG------TTFNAA-MGI---G--A-----PS-R---Y---R-------DF------------------------LTM-----H---
-R----K---D-------------VRARI--KA--MYVLVVDECSM--MSGEMFAI-VEFMLRTIR---------------K---------N-----------------------S-
--------------------------RPAGG-----LQLILCGDFFQLPPIT----------------KI---S--M-ADP------------------------------------
---PPQRDA--------------FTNY-----G-----------YAFQA-----------------PSWRQ----VFSEG-NHIVLTRIFR----Q-S-----------D--E----
---------SFAAVLNSIRLG--EEGV-K------Q-------------IT------------------------------AR--LV-AE--CS---------REVT----------
--------------------------------------------------------------------C-AE---------------------G---------I-K-----------
P-------------------TQIFARNADVDRINMAELA-AL----P---G-------------------------D-----------------------A-V-Q-F----------
----------R---------------------SVD------EYALK----------------AG------------------VEE---------------------------S-K--
-----T-------------------------------------------------------Q-K--------------D--------------------FL-----------R-D--
------CIA----------AHDI-----SLKEGAQVMLLKNLD---------P----M--G-----G-----LVNGSRGVVTGY
>M_polymorpha2
-----PFSPEQQR-VIKLV------N-------------------------------EG-----KNIFFTGAGGTGKTYVLKYIIAS---L-K--K--
KFGIKHRAPQPGAVEGSIVYCTCFT-CSVAVTAATGIAALAI------GG------TTLHSA-TGI---G--V------PR-R---I---R-------DF-----------------
-------ARM--Y-------Q----T---R-------------VKTKW--RN--LKVLIIDEISM--ISAEVFEY-LEQTITEVR---------------K---------A------
-----------------SEEEMDLEEATALRSVNEADVRPKEEQFRPFGG-----LQVILAGDFFQLLPVL----------------NR---F--D-DFTV----------------
----------------------KPTWDE-------------LTNR-----G-----------LAFQA------------------PAWQK----AE--L-EVVVLQMMFR----Q-
N-----------D--K-D-----------YFVKLLQNIRTG---L-N-P------D-------------SV------------------------------EE--IV-LK--CS---
```

```
------RELK-------------------------------------------------------------------------------C-EH---------------------G---
------I-K-----------P-------------------TQLYPRNKEVKELNEKELT-KL----R---T--------------------------R--------------------
---E-E-V-I--------------------I---------------------SVD------TFETSE--EKLLLK--------PS--------------------ELDN----
LQAVCRRH------------E-R-------H----------------------------------------------------------K--------------RM------------
-------LQDH------GFWH-A--------CIA----------DQVL-----RLKTGAQVMLIRNIKRP------GSQ----K--L-----S-----LVNGSRGIIVGW
>KAG6541057.1_Marchantia_paleacea
-----PFSPEQQR-VIKLV------N--------------------------------EG-----KNVFFTGAGGTGKTYVLKYMIAS---L-K--K--
KFGITHRAPQPGAVEGSVVYCSCFT-CSVAVTAATGIA-----------GG------TTLHSA-TGI---G--V------PR-R---I---R-------DF-----------------
-------ARM--Y-------Q----T---R-------------VKTKW--RN--LKVLIIDEISM--ISAEVFEY-LEQTITEVR----------------K---------A------
-----------------SEEEMDLEEVTALRSITEASVRPKEDQYRPFGG-----LQVILAGDFFQLLPVL-----------------NK---F--E-DITV----------------
----------------------KPSWDE--------------LTNR-----G-----------LAFQA------------------PAWQR----AE--L-EVVVLQMMFR----Q-
N-----------D--K-D-----------YFVKLLQNIRTG---L-N-P------D--------------SV------------------------------EE--IV-LK--CS---
------RELK-------------------------------------------------------------------------------C-EH---------------------G---
------I-K-----------P-------------------TQLYPRNKEVKELNEKELT-KL----R---T--------------------------R--------------------
---E-E-V-I--------------------L---------------------SVD------TYETSE--EKLLLK--------PS--------------------ELDN----
VQAVGRRH------------E-R-------H----------------------------------------------------------K--------------RM------------
-------LQDH------GFWH-A--------CIA----------DQVL-----RLKTGAQVMLTRNIKRP------GSQ----K--L-----S-----LVNGSRGIIVGW
>KAG0609116.1_Ceratodon_purpureus
-----TLTPEQQQ-VLDFV------Q--------------------------------AG-----KNVFFSGPGGTGKTVVLREIVEF---F-K--W--RFDKHHSDYLHLG-----
HTCGCFA-CNVAITAPTGIAAIPI-------GG------STLHRA-TGI---G--I------PR-R---P---R--------DF-----------------------NRM--W-----
--D----K---P-------------IRLKW--RN--LSVLIIDEISM--VSAELLEY-LEQTIRRIR----------------T---------K-----------------------
K--------------VN----LPG----EPFGG-----LQVIIAGDFFQLQPVE-----------------DK---D--T-KTC----------------------------------
--------SNQ--------------FLNR-----G-----------LAFEA------------------PAWDR----AN--L-KTVILKRVFR----Q-K-----------D--D--
-----------HFVALLNGIRTG---E-N-K------A--------------AL------------------------------EE--IV-EN--CS---------RPLP--------
-----------------------------------------------------------------------V-KN---------------------G---------I-Q---------
--P-------------------TVLYPRNVEVDQFNKQKLN-GL----M---S--------------------------R-----------------------E-V-V-I--------
------------N---------------------ADE------EILTE-----------------EG----------------------------LKAVEERN------------E-
D-------LRR------------------------------------------------------VQE--------------RI-------------------LIDA------EFWK-
D--------CIA----------PDQV-----KLKVGAQVMLLRNLD----------QK-GNE--N-----D-----LVNGSRGILVGW
>P_oligandrum2
-----RLTEDQQR-VIDLI------K--------------------------------SR-----CNVFFTGSAGTGKSFLLQQILQPNGPL-R--SYLQ------------------
-----G-KRIYATATTGIAAYNI-------NG------MTLHHF-AGL---DPRA------AS-A---G---M-------KE---V--------------L-----VHV--R--R---
-N----R---D-------------ALQRW--RT--ADVLVIDEVSM--LDGRLFDT-LEALARELR----------------P---------E---------------------HHQ-
--------------------------RFFGG-----IQLVLSGDFFQLPPVA-----------------SR---N--E-RD-------------------------------------
-------------------------KM-----T-----------LCFES------------------SAWQS----GI--D-EIVQLSQVFR----Q-T-----------N--T----
---------AFVDILNAFRVG---Q-P-S------R--------------AM------------------------------LD--NL-NE--RC---------TRSI----------
------------------------------------------------------------------GTNE-ED---------------------D---------D-D-----------
A-------------------IRIFTHNNDVLEINSKRLD-EL----P---S--------------------------K-----------------------K-F-N-Y----------
----------I---------------------SAD------T----------------------G--------------------K--------------------------------
----------------------------------------------------------------R--------------E--------------------YL-----------A-G--
------CPA----------PPTL-----SLKKHARVMLIKTIN----------P----A--S-----G-----LVNGCRGVITGF
>XP_009828150.1_Aphanomyces_astaci
-----KLTMKQAQ-VLQAI------Q--------------------------------KK-----ENVFFTGRAGTGKSFLLGHIRRA---M-P--K---------------------
-------QGLFLTATTGIAAFNI-------NG------MTLHHF-AGL---P--Q------VD-T--FD---V-------TM---L--------------M-----AAV--Q--R---
-N----R---Q-------------ALIRW--RD--AVLLVIDEVSM--LDGQMFDA-LETIARIVR----------------Q---------S-----------------------K-
--------------------------LFFGG-----IQLVLSGDFYQLPPVT-----------------KG-----------------------------------------------
-------------------------EP-----T-----------FCFES------------------QAWQR----GI--N-TSICLDQVFR----Q-S----------DD--P----
---------EFVAMLNAIRVG---T-H-T------S--------------AM------------------------------IK--TI-NA--RC---------VDRR-----R----
---------------------------------------------------------------------H-SA---------------------S---------N-E-----------
A-------------------IHIFSHNAEVLAMNNARLE-HL----D---G--------------------------D-----------------------I-H-D-F----------
----------F---------------------AID------T----------------------G--------------------D--------------------------------
----------------------------------------------------------------K--------------G--------------------LL-----------K-G--
------SPI----------PVRI-----QLKQGARVMLTKNLS----------V----A--A-----G-----LVNGSRGEVVGF
>P_multistriata2
-----ILSAEQTL-ALKLI------T--------------------------------EG-----RNVFVTGVAGTGKSLVLRKTLEY---V-Q--E--VYE----------------
-----P-NEYVAMAPTGSTAIAL-------EG------QTVHSF-AGI---G--I------PK---------I-------YK---D--------------F-----KRM--K--T---
-N----K---N-------------IRKRW--EE--LQVLILDEVSM--ISGEFFDS-LSKVVSDIR----------------N---------D-----------------------P-
--------------------------RPFGG-----IQLIVCGDFLQLPPIS-----------------PR---Q--W-EVD----------QTVKALQERE-------GL-------
---ETPEEARD----------WLFLNR-----G-----------FCFQS------------------VAWKE----AN--F-ELVELNHVFR----Q-R-----------N--E----
---------DFVRALQDIRVG---N-V-T------P--------------ET------------------------------IR--YL-RE--NC---------ERPL-----P----
---------------------------------------------------------------------E-NDL--------------------G---------I-Q-----------
P-------------------TILHSKNIDVARENLVDLN-KL----S---G--------------------------D-----------------------T-V-S-Y----------
----------E---------------------ASD------AVE-----PE------------KGVGP--------------WVKKDL-------------------------E-N--
----------------------------------------------------------------N--------------S--------------------FF-----------R-S--
------CLA----------ERKL-----QLKIGAQVMLIRNLS----------Q----N--S-----G-----LVNGSRGTIVGF
>MBS3922931.1_Nitrosarchaeum_sp.
---------TQDK-ALLIL------K--------------------------------TG-----ANVFLTGEPGAGKTYTINKYVAY---L-R--E--H------------------
-----G-VDYAVTASTGIAATHI-------GG------MTIHSW-SGI---G--I------KE-S---L---T-------KY---D--------------L-----DKI--A--T---
-S----E---Y-------------LNRRI--RK--TKVLIIDEVSM--LHADTLSM-VDAVCREIK----------------Q---------V-----------------------S-
--------------------------EPFGG-----IQVVLVGDFFQLPPIQ-----------------KK---A--I-EQK-------------------------QAEL-------
-----LYEKPASL-------IGAGRPA-----H-----------FAYES------------------DAWKR----LA--P-VVCYISEQHR----Q-E-----------D--E----
---------AFLELLLSIRRG---T-L-E------E--------------EH------------------------------YE--FL-KT--RY---------VERD-----EM---
----------------------------------------------------------------------------------------------P---------E-E-----------
V-------------------TKLYSHNLNVDRVNDEELD-KI----D---E--------------------------Q-----------------------E-K-I-F----------
----------E---------------------MTS------S----------------------G--------------------S---------------------------A-S--
-----L--------------------------------------------------------V-T--------------A--------------------LK-----------K-G--
------CLS----------PETL-----ILKKGSIVMCTKNNP---------------K--E-----H-----YVNGTLGTVVGF
>C_Zambryskibacteria
```

```
---------TQKQ-ALEIL------K--------------------------------TG-----ANVFLSGEPGSGKTYTVNQYVSY---L-R--S--R------------------
-----K-VEVAITASTGIAATHI-------GG------MTIHSW-SGI---G--I------KR-N---L---D-------KY---E--------------L-----DRI--A--S---
-N----E---R-------------IAKRI--RS--SKALIIDEVSM--LGPRTLSM-VDMVCREVK----------------Q---------S-----------------------D-
--------------------------QAFGG-----LQVLLVGDFFQLPPVV-----------------RR---G--E-SEF-------------------------QKTL-------
-----IEEA----------------TA-----R-----------FAYDA------------------PCWLT----TG--F-ITCYLTEQYR----Q-D-----------D--R----
---------NFLSILSAIRHN---A-Y-N------D--------------TH------------------------------HS--HI-EK--RR---------VTSE-----NA---
----------------------------------------------------------------------------------------------P---------D-D-----------
I-------------------PRLFSHNEEVERVNDQELA-KI----S---E--------------------------K-----------------------E-R-I-F----------
----------E---------------------MTS------Q----------------------G--------------------A---------------------------S-S--
-----L--------------------------------------------------------V-A--------------T--------------------LK-----------K-G--
------CLS----------PEVL-----RLKVGAKVMFTKNNP---------------Q--A-----G-----FVNGTLGEVREL
```

```
>C_Falkowbacteria
---------KQKE-AIAIL------E--------------------------------AG-----HNVLLTGPAGSGKTFLLNQFIAY---L-K--K--K------------------
-----G-IGVAVTASTGIAATHI-------GG------RTIHSW-AGI---G--I------KD-H---L---S-------SR---E--------------I-----QTL--S--K---
-R----S---Y-------------MKKQF--EK--TEVLIIDEISM--LHAHRLDM-VDAVCRAMK----------------K---------N-----------------------A-
--------------------------LPFGG-----IQVVMSGDFFQLPPIT-----------------PG---S--D----------------------------------------
-------------------------EA-----D-----------FVYKA------------------NVWPE----MD--V-RICYLEEQHR----Q-N-----------D--E----
---------KMIQILKSMRED---A-V-S------D--------------DI------------------------------LG--LL-NE--RL---------KEKP-----KF---
----------------------------------------------------------------------------------------------G---------L-R-----------
P-------------------VRLFTHNIDVDSINNTELE-KI----E---A--------------------------E-----------------------E-Y-V-Y----------
----------R---------------------MTG------D----------------------G--------------------E---------------------------K-K--
-----L--------------------------------------------------------M-E--------------S--------------------LK-----------K-N--
------CLA----------PDTL-----ILKEGAKVMFVKNKF----------KD--EK--V-----I-----YVNGTTGEVVGF
```

```
>MBP6882245.1_Candidatus_Levybacteria_bacterium
---------TQKD-ALNLL------K--------------------------------LG-----HNVYLTGPAGSGKTHLLNQYIDY---L-K--Q--Q------------------
-----K-VSVGITASTGIAATHM-------GG------TTIHSW-SGM---G--I------KD-T---I---T-------TP---E--------------I-----HDL--M--K---
-R----S---Y-------------LRKRF--LL--AKVLIIDEVSM--LHAHQLDI-VDAICRGFK----------------R---------N-----------------------Y-
--------------------------EPFGG-----MQVIMCGDFFQLPPVV-----------------KG---G--E----------------------------------------
-------------------------KP-----S-----------YVIDA------------------EVWNN----MR--L-QICYLDEQFR----Q-S-----------D--R----
---------SFLRVLSDIRSG---E-V-N------E--------------DT------------------------------VE--VL-SE--RL---------DKNP-----EG---
----------------------------------------------------------------------------------------------Y---------S-K-----------
P-------------------TKLFTHNADVDAINKKELD-EL----K---G--------------------------E-----------------------S-H-D-F----------
----------L---------------------MVG------R----------------------G--------------------S---------------------------P-K--
-----I--------------------------------------------------------V-E--------------T--------------------LR-----------K-T--
------CLA----------PERL-----SLKVGAQVMFVKNNW---------------D--V-----G-----YVNGTLGEVIGF
```

```
>NKQ38702.1_Methanosarcinales_archaeon
---------TQNE-ALDIL------K--------------------------------LG-----YNVFLTGPPGSGKTFLLNKYINY---L-K--K--Y------------------
-----R-RGVAITASTGIAATHM-------GG------VTIHSW-SGL---G--I------KE-K---L---S-------EQ---D--------------L-----KKL--L--R---
-K----S---Y-------------LKKRF--KN--TGVLIIDEVSM--LHAFQLDL-INKICQAFK----------------G---------N-----------------------S-
--------------------------KSFGG-----IQVICSGDLFQLPPVQ-----------------KG---G--G----------------------------------------
-------------------------VA-----K-----------FITES------------------EIWEN----MN--I-KICYLEEQYR----Q-E-----------S--G----
---------ELLNLLNHIRNN---A-V-N------E--------------AR------------------------------EI--LL-NN--KY---------KEN------TF---
----------------------------------------------------------------------------------------------S---------F-T-----------
S-------------------TKLYTHNIDIDTINSFELN-KI----D---E--------------------------K-----------------------K-F-V-Y----------
----------R---------------------MSS------T----------------------G--------------------D---------------------------K-N--
-----I--------------------------------------------------------V-A--------------I--------------------LK-----------Q-S--
------CLA----------PEKL-----VLKKGAKVMFVKNNF---------------D--K-----G-----YVNGTLGNVVDF
```

```
>MBD3359246.1_Candidatus_Buchananbacteria_bacterium
---------KQRD-ALNIL------K--------------------------------LG-----YNVFLTGAAGSGKTFLLNKYIKY---L-R--K--N------------------
-----D-IAVAITASTGIAATHM-------NG------RTIHSW-CGM---G--I------NL-K---M---N-------KS---Q--------------I-----NEI--V--N---
-K----D---Y-------------IYDNI--LN--TKVLIIDEVSM--LHSSQLDL-VDKICKTIK----------------N---------N-----------------------D-
--------------------------EPFGG-----IQVILCGDFFQLPPVS-----------------KE---S--D----------------------------------------
-------------------------TS-----E-----------YAFES------------------DIWNN----MD--L-KVCYLTEQYR----Q-N-----------D--K----
---------DFLGILKKIREN---S-V-D------Q--------------DV------------------------------KN--KL-IK--RI---------NAKA-----KS---
----------------------------------------------------------------------------------------------K---------L-H-----------
I-------------------TKLYSHNIDVDRINHNELK-KI----K---G--------------------------K-----------------------E-I-V-Y----------
----------E---------------------MHS------E----------------------G--------------------I---------------------------R-T--
-----M--------------------------------------------------------V-D--------------S--------------------LK-----------K-S--
------CLA----------PERL-----TIKKGAIIMFVKNNF---------------R--E-----G-----YVNGTLGEIIDF
```

```
>OGH84178.1_Candidatus_Magasanikbacteria_bacterium_RIFOXYA2
---------NQSQ-ALKIL------Q--------------------------------SG-----ANVFLTGSAGTGKTFLLNQFIDY---L-K--S--K------------------
-----K-IKVGVTASTGIAATHL-------NG------RTIHSW-CGM---G--I------ER-K---L---N-------DK---K--------------L-----KKI--L--R---
-R----E---E-------------VVDRI--SN--AQVLIIDEISM--LDADRLDL-VDKICRAVK----------------S---------P-----------------------F-
--------------------------SPFGG-----IQIVLCGDFFQLPPID-----------------P------------------------------------------------
-------------------------DS-----L-----------FAFSA------------------FSWRN----SD--I-KVCYLDEQFR----Q-D-----------D--D----
---------RFLNILNKIRAN---E-A-G------E--------------KE------------------------------LE--FL-KS--RL---------YQSV-----DC---
----------------------------------------------------------------------------------------------A---------S-K-----------
P-------------------TKLYTHNVNVDALNNFELA-RL----A---A--------------------------E-----------------------E-Q-V-Y----------
----------Q---------------------MTE------E----------------------G--------------------P---------------------------V-E--
-----L--------------------------------------------------------V-R--------------L--------------------LK-----------K-N--
------YPA----------HPEL-----KLKIGAIVMFIKNNF---------------D--S-----G-----YVNGTLGEVIEF
```

```
>C_Moranbacteria
---------KQDI-AFKIL------K--------------------------------DG-----YNVFLTGPAGSGKTYLLNQYITH---L-K--K--E------------------
-----N-VRYAVTAATGIAATHL-------SG------RTIHSW-SGV---G--I------HN-S---L---S-------ER---D--------------I-----KDI--L--K---
-N----H---L-------------IKERL--KN--TRVLIIDEISM--IHAHQLDL-INKITRLAR----------------A---------S-----------------------W-
--------------------------EPFGG-----MQVVFSGDFFQLPPII-----------------TK---D--S-DI-------------------------------------
-------------------------KR-----R-----------FVFDA------------------QIWKE----MD--V-KICYLSEQFR----H-C-----------D--N----
---------QIIQILGDIRNN---T-V-N------E--------------GT------------------------------VE--KL-QE--TG---------TDF------DS---
```

```
-----------------------------------------------------------------------------------------------S---------D-D-----------
V-------------------TRLFTHNIDVDKINERKLA-KI----P---G---------------------------K-----------------------S-Y-V-Y----------
----------D---------------------MLD------D-----------------------G--------------------Q---------------------------E-R--
-----I---------------------------------------------------------V-Q--------------A--------------------LK-----------K-S--
------CLA----------PEKL-----ILKEGALVMFIRNNF----------------D--E-----G-----YVNGTIGTVVDF

>MBI2830749.1_Chloroflexi_bacterium

---------IQSE-ALEIL------K---------------------------------NG-----HNVYLTGAAGSGKTYLLNAYIQY---L-K--A--N------------------
-----R-VNVGVTASTGIAATHM-------EG------ITIHSW-AGI---G--L------LR-T---A---S-------DK---E---------------I-----QAI--I--E---
-N----K---R-------------IVKRF--QK--TQTLIIDEVSM--LDADRLDL-LEKVARLAR-----------------G---------S-----------------------W-
--------------------------EPFGG-----MQVVLCGDFFQLPPVA------------------KA---G--E----------------------------------------
------------------------PLP-----R-----------FVYKS-------------------AAWEN----MN--L-KVCYLHGQYR----Q-G-----------E--E----
---------EFLRMLNAIRDA---S-V-D------E---------------TV------------------------------VS--RL-HQC-RA---------TAFSP----DS---
-----------------------------------------------------------------------------------------------L---------G-R-----------
V-------------------VRLYSHNLNVDLENNRELA-KL----P---G---------------------------K-----------------------E-Y-V-Y----------
----------L---------------------MEM------S-----------------------G--------------------I---------------------------P-A--
-----I---------------------------------------------------------A-E--------------S--------------------LK-----------R-G--
------CLA----------PEKL-----VLKIGAAVMFVKNNF----------------E--Q-----G-----YVNGTLGTVASF

>Flavobacteriaceae

---------QQEK-ALAIL------K---------------------------------SG-----KNVFLTGSAGTGKTYVLNEYIKY---L-R--A--R------------------
-----K-VPVAVTASTGIAATHM-------NG------MTIHSW-SGI---G--V------KE-H---L---T-------QG---N---------------L-----ASM--K--A---
-K----K---Y-------------LKKNL--GK--AEILIIDEISM--LHKNQLNL-VDRVLRYFK-----------------D---------N-----------------------Q-
--------------------------DPFGG-----IQVVLSGDFFQLPPIG------------------KY---N--E-KS-------------------------------------
-------------------------RD-----K-----------FSFMS-------------------EAWVN----AN--F-NVCYLTEQYR----Q-S-----------D--S----
---------SLNDILNEIRTG---N-V-S------Q---------------QN------------------------------LQ--IL-KE--AT---------EHTL-----EK---
-----------------------------------------------------------------------------------------------K---------E-V-----------
P-------------------TKLFTHNTDVDKINTEHLV-EL----E---G---------------------------R-----------------------T-K-T-F----------
----------K---------------------ATA------K-----------------------G--------------------N---------------------------I-K--
-----L---------------------------------------------------------I-D--------------T--------------------LK-----------N-S--
------VLA----------SENL-----QLKIGAKVMFVKNNS----------------E--K-----G-----FVNGTLGKVTGF

>WP_096064617.1_Psychrobacter_sp._FDAARGOS_221

---------KQST-ALDIL------K---------------------------------TG-----KNVFLTGSAGSGKTYTLNQYIHY---L-R--A--R------------------
-----R-VPVATTASTGIAATHM-------NG------TTIHSW-SGI---G--I------KD-E---L---T-------ER---D---------------L-----SNL--S--R---
-K----K---I-------------LKDRL--QG--TSVLIIDEISM--LHAKQLNL-VNQVLKHIR-----------------Q---------S-----------------------D-
--------------------------KPFGG-----IQLVAAGDFFQLPPVG------------------SR---G--E-SN-------------------------------------
-------------------------RD-----K-----------FAFMS-------------------EAWLD----AG--F-KVCYLTEQHR----Q-Q-----------A--D-DQ-
-AKDVQQQITLDAILNQIRGD---QGV-T------A---------------EA------------------------------IL--AL-QN--TF---------YQDV----------
-----------------------------------------------------------------------------------------------D---------V-N-----------
R-------------------TRLYTHNVNVNKINENELA-QL----S---G---------------------------E-----------------------T-V-T-Y----------
----------H---------------------AIA------H-----------------------G--------------------D---------------------------N-K--
-----L---------------------------------------------------------V-E--------------T--------------------LK-----------K-S--
------VRT----------SDEL-----TLKIGAKVMFIKNNT----------------E--L-----G-----VSNGTMGELVGF

>ONG38169.1_Enhydrobacter_sp._H5

---------KQAT-ALDIL------K---------------------------------TG-----KNVFLTGSAGAGKTYTINQYLHY---L-R--A--R------------------
-----D-VAVAVTASTGIAATHM-------NG------MTIHSW-AGI---G--I------SN-E---L---T-------AK---D---------------I-----ARI--K--K---
-R----T---V-------------VVERI--ER--TKVLVIDEISM--LHRQQFEL-INQVLQAIK-----------------E---------N-----------------------T-
--------------------------LPFGG-----IQLLVAGDFFQLPPIG------------------EP---H--E-SN-------------------------------------
-------------------------RD-----K-----------FAFMA-------------------QAWLD----AD--F-QICYLSEQHR----Q-K-----------T--D-
KTAVAGNTYYGLDLNAILNQIRSQ---Q-F-T------P---------------HI------------------------------MP--AL-TA--TA---------EHVL-------
--------------------------------------------------------------------------------------------------A---------D-N--------
---R-------------------TRLFTHNVNVQAINEQELG-KL----T---T---------------------------A-----------------------A-H-T-F-------
-------------R---------------------AWG------E-----------------------G--------------------D---------------------------E-
K-------L---------------------------------------------------------V-E--------------T--------------------LK-----------K-
S--------VRN----------TPEL-----VLKLGAKVMFIKNNT----------------E--L-----N-----VSNGTMGKVVDF

>Aalborg_AAW1

---------NQSL-ALSLL------K---------------------------------SG-----RNVFLTGQAGAGKTYVINQYIQW---L-R--S--C------------------
-----D-IPVAITASTGIAATHI-------GG------VTIHSR-AGI---G--I------KD-R---L---T-------DH---D---------------M-----ELI--Q--Q---
-K----E---H-------------LHKNI--TK--AKVLIIDEISM--ISANTLDM-VDRVVQMIR-----------------R---------D-----------------------G-
--------------------------RPFGG-----LQVILVGDFFQLPPVM------------------SS---Q--D-ANN------------------------------------
-------------------------TK-----R-----------FAFAA-------------------KAWKE----LN--L-AICYLHTQHR----Q-D-----------E--G----
---------DFSIVLNELRKG---Q-A-S------Q---------------ES------------------------------IA--LL-RT--RM---------DAKI-----T----
-----------------------------------------------------------------------------------------------H---------T-N-----------
P-------------------VKLYTHNIDVDRINDEKLE-EL----T---G---------------------------D-----------------------E-K-S-Y----------
----------I---------------------ATG------A-----------------------G--------------------D---------------------------K-K--
-----L---------------------------------------------------------L-D--------------T--------------------IK-----------K-S--
------MLA----------PEVL-----YLKVGAQVLFVKNNP----------------V--K-----G-----YYNGTTGEVVGF

>WP_033523136.1_Bifidobacterium_merycicum

---------QQSE-ALAIL------N---------------------------------VG-----ANVFLTGAPGAGKTYVLNEFVRA---A-R--A--E------------------
-----G-ANVAVTASTGIAATHI-------NG------QTIHSW-SGI---G--L------AT-S---L---S-------DR---L---------------F-----KTI--R--M---
-R----R------------------KRKL--QA--ADILIIDEVSM--MHAWLFDM-VDQVCRRIR-----------------L---------D-----------------------P-
--------------------------RPFGG-----LQVVVCGDFFQLPPVS------------------TS---N--R-NHD-------------------------LIAPTP----
EF--VASRERYASL---------GKDPE-----G-----------FITES-------------------LVWDE----LD--C-TVCYLTEQHR----Q-D-----------D--G---
----------RLLGVLTDIRQG---N-V-N------D---------------DD------------------------------RA--AL-AT--RL---------GVLP-----EP--
------------------------------------------------------------------------------------------------G---------Q-Q----------
-A-------------------VNLFPVNKQADTLNDMRLF-EI----P---E---------------------------E-----------------------P-H-E-Y---------
-----------V---------------------ATA------A-----------------------G--------------------P---------------------------A-N-
------L---------------------------------------------------------V-E--------------R--------------------LK-----------R-N-
-------MLA----------PERL-----QLKTGAAVMAVRNDQ----------------N--H-----Q-----FVNGSLGTVRAF

>YP_009595951.1_Acinetobacter_phage_vB_AbaM_ME3
```

```
---------NQDT-ALKVM------K--------------------------------SG-----ANVFLTGKAGSGKSHTIRQFLEY---H-R--Q--K------------------
-----E-TNVAITASSGIAATVV-------GG------STIHSY-IGM---G--I------KS-R---I---S-------NA---D--------------L-----MDI--R--K---
-R----R---G-------------MTAKL--KA--LEVLIIDEISM--LHKDQLSS-VDFILRSLR----------------R---------D-----------------------P-
--------------------------RPFGG-----VQIIVVGDFNQLPPVG-----------------PE---D--I----------------------------------------
-------------------------NA-----R-----------LCFMS------------------SAWVS----AE--F-KICYLTTTYR----Q-E-----------N--G----
---------ELLEILNAIREG---T-I-T------Q--------------EH------------------------------KN--KI-IN--TV---------DNDL----------
----------------------------------------------------------------------------------------------D---------D-L-----------
P-------------------SQLYTHNASVDYINERELN-LI----E---G--------------------------K-----------------------A-R-I-Y----------
----------K---------------------AIT------S----------------------G--------------------S---------------------------E-A--
-----N--------------------------------------------------------V-K--------------F--------------------LC-----------E-N--
------VLS----------PPVL-----TLKVGAKVLFTKNDP---------------Q--G-----N-----FVNGTLGIVTRL
```

```
>DAJ82417.1_Podoviridae_sp.
```

```
---------EQEE-ALGIM------L--------------------------------DG-----NSVILAGSGGSGKSHTLRQFIER---N-R--L--L------------------
-----G-RKTAVTATTGLAASHI-------NG------QTLHSW-ARV---G--L------GK-E---L---P-------DD---W--------------Q-----FTI--S--K---
-K----------------------KRKEF--QT--TATLVIDEVSM--MPDFVFDM-LDTVLRWAR----------------N---------D-----------------------D-
--------------------------RPFGG-----IQLILCGDFYQLPPVE-----------------G------------------------------------------------
--------------------------------K-----------FITNS------------------RVWNE----LN--I-RSCYLTKVYR----Q-K-----------D--D----
---------RLRDLLEGVRGG---N-L-F------K--------------RH------------------------------IA--YI-QS--RM---------VKPD----------
--------------------------------------------------------------------------------------------------------R-Q-----------
V-------------------PRLYSLNRKVDSENAHQLS-RL----K---G--------------------------D-----------------------S-I-F-Y----------
----------M---------------------MTE------K----------------------G--------------------D---------------------------I-N--
-----I--------------------------------------------------------I-N--------------G--------------------LK-----------G-S--
------IQS----------PELL-----ELKVGAPVIATKNNS---------------E--G-----L-----YHNGSLGKVIAL
```

```
>Rickettsiales
```

```
-----NLSPDQQK-ALDLI------K--------------------------------EG-----HHILITGPGGTGKSFLVNLIQKQ---L-P------------------------
--------NTALTATTGIAAVNI-------GG------RTYHSW-SGM---G--L------GK-E---T---V-------DE---L--------------V-----EKV--L--K---
-S----RWCEA-------------QRKTI--KQ--TKHLIIDEISM--MGADHFVK-LDQICRAVR----------------Q---------S-----------------------D-
--------------------------EPFGG-----IQLIMFGDFLQLPPVK-----------------D------------------------------------------------
--------------------------------E-----------YVFTS------------------SLWDY----LL--P-VVIVLTTIHR----Q-K-----------D--K----
---------EFAELLHRVREG---L-H-T------K--------------ED------------------------------MV--FL-EA--CG---------SKKL----------
---------------------------------------------------------------------E-QNL------------------------------D-D-----------
Q-------------------LILHSHNDAVDQFNKKMLN-DI----Y---S--------------------------D-----------------------E-Y-V-Y----------
----------N---------------------AND------T----------------------G--------------------K---------------------------G-A--
-----P--------------------------------------------------------L-K--------------A--------------------LQ-----------R-D--
------CIT----------PAEL-----TLKVGARVMLTKNLG------------------N-----G-----LCNGSLGTVVRL
```

```
>MBL6664806.1_Rickettsiales_bacterium
```

```
-----ELSPDQQD-VIRAF------E--------------------------------SG-----YNIFVTGSAGSGKSHLLNYLKRY---Y-S--H---------------------
-------QGLEITASTGIAAVNI-------GG------STIHSW-SAI---G--V------AN-L---P---V-------DK---I--------------I-----ANL--F--G---
-A----KF-SK-------------IRRRI--KR--TKALAIDEISM--ISSETLEI-LDRVFKSIR----------------E---------N-----------------------D-
--------------------------APMGG-----LQILFFGDFLQLPPIA-----------------KF---N--S----------------------------------------
-------------------------QI-----N-----------FCFES------------------NCWNE----LD--L-KTFNLKEIFR----Q-K-----------D--R----
---------KFINILNNIRKG---E-L-N------E--------------EN------------------------------IA--DL-QK--RV---------GLID----------
---------------------------------------------------------------------K-NQ---------------------A---------I-K-----------
P-------------------TILTTHNYKVDKINEEKIK-HI----P---K--------------------------S-----------------------E-Q-V-Y----------
----------K---------------------AEY------F----------------------G--------------------V---------------------------Q-S--
-----K--------------------------------------------------------I-D--------------F--------------------LK-----------K-N--
------SIV----------PEFL-----QLKIGAQVMMIKNTY----------Q----K--E-----G-----IINGSLGIIKDF
```

```
>NBR95534.1_Proteobacteria_bacterium
```

```
-----ELSNLQQN-AVNYF------L--------------------------------LG-----ENVFVSGGAGCGKSYLINFLKNN---Y-S--Q---------------------
-------LGLEITASTGIAAVNI-------GG------STIHSW-AGI---G--L------AN-Q---P---L-------EH---I--------------L-----ENL--N--S---
-F----KF-SK-------------IKQRI--RA--TNCLIIDEISM--ISAEVLDL-LNKVLQNIR----------------K---------N-----------------------Q-
--------------------------KPMGG-----LQILLFGDFLQLPPVG-----------------NH---K--D----------------------------------------
-------------------------GA-----K-----------YCFDS------------------QVWQD----LN--L-KNIILNQSFR----Q-S-----------D--A----
---------KFVEVLNHIRFG---N-I-N------D--------------EV------------------------------KQ--LL-TA--RI---------AVYD----------
---------------------------------------------------------------------N-SP---------------------A---------I-K-----------
P-------------------TVLTTHNHRADEINQQFLQ-QI----N---G--------------------------E-----------------------A-K-D-F----------
----------S---------------------ATY------K----------------------G--------------------N---------------------------E-N--
-----K--------------------------------------------------------I-V--------------F--------------------LK-----------K-N--
------CLA----------YENL-----TLKIGAQVMMIKNSL----------Q----K--E-----G-----VVNGSIGIVKDF
```

```
>MBN8828841.1_Sphingobacteriia_bacterium
```

```
-----GLSFDQLQ-VLEAI------R--------------------------------NG-----RNVFITGHAGTGKSYLLKCIRDL---Y-Y------------------------
-----G-KGLHITASTGIAAVQI-------GG------HTLHSW-AGL---G--N------GQ-A---N---V-------EY---L--------------I-----DYI--L--S---
-G----KG-TY-------------VRRKI--KN--CKMLAIDEISM--LPGDIFNK-LNTVLKAVK----------------N---------S-----------------------P-
--------------------------KPFGG-----IQLILSGDFFQLPPVT-----------------KD---N--E----------------------------------------
-------------------------TP-----I-----------FCFET------------------RAWQE----GQ--I-TTFCLQKIFR----H-S-----------E--Q----
---------LFIDFLSNLRKG---R-L-N------D--------------ND------------------------------VA--LI-KS--RT---------ITR-----------
-----------------------------------------------------------------------PN---------------------N---------I-T-----------
P-------------------TFLATHNYQIEQINNTHLK-SL----S---S--------------------------K-----------------------S-F-I-Y----------
----------E---------------------MSS------Q----------------------G--------------------D---------------------------E-K--
-----K--------------------------------------------------------S-E--------------F--------------------LA-----------N-N--
------CIA----------PKVL-----ELKIGALVMMLKNNY----------Y----K--D-----G-----IINGSIGKIIDF
```

```
>MBQ7287145.1_Candidatus_Gastranaerophilales_bacterium
```

```
-----DDDENFLR-IMHLI------K--------------------------------TR-----KNIFITGHAGTGKSYLLNKIKEN---V-P------------------------
--------NLVITSTTGIAAVNV-------KG------QTLHSW-AGV---G--I------CN-K---T---V-------EQ---T--------------V-----EKI--L--T---
-K----S---S-------------IKKQI--QK--CKILAIDEISM--LDIKTFEF-VNEVLKQVR----------------S---------C-----------------------D-
--------------------------EPMGG-----IQVIFIGDFFQLPPVE-----------------KD---T--D-K--------------------------------------
-------------------------EE-----K-----------YCFES------------------KLWQE----LD--L-QTILLKKSYR----Q-N-----------E--E----
---------NFIKALANMRTN---S-L-T------K--------------DD------------------------------VN--LL-KT--RE---------FEKS-----SIL--
```

```
----------------------------------------------------------------------------------------------------------D-N-----------
V--------------------LHIFATNLEADNYNNLKFK-SV----N---S--------------------------K------------------------E-Y-KLF----------
---------------------------------AID------GVY-------------------KG---------------------EKL----------------VETPTNAKEE-N--
-----I---------------------------------------------------------L-K--------------R---------------------ID-----------V-V--
------CSA----------EKSI-----SLKIGARVMLLVNLD-----------F----D--K-----G-----LINGSCGNVKEI

>MBE7709962.1_Cyanobacteria_bacterium_SIG32

-----DNSLVIKN-IIRLI------E---------------------------------NK-----HNVFITGHAGTGKSYILEKLKSR---F-R-------------------------
--------KMVVTSTTGIAAVNV-------KG------QTLHSW-AGV---G--I------CK-V---P---V-------DI---T----------------I-----QHI--L-SS---
-R----S---E-------------VVKRI--KK--TSLLAIDEISM--LKKDTLEY-VDKVLKAIK-----------------D---------S------------------------D-
--------------------------KPFGG-----IQVVFIGDFFQLPPVE------------------DE---Y--K-N---------------------------------------
-------------------------DE-----L------------YCFES------------------ELWEK----FN--F-KNVVLTENYR----Q-H------------E--E----
---------DFITALANMRKN---C-L-T------T---------------QD-------------------------------VE--LL-KS--RI---------IFDA-----DSY--
----------------------------------------------------------------------------------------------------------K-N-----------
V--------------------LHIFSTNKETDLYNEINFN-SL----Q---T--------------------------P------------------------I-Y-E-F----------
----------A----------------------ARD------GVM-------------------KG---------------------EKF----------------EYETLTEKDV-K--
-----I---------------------------------------------------------L-D--------------I---------------------LD-----------K-N--
------CKV----------NKHI-----KLRQGCRVMLVMNLS-----------F----N--E-----G-----LINGSCGTVDKI

>QBK85639.1_Marseillevirus_LCMAC101

-----IPDKKFQD-VLTAI------D---------------------------------NG-----QNVILYGPGGVGKTVALREIAAY---L-Q--E--K-------------------
-----G-KNIGVTATTGVAAINLNIPERKIRG------RTLHSW-AGV---G--L------GD-N---V---A-------AK---L----------------A-----AKI--M--C---
-Q----P---R-------------AKERW--LT--TDILIIDEVSM--LGGDFFDK-LDYIGRTLR-----------------Q---------R------------------------E-
-------------------------MDPIGG-----LQLILSGDFLQLPPVK------------------D-------------------------------------------------
--------------------------------E------------FCFQS------------------LAWKE----LA--L-APFIFLDPKR----Y-D------------D--V----
---------EYFQLLLRVRDG---E-P-T------M---------------ED-------------------------------IK--CL-YA--RV---------QAYE-----
HFCKMMED----------------------------------------------------------------CSDE-TK----------------------I---------I-K------
-----P-------------------TKAFSHKAKVEYTNDKELE-KL----P---G----------------------------E-----------------------T-F-D-F-----
---------------N----------------------CMD------SLKKYT----------------KN-------------------FKK---------------------------
D-Y-------Y---------------------------------------------------------L-R--------------Q---------------------LE-----------
-----------DAA----------PQQI-----SLKVGAQVMLKCNMS-----------V----E--Q-----G-----LVNGSRGVITEI

>QBK86258.1_Marseillevirus_LCMAC102

----------EHS-IFTAL------D---------------------------------NH-----ENIILHGPGGTGKTTVLKKIASH---A-Q--D--N-------------------
-----N-KIVCCTATTGVAAINLNVPEKKIAA------STLHRW-AGV---G--L------AQ-G---V---V-------DK---L----------------Y-----TKV--Y--H---
-D----E---L-------------ARKRW--LK--TDVLIVDEISM--LGADLIEK-LDFIGRKIR-----------------N---------N------------------------Q-
-------------------------EVSFGG-----LQLVFSGDFLQLPPVK------------------D-------------------------------------------------
--------------------------------K------------WAFQS------------------FAWKE----II--F-VPFIFTEPKR----Y-D------------N--Q----
---------DYFQLLLRIREG---K-H-T------I---------------ED-------------------------------LK--KL-RN--RV---------RSYE-----
KLFSILDD-----------------------------------------------------------------TKT-LD----------------------V---------I-R------
-----P-------------------TILHSLRVDVDSHNEKELA-KL----P---D----------------------------K-----------------------T-H-E-F-----
---------------I----------------------ADD------TFSASN----------------NN-------------------VKS---------------------------
D-Y-------Y---------------------------------------------------------I-R--------------L---------------------LD-----------
-----------EAI----------PKAI-----ALKVGAQVMLKCNLD-----------V----K--G-----G-----LVNGSRGVILKI

>QBK87070.1_Marseillevirus_LCMAC103

------FFSQYEC-VLEAI------H---------------------------------NK-----WNILLHGPGGCGKTYTIAKLINA---L-T--I--ANP-----------------
-----D-AVIACTALTGVAAANL-------RGSIPVDAQTLHRW-AGV---Q--L------AH-G---P---A-------DQ---L----------------V-----RKL--Y--R---
-N----R---E-------------ALDRW--RT--TDILFVDEISM--LGKELFEK-FDYIARSVR-----------------T---------M------------------------R-
--------------------------KPFGG-----IQLVLSGDFLQLPPVD------------------D-------------------------------------------------
--------------------------------E------------WVFTS------------------RRWAE----LDLVP-YIFEDGVGKR----Y-D------------D--P----
---------AFFAMLLRARVG---K-L-T------A---------------DD-------------------------------AR--RL-AA--RD---------QAYR-----
DYLSEEEA----------------------------------------------------------------QPGR-AE----------------------S---------V-K------
-----P-------------------TLLFPTNRDADIHNSGKLA-EL----D---T----------------------------P-----------------------V-R-T-Y-----
---------------A----------------------AAD------VVNVHP----------------RA---------------------R---------------------------
Q-A-------T---------------------------------------------------------L-D--------------Q---------------------LD-----------
-----------KII----------PARI-----DLRVGAQVMLRANLD-----------V----A--A-----G-----LTNGSRGVVVDL

>BCU09408.1_Sicyoidochytrium_minutum_DNA_virus

-----PLNQEQAF-ALDLV------K---------------------------------RG-----KNVFITGVAGTGKSFTVARIVEW---A-E--K--I-------------------
-----G-KKIDVTASTGLAAFLL-------SGPCKKYCSTFHSW-AGI---G--L------GK-G---N---A-------DK---L----------------A-----RNM--L--S---
-K----V---D-------------VCAHL--RE--VDIVVIDEISM--MNAEYMAK-VDVVMKAVR-----------------K---------N-----------------------PR-
--------------------------SPFGG-----IQIIFCGDFGQLPPVR------------------RD---G--S-----------------------------------------
-------------------------AP-----I------------YLFEH------------------PIWKD----TV--D-HCVLLKKVYR----Q-E------------Q--E----
---------EFVDILTRMRDG---E-T-T------E---------------ED-------------------------------LK--LL-RE--TG---------PVE-----------
----------------------------------------------------------------------E-IN----------------------G---------V-R-----------
P--------------------TVLYSRNRDVDYMNHLELS-RL----P---G--------------------------E------------------------S-K-V-Y----------
----------N----------------------AND------IFK--------------------H---------------------P---------------------------K-A--
-----N---------------------------------------------------------Q-E--------------F---------------------VK-----------K-K--
------FSL----------PETL-----ELKPGAQVMLLMNYM-----------P----G--A-----G-----LVNGSRGVVTEL

>ADX05998.1_Organic_Lake_phycodnavirus_1

-----DFSSTQQI-AYDHY------L---------------------------------NG-----DNVFITGPGGTGKSYFIKKVYEN---A-K--K--R-------------------
-----N-LNVSVTAMTGCAALLL-----DCNA------KTIHSW-GSI---G--L------GT-D---P---I-------EM---I----------------Q-----SRI--V--K---
-Y----R-----------------KRDIW--LN--TDILIIDEVSM--MSCELFEL-LFKIAQHFR-----------------R---------N------------------------K-
--------------------------KPFGG-----IQVIFSGDFHQLPPVT------------------KD---S--------------------------------------------
--------------------------------A------------FCFES------------------LLWEE----CF--K-HSVILKENFR----Q-T-----------SD--P----
---------VYQVILNEIREG---V-I-S------K---------------QS-------------------------------KD--IL-NT--CL---------NKPQ----------
------------------------------------------------------------------------LD----------------------N---------V-S-----------
P--------------------TLLYPVKRLSEQVNISEHV-CL----E---G--------------------------K------------------------E-H-I-F----------
----------K----------------------MKY------IEP-------------------PN---------------------KKI-------------------------E-E--
-----E---------------------------------------------------------L---------------IK---------------------QK-----------K-N--
------MIV----------DESL-----RLKVGSQVMCIINLD-----------Q----D--N-----G-----IVNGSQGKVVGF

>ADX06411.1_Organic_Lake_phycodnavirus_2
```

```
-----TFSESQQS-AYKNY------L--------------------------------KG-----ENVFITGPGGTGKSYFIKKVYED---A-K--S--R------------------
-----G-LNVSVTAMTGCAALLL-----DCNA------KTIHSW-GSI---G--L------GT-E---P---I-------ES---I--------------K-----QKI--V--K---
-Y----R-----------------KRDVW--IK--TDLLIIDEVSM--LSCELFEL-LYRIAQDFR----------------R---------S-----------------------E-
--------------------------KPFGN-----MQLIFSGDFHQLPPVS-----------------KD---S-------------------------------------------
--------------------------------K-----------FCFES------------------PFWNG----CF--Q-HKIVLKENFR----Q-K----------GD--K----
---------VYQTILNEIREG---N-I-S------E--------------ES------------------------------KD--IL-RS--CL---------NKKN----------
-----------------------------------------------------------------------NE---------------------H---------L-S-----------
P-------------------TLLYPVKRLSEQVNLFENI-SL----K---G--------------------------E-----------------------E-K-L-Y----------
----------K---------------------MKY------IIQ-------------------PT--------------------KKI-------------------------E-Q--
-----E--------------------------------------------------------L---------------IK--------------------QK-----------R-N--
------LIV----------DEEL-----RLKIGSQVMCAVNLD----------Q----E--Q-----G-----IINGSQGKVIGF
>YP_009173733.1_Chrysochromulina_ericina_virus
-----NLNNQQQD-IFDKY------L--------------------------------KG-----ENIFITGPGGTGKTYLIKAIVED---A-K--K--N------------------
-----N-KAYHVCALTGCAAILL-----QCGA------TTLHGF-SGI---G--L------AS-G---T---I-------SQ---V--------------V-----DRV--V--K---
-N----R---Y-------------KKPNW--AK--TELLIVDEVSM--LSLKIFTI-IDLIAKRVK----------------R---------Q----------------------RD-
--------------------------IPFGG-----MQIIFAGDFYQLPPVG-----------------DE---E--E-IE-------------------------------------
-------------------------TT-----Q-----------FCFES------------------PLWNE----VFPSS-NQIVLETIFR----Q-T-----------D--N----
---------NYAKILNKLRVG---E-I-T------K--------------NG------------------------------IK--AL-EQ--CV---------NKKF----------
-----------------------------------------------------------------------ND---------------------E---------L-N-----------
P-------------------TILLPRRKDVDNINIKEYN-KL----D---K-------------------------IS-----------------------E-K-T-YT---------
---------MK---------------------PVD------MLDLPL--------------------------------------SKEH------------IQNITLFTDTERQ-Q--
-----E--------------------------------------------------------I-D--------------Y--------------------LA-----------D-N--
------LMA----------EKTL-----NLRIGTIVMCISNLD----------V----E--A-----G-----IINGSQGIVVDF
>YP_008052747.1_Phaeocystis_globosa_virus
-----QLNAEQEL-IFQKY------K--------------------------------NG-----ENIFVTGPAGSGKSFLIKTIVND---S-V--E--N------------------
-----D-YNLQVCALTGCASILL-----NCKA------TTLHRF-AGI---G--L------AN-K---S---I-------DA---V--------------V-----EDV--F--E---
-K----R---Y-------------KLKKW--YD--LKCLIIDEVSM--MSLKILLI-LDKMARKIYK---------------K---------E-----------------------N-
--------------------------TPFGG-----LQVIFSGDFYQLPPIK-----------------SN--DG--D-KE-------------------------------------
-------------------------SS-----M-----------FCFED------------------PLWNQ----LFPAD-NQILLKSIFR----Q-D-----------E--K----
---------EFLKVLKYVREG---R-I-T------K--------------ST------------------------------RE--TL-EK--RV---------FTEA-----EI---
-------------------------------------------------------------------DKV-RE---------------------E---------N-V-----------
V-------------------TIISPYKKDTDNINAAAYK-ML----S---N------------------------DVE-----------------------K-K-M-Y----------
----------S---------------------IKY---------------------------LKG--------------------SRK-----QDGAVESAVNNLLIDSNASLK-A--
-----D--------------------------------------------------------Y-E--------------F--------------------LA-----------N-N--
------IMA----------NTSL-----ELKIGTHVMCIANIS----------LE--SE--I-----Q-----LANGSQGVVVGF
>ANS04235.1_uncultured_Mediterranean_phage
-----QLSSEQQE-VLGLV------R--------------------------------QG-----LNVFISGPGGTGKSYLIKLICEL---Y-R--D--K------------------
-----I---VQVCALTGCAAELL-----GCGA------RTIHSW-SGT---G--M------SR-G---D---K-------YR---I--------------I-----NRV--C--S---
-K----K---K-------------NRGAW--KK--VDILIIDEVSM--MSVKYFEL-LDEIGKTIR----------------N---------S-----------------------T-
--------------------------QPFGG-----IQLIFSGDFHQLPPIG-----------------DE---S--E-PD-------------------------------------
-------------------------TC-----K-----------FCFES------------------ERWKT----TF--L-NVVLLTHIFR----Q-S-----------D--K----
---------TFTKILRQVRKG---G-I-T------Q--------------KT------------------------------HD--IL-NT--RL---------MKKS-----N----
------------------------------------------------------------------KLSY-GT---------------------G---------R-K-----------
P-------------------TIISPIRKEVKSVNDRNMS-RL----D---S--------------------------E-----------------------L-L-T-Y----------
----------E---------------------YQI------V----------------------K--------------------DKD------YKPPTIVVNDVTKIDQKLID-Y--
-----E--------------------------------------------------------I-N--------------Q--------------------LK-----------Q-R--
------MNG----------ELSL-----ELKLGAHVMCVANLD----------ME--GK--Q-----Q-----IVNGSQGIIEDI
>QPI16828.1_Virus_NIOZUU159
-----SLNTKQRE-AVDAV------L--------------------------------NG-----RNILITGPGGTGKSFTIKYITEL---L-N--K--N------------------
-----N-KYYGLTATTGTASVLI-------GG------QTINSY-LGI---G--L------GN-D---K---V-------SD---I--------------I-----KNI--I--T---
-N----K---N-------------IRERI--VK--LEVLIIDEISI--LEDKLFEK-ISEILSTIR----------------G---------QFID----------------KKLAE-
--------------------------KPFGG-----IQMIFVGDFCQLAPVK-----------------G------------------------------------------------
--------------------------------L-----------YCFLS------------------KIWEK----SE--V-DIIVLEELVR----Q-T----------GD--Q----
---------LFQKILGIVRKG---K-C-T------D--------------NI------------------------------IK--VL-ER--LK---------DTQF----------
-----------------------------------------------------------------------SD---------------------N---------I-I-----------
P-------------------TKLYPVNIDVNKINNIEIA-KL----K---E--------------------------KG--------------------YKS-S-L-Y----------
----------K---------------------ATC------S---------------------KG-----------------------------------------------------
----------------------------------------------------------------N------------------------------------------------E-K--
------AAL----------NYDI-----ELTENAQIIITRNID----------I----S--Q-----G-----LINGTRGVIKHL
>YP_009010863.1_Invertebrate_iridescent_virus_22
-----IPNPEQQY-TLRLI------E--------------------------------EG-----KNIFINAPAGTGKSALIKYFWQQN--F----N--K------------------
-------KVLGLTSTTGISALNI-------GG------STLHSF-LGI---G--L------GK-E---N---V-------DD---L--------------Y-----DKI--I--K---
-N----R---E-------------KHELW--LK--LDLLIIDEISM--LHPELFNK-LEKVARLVR----------------E---------N-----------------------K-
--------------------------KKFGG-----IQLIVTGDLFQLPPVS-----------------QD---S--T----------------------------------------
--------------------------------------------LIINS------------------PKFNK----CI--D-TIVEFRNIIR----Q-I-----------D--P----
---------IFKNILNKIRIG---I-V-D------A--------------QV------------------------------KK--LL-KK--RF---------IKAP-----KQ---
---------------------------------------------------------------------P-DI---------------------Q---------I-K-----------
P-------------------TKLYCTRKSVDHLNENELN-KL----A---N--------------------------K-----------------------G-Y-T-F----------
----------R---------------------EYIM-----EFV------------------NQN-------------------------------------------CPISFD-Y--
-----I--------------------------------------------------------I-K--------------N--------------------FV-----------K-N--
------STT----------PSTL-----QICEQTQVMLTYNIS------------------P-----T-----LVNGSRGIVTGF
>QNH08436.1_Invertebrate_iridescent_virus_Kaz2018
-----KLNKQQSR-ALALM------C--------------------------------QD-----KNIFITAPAGAGKTLLINHYCDY---V-R--Q--HEPF---------------
-------KKIAITSTTGVSAILI-------GG------STLHSY-LGI---G--L------GY-G---T---I-------EE---L--------------V-----QRI--K--K---
-A----S---K-----------GIKERVW--KE--LTTLIIDEVSM--LNPVLFDK-LEKIARIIR----------------G---------S-----------------------N-
--------------------------LPFGG-----IQLILSGDLLQLPVVK-----------------GA---G--A-GNK------------------------------------
----------------------NDHNM-----E-----------FVTDA------------------NSWKK----CI-GN-NIVLLTEIMR----Q-K-----------D--F----
---------HFKEILLKIRVG---N-I-D------K--------------QV------------------------------RS--VL-SQ--HM---------KKYS-----KL---
```

```
---------------------------------------------------------------------K-KE---------------------E---------I-Q-----------
P-------------------TRLFCLKKYVQDLNDSELK-KL----E---D--------------------------S-----------------------G-K-K-F----------
----------I---------------------NFN------ALVKKY--SEEAIL------TNKG--------------------RSRC-------------------TDLQFK----
-----F--------------------------------------------------------LSD--------------R--------------------FV-----------K-D--
------STT----------PQHL-----RVCEGAQVMLTYNID----------Q----L--S-----G-----LVNGSRGVIIGF
>YP_009046811.1_Armadillidium_vulgare_iridescent_virus
-----KLNNKQEK-AYKMM------V--------------------------------AG-----ENIFITAPAGTGKTFLINYFCKT---I-D--P--I------------------
-------RTVAITSTTGVSSLLI-------GG------STLHSY-LGI---G--L------GD-G---T---T-------DQ---L--------------F-----HKI--V--N---
-C----S---K-----------GIKAAVW--RK--LQTLIIDEVSM--LSPILFDK-LECLARQIR----------------G---------N-----------------------N-
--------------------------KPFGG-----IQLILSGDLLQLPVVK-----------------GG---G--V-ADN------------------------------------
-----------------------GSPL-----D-----------FVTDA------------------SSWGR----CV-GN-NVVLLTEIMR----Q-K-----------D--P----
---------LFKEILLKIRVG---C-I-D------A--------------QV------------------------------KE--VL-NN--HI---------SKNV-----
SIGEYDGDYDEED--------------------YEEDYNPGVA------------------------PAPDVEAN-KEES-----------------ELK---------I-Q------
-----P-------------------TKLFCLKRYVKALNDKELQ-KL----E---N--------------------------A-----------------------G-V-K-F-----
---------------K---------------------NFN------ALIKAF--SSEEIN------SVKG--------------------RSKS-------------------
SEAQFN---------F--------------------------------------------------------LKD--------------R--------------------FV------
-----K-D--------CTT----------PQQL-----RICEGAQVMLTYNIN----------Q----P--M-----G-----LVNGSRGVITSF
>KAF0979914.1_Naegleria_fowleri
-----KLSDEQLN-VLKCA------I--------------------------------EG-----HSMFITGVAGTGKSFLLECIIKT---L-S--N--VH-----------------
-----Q-KKVVVTASTGIAAVNI-------GG------STIHSF-AGI---R--T-----LDN-G---Q---V-------DS------------------------KTAW--------
----------R-------------NDKEW--QS--TDVLIIDEISM--IDAQYFDQ-LEAVATEIRCFFEIATSKAPKEMVMK---------Q-----------------------L-
--------------------------PAFGG-----IQVILCGDFLQLPPVA-----------------KP---F--K-NEH------------------------------------
------GET-------------VYQKK-----E-----------MCFKA------------------KCWQK----II--K-YTFELTNVFR----Q-E-----------E--N----
---------EWVSILNSIRTC---R-I-D------S--------------NA------------------------------IS--QL-SK--LQ---------HNRF-----E----
----------------------------------------------------------------------------------------------G--------ME-K-----------
S-------------------TVIHTLNKNVDGVNESELL-KL----D---P--------------------------P-----------------------H-F-I-Y----------
----------K---------------------DHT------YFSYGE--YD------------PG-VD-----------------PPS------TK-----------------E-S--
-----I--------------------------------------------------------R-N----------AIL-S--------------------NF-----------N-S--
------SNA----------APEI-----NLRVGAQVMMIKNDF---------------T--N-----Q-----LVNGTRGEVIGF
>H_opuntiae1
-----ELSKEQSV-VYDLI------VK-------------------------------GG-----RNVFFTGPAGSGKTTLLKTIIHG---L-K--V--KHDAFE-------------
--DSKA-LRVGVTASTGLAAMNL-------KG------LTFHSF-LQI---G--L------GT-L---K---A-------EA---I--------------A-----KNL--L--S---
-D----I---N-------------FNLVW--NS--LRVLIIDECSL--INSKLFQK-LEKVARLVR----------------K---------N-----------------------H-
--------------------------KPFGG-----IQLVLVGDFYQLPPII-----------------ED---Y--D-ILA-------------------------KIGI-------
-----VKDKTDY---------HEFKRK-----R-----------FAFCS------------------PAWKK----CI--E-FELGLKEVHR----Q-K----------GD--P----
---------KFIEYLNQIRLG---N-V-T------K--------------EI------------------------------DQ--EM-QK--LT---------RELS-----P----
-----------------------------------------------------------------------IE---------------------G---------V-E-----------
P-------------------TYLFPTKFKANNYNLQQMN-KI----K---S--------------------------R-----------------------T-Y-R-Y----------
----------K---------------------AAL------D----------------------G--------------------K-L------KG-----------------T-N--
-----E--------------------------------------------------------F-A--------------K--------------------MV-----------E-A--
------CMF----------FKTL-----DLKVGSQVMLVKNNF---------------P--E-----G-----VINGTKGVVVGF
>XP_005717394.1_Chondrus_crispus
-----VADRYQAE-AIRAA------K--------------------------------QG-----ESFLLTGSAGTGKSFVLKHVISV---L-R--S--M------------------
-----G-KVVGVTASTGCAAVGI-------GG------GTIHSL-SGV---G--I------GM-D---P---I-------EK---L--------------V-----RKG--H--T---
-D----R---V-------------LRKRL--KQ--LDVLVIDEISM--IDSFLFDK-LNAIIAAAR-----------------C--P---PPK----------------------RD-
----SGITRGIRTLN--SGPGGLFTLKPFGG-----LQVILCGDFFQLPPVA-----------------AS---D--T-RFV------------------------------------
----------------------NSSEK-----F-----------FAFEA------------------KTWKR----II--K-NTYVLRVVHR----Q-A-----------D--R----
---------QFAGLLNEVRQG---V-V-S------E--------------ST------------------------------MQ--VL-NA--CL---------VNPL-----
KPLVE------------------------------------------------------------------VEEN-GR---------------------R---------V-A------
-----F-------------------TKLFSYRRQVASENSSQLK-KL----K---T--------------------------K-----------------------G-I-R-Y-----
---------------D---------------------AYD------QIH-----RSE-----------LG--------------------TLT-------------------------
A-R-------H--------------------------------------------------------V-Q--------------Q--------------------ML-----------
D-N--------TNC----------AQSI-----ELRRGCRVLCTKNLD----------T----G--L-----G-----IVNGAPGIVVGW
>PXF40737.1_Gracilariopsis_chorda
-----SLDRSQEI-VIAEA------I--------------------------------KG-----RSLFITGSAGTGKTFTLLKLIRT---L-R--S--Q------------------
-----G-KQVAVTASTGCAAVAI-------RG------STIHSF-LRL---G--M------GN-L---S---L-------SK---A--------------R-----AIT--D--A---
-N----V---S-------------FQRLL--QK--TDTLVVDEVSM--IEGHLFDL-MDVVCTTAR----------------KCN----S--S---------HKPGEYDVNLCNNTR-
--------------------------ATFGG-----LQIIVCGDFFQLPPVR-----------------SK---S--S----------------------------------------
-------------------------NL-----C-----------FAFES------------------AAWKE----TN--L-QVHVLPRAHR----Q-S-----------C--N----
---------SFVGMLGEVRRG---I-L-S------Q--------------YT------------------------------RR--VL-NA--SV---------IGTR-----
NLQPE------------------------------------------------------------------LSRK-GD---------------------M---------L-Q------
-----F-------------------TKLFPLRAQAQSENMYRLN-AL----P---G--------------------------F-----------------------T-V-R-Y-----
---------------K---------------------SQFF-----SIK--------------------G--------------------Q---------------------------
---------------------------------------------------------------------N--------------S--------------------QN-----------
E-F--------GSV----------ECLI-----DLKQGCPVLCTKNID----------E----S--K-----G-----LVNGTSGFVVGF
>G_muris
-----DLSFEQKL-LFRAA------VC-------------------------------DR-----RPLFFSGSAGTGKSHLLRAIISG---F-N--D--E------------------
----HP-DGLAVTASTGTAAVNI-------AG------CTIHSF-SGL---N--A------DTVG---D---P-------RQ---L--------------Q-----AQI--R--QM--
-R----K---A-------------ISERW--KA--TEVLIIDECSM--LQAEFFDA-LEQVAREKK----------------R---------R-----------------------T-
--------------------------SFFGG-----IQVILCGDFLQLPPVT-----------------KN---S--K----------------------------------------
-------------------------PF-----T-----------WLFES------------------SSFKE----IK----LKTSLVKSFR----Q-Q-----------D--P----
---------DFLTLLNELRVA---K-L-S------P--------------LS------------------------------KA--RL-NQ--RL---------VRQE-----
DIEAEQKREIER---------------------ARKNYEEKEAYLKEEIARLEVEVDE------------SLLAL-EK----VTTQESFRNIFFRLLSSNRE--LITTKVTK-
SVTFYTCPTFP-------------------VSLKTHKKDVESVNQQRLR-QT----K---E--------------------------T-----------------------I-F-N-
M--------------------L---------------------AKD------------------------------------------------------------------------
--------H------------------------------------------------------------------N--------------D--------------------TD-----
------K-Q--------EDP----------PKCI-----TLAIGAQVLITKNLD----------V----Q--K-----G-----ICNGSQGVVIGI
>EET02286.1_Giardia_intestinalis_ATCC_50581
```

```
-----NLSFEQKL-LFNAA---V--I--------------------------------RR-----KSLFFSGSAGTGKSHLLRAIIKG---L-S--R--LED----------------
-----D-EKVVVTAPTGTAAVNI-------SG------CTIQSF-AGFLDEN--L------SE-Q---K---F-------PD--M---------------L-----ARA--R--R---
-V----K---Q-------------TRKRW--ID--ADVLIIDECSM--LQGTYFDC-LEYVARNLR----------------GGS-------S-----------------------K-
--------------------------SFFGG-----IQLILCGDFLQLPPVV-----------------RS---N--N----------------------------------------
-------------------------PL-----V-----------WLFEA------------------KAFQL----IP----LKASLTHCFR----Q-S-----------D--K----
---------SFISMLNETRIG---C-V-S------P--------------QT------------------------------DS--LL-QS--LL---------IHES-----
ELNNEHDRISEK---------------------AYRDYSAEIKTLEEEGRKTEEARQF------------ALEKL-DDA----ADELSFRQSLLNLIHINRR--RCELEV-
KRIRTTYVSPSFP-------------------VRLYTHRQAVDEYNNSNLL-K---------G--------------------------K------------------------Q-V-
I-FR------------------LQ---------------------AVD----------------------------------------------------------------------
----------E------------------------------------------------------------------N-----------------------------------LTDH-
--------Q-H-------LLDP----------PPVI-----SISIGAQVIITKNID----------V----Q--R-----G-----LCNGRQCVVKDI
```

```
>A_deanei
```

```
-----EWTSEQRR-ATQLF------Q--------------------------------SG-----RNVFVTGAAGTGKTQWLLHLIRQ---VIP--N--S------------------
-GTVYQ-GGLAITGTTGAAARLI-------GG------TTVHSF-AGI---G--R------GE-G---T---V-------EA---L--------------L-----EKV--K--S---
-R----G---D-------------AMRAW--RA--CQVLIIDEIGM--LPAHIVTK-LDYIARHVR----------------K---------E----------------------LQ-
--------------------------KPFGG-----IQVVVVGDFLQLPPVA-----------------KG---G--E----------------------------------------
-------------------------EV-----K-----------AAFAS------------------ASWSD----AK--F-AAVEFTHTFRFGTSQ-S-----------N--K----
---------LFVQCLSHIRRG---L-Y-T------R--------------AV------------------------------HT--VL-TE--CL---------HRPL-----D----
---------------------------------------------------------------------E-RG---------------------G---------V-H-----------
P-------------------TVVMARRNDVETHNQIKLD-EL----E---D--------------------------P-----------------------Y-F-QRY----------
----------A---------------------SED------YAAY------------------PG-----------------------------------------------------
----------------------------------------------------------------D--------------S--------------------VD-----------S-E--
------VSL----------PAVL-----TLKLGAQVVLLVSLQ-----------G---Y--E-----G-----LTNGSLGVVMDF
```

```
>KAG5479457.1_Leishmania_martiniquensis
```

```
-----SWTREQQR-AMQLV------R--------------------------------AG-----HNVFVSGAAGTGKTEWLLHVLQHV--LPRTRQ--RQGLKSGAHPGAEEG--
KEEYAVDT-ARVAVTAATGIAARLI-------GG------KTVHSF-SGI---G--R------SE-G---D---P-------DV---I--------------L-----QRV--Q--S-
---R----P---D-------------IVRAW--QQ--CEVLVIDEISM--LSSRTFAL-LDRIARALR----------------A---------SMPPP------
ASSSQRRHPTNNAA---------------------------LPFGG-----IQLLVVGDFLQLPPVS-----------------RG---A--G-------------------------
---------------------------------------EEV-----Q-----------PAFMA------------------SAWRS----CN--F-QTLLFTKDYR----H-A----
------ED--P-------------RFAECCAAVRRG---E-C-T------P--------------LV------------------------------QE--VL-EA--CL--------
-GREL-----E-------------------------------------------------------------------------E-RF---------------------G--------
-V-E-----------A-------------------TTLLARRKDVDRYNAQRLQ-QL----E---S--------------------------M-----------------------Q-
F-HRY--------------------A---------------------SED------YAAV------------------PG--------------------------------------
-------------------------------------------------------------------------------A--------------N--------------------ID-
----------D-E--------VSL----------PPVL-----TLKVGAQVVLLASLP-----------N----E--P----S-----LANGNLGVVVGF
```

```
>CUF06097.1_Bodo_saltans
```

```
-----VLDASQQA-AVEAA------G--------------------------------RG-----ENLFVTGGAGTGKTLVVKRIVDS---L-R--A--A------------------
-----G-KTVAVTATTGVAALNC-------GG------TTLHHF-AGM---S--Q------SF-Q-DLP---P-------EE---C--------------A-----RRI--N--A---
-K----R---H-------------VVHRL--SK--TDVLVIDEISM--LEASTLEK-VHVAAQMAR----------------N---------N----------------------FF-
--------------------------KPFGG-----LQLIFCGDFLQLPPIS-----------------AR---G--Q----------------------------------------
-------------------------AI-----P-----------YPFFS------------------PVWQQ----LN--L-NVVTLATKFR----Q-Q----------SD--T----
---------SFQSVLDAVREA---K-L-E------Q--------------EH------------------------------ID--AL-QQ--CV---------RRHQ-----DA---
---------------------------------------------------------------------------------------------ID---------D-S-----------
Y-------------------VRLYGSNREVDAYNLQCFS-FL----SPRLG----------------------ELVTDD--------------------KPM-L-L-Y----------
----------N---------------------AMD---------------------------LKS--------------------S-----------------------------K--
-----A--------------------------------------------------------A-S-----------------------------------IN----------LN-D--
------GRL----------AQTI-----PLKIGTKAMLLTNLN-----------V---R--A-----G-----LVNGAVGVVTGF
```

```
>Phytomonas_sp
```

```
-----RLSAEQAQ-TLSLA------L--------------------------------NG-----ASLFVGGKAGTGKSFLLREIVHK---M-R--L--R------------------
-----G-IRVAVTASTGIAALNI-------GG------NTFHSV-FGV---P--V------YQ-D---D---EAVG----KRRT-LS-----------TRTPKAYEKKL---------
------T---Y-------------DEKVL--SQ--VDVILIDEISL--LHAGYLEA-LERAARGAK----------------G---------K---------------------NPS-
--------------------------KPFGG-----VQIILSGDFMQLTSFQFQRGGCSGRASASSNIINN---K--D-RLVCAQVAK-------------------ECDIAV----
QY--VAIKDALERV---HKDSAHKSVER-----RY--CVGYYCALPMYES------------------YAFRN--------FLLHVQLSESTR----H-R----------ID--A---
----------GFLQDLNLLRVG---I-L-T------Y--------------RL------------------------------SRSFVL-NR----------------------E---
-----------------------------------------------------------------------------------------------D---------D-S----------
-A-------------------IRLFATRRSVKAYNEQQIV-GL----N---G--------------------------R-----------------------E-V-V-F---------
-----------KSHLMLLGVGKGDSAGCFSSVEAQK-----CKYTKQKFWSDVILL---HFTNREGFSIRFGHGRGDLGSRRWARGKPRE----ITYSEVQSIVH-
EICQAKTLDSER---FFAYV--LPFAYCYSPNIHSVAVRTFGHNRKEATMQLKGFLASASERM-------
QGDASLKNSVAVGAEYMGFMFASAPRWMLIRFERFSAHKFASHLQPRFHHYFRYDIQ-N--------DLV----------TQSK-----KLKVGCRVMLLRNLN--------------
----A-----Q-----YVNGSLGTIVDF
```

```
>S_culicis
```

```
-----HLTSEQAT-VLRLA------L--------------------------------EG-----ASMYIGGKAGTGKSHLLRVISTE---L-R--N--K------------------
-----G-LCVLVTASTGVAALNI-------DG------NTFHST-FRV---P--VLAPVGASR-E---K---A-------EE---M-------------------DETT--E--SP--
-SH---R-DHQRHSTVL---Y---DTQVL--AQ--ADVIILDEVSL--LHAGYLES-LDEAARAAAGK--------------E---------K-----------------------D-
--------------------------KYFGG-----IQMVLSGDFLQLTAFD-----------------GA---A--Q----------GVGTSDRFVCQRVEPA-ESEQGLR-----
ENGKAPRDDACGANGN-ESDSEMSASSTVAVPVERLASPSYY-NLPMYSS------------------YCFHR----AL----LHVQLTKSAR---HQ-T-----------D--P---
----------VFLRELNELRVG---R-L-P------Y--------------RL------------------------------SRSAFL-NP--------------------------
----------------------------------------------------------------------------------------------YD---------P-T----------
-A-------------------IRLFAVKHAVKTFNDQKML-AL----P---GRFLAFPT----EVALLELTRGNEMGSE---RISHWSAITLLHFFF-----A-H-VRY--------
HFNSHDAHAL-VKRVLAS----------------AKD-----ASLAGLLA-ARFYVYVCACSPVVHG--------------------SAARVAVRFRGDTEKQAQHCHTQFVTLVE-
T-------YM--------------------------------RTHLAGAVDT--------KRTTAVTSS------GSGARRSKKTRSA--------AHADATLVK-----------Q-
EPMTMRQLLAAVLPSCQKEKPHSSDLLLQNKYLKVGCRVMLLRNLN------------------H-----K-----YVNGSLGTIEQF
```

```
>DAC81588.1_Hydra_MELD_virus
```

```
-----KLNKVQQE-IMAAF------D--------------------------------HG-----DNIFITGPGGVGKSMILKHIADT-----D--Q--L------------------
-----Y-KKIAVTATTGVAAHLI-------GG------MTIHSF-AGI---Q--R------GE-K---D---Y-------SY---Y--------------V-----KNM---------
-T----V---D-------------VKKRW--FE--TDVFIIDEASM--WTAKLFKL-VHEIACAAR----------------Q---------N----------------------DD-
--------------------------ELFGG-----ILLILSGDFYQFPPVE-----------------GG-----------------------------------------------
--------------------------------------------FIFTC------------------NLWLK----IE----KVFVLTESYR----Q-K----------DD--E----
```

```
---------AFFKTLNNVRVG---K-L-N------S--------------QD------------------------------VD--FL-MQ--QH---------RGHA-----SD---
---------------------------------------------------------------------------------------------IP---------K-T-----------
F-------------------PRLYFTNKKVDAYNQVMLN-SM----Q---T--------------------------E-----------------------E-K-L-F----------
----------R---------------------SVD------EIK--------------------------------------------------------------------------
--------------------------------------------------------------V-S--------------E--------------------VD-----------F-T--
------FQI----------PSET-----RIKVNALVMITKNID----------I----D--N-----G-----LCNGAMGKVVSF

>H_opuntiae2

-----KFNKEQQY-VIDLV------VN-------------------------------KQ-----ESIFLTGAAGTGKTVLLRELIER---L-K--Q--KHGVKN-------------
NSNKYE-YNVLVTATTGLAAYHI-------GG------QTYHSA-LGL---MD-L------NK-N---N---T-------GR------------------------QKI--K--L---
-N----A---A-------------KSNAW--KQ--CKVLIIDEVSM--MEASTLDF-IDKTAKEMR----------------K---------N-----------------------Y-
--------------------------SPFGG-----IQVILCGDFFQLPPVD-----------------NK---I--L-EGI----------------------------I-------
------KKYDELEG-A-IDPTLVNENKELPICE-----------YAFKS------------------KVWSH----GI--K-HCLSLSTVFR----QLD-----------D--P----
---------EFVKCLNELRLG---I-V-S------P--------------TT------------------------------EA--LM-KR--VE---------SKSYV-----A---
----------------------------------------------------------------------------------------------S---------S-D-----------
V-------------------ITLFGTRRETSVHNSRILK-TM----K---G--------------------------P-----------------------M-V-V-F----------
----------E---------------------ACH------G----------------------G--------------------K-------------------------LKDTED-
-----Y----------------------------------------------------------EI----------------------------------IN-----------K-A--
------SML----------ADSI-----PLKIGSKVMITKNVD------------------H-----T-----LYNGTTGTIVGF

>MBQ6280177.1_Mycoplasma_sp.

LTNIEFLNILKSE-IID------------------------------------------KK-----HSIFLTGSAGVGKTTLLQNLKNE---L-N--A--V------------------
-----G-ANAVLTSTTGLSSFHI-------GG------VTIHKF-MGI---N--I------QK-N---V---N------YLN---Y--------------F-----SHT--F--Q---
-F----Q---A-------------LKKRL--AK--FDVIIIDEISM--LRADQFTL-IDSVLKKAS----------------E---------N-----------------------N-
--------------------------QPFGG-----KIVVFSGDFFQIPPVV-----------------LE---N--E--------------------------------I-------
------------------------KKN-----Q-----------WIFTS------------------DPWIN----SN--I-KIYKLVHVHR----Q-S-----------D--N----
---------DFVNCLDEIKEG---K-V-N------S--------------KK------------------------------VK--DL-IE--KC---------EKRK------P---
----------------------------------------------------------------------------------------------T---------I-N-----------
D-------------------TVFFATNEECDEFNKTKIA-KL----P---G--------------------------K-----------------------Q-I-T-Y----------
----------I---------------------ATV------A----------------------G--------------------K---------------------------R-KQ-
-----Y----------------------------------------------------------NK------------DT--------------------II-----------R-E--
------CIA----------KEKL-----DLKIGAKVIIIYNDP---------------K--N-----R-----FVNGTKATVTKL

>P_grisea2

-----VLDPAQSA-LVDRI------V--------------------------------RG-----ENIFFTGSAGSGKSTVLKAFVKR---L-R--A--I------------------
-----G-KRVDVVAPTGRAALEV-------EG------STVHSY-AGW---D--A------TA-L---S---L-------EQ---A--------------T-----GRA--R------
-T----R---F-------------VKNRL--RR--TDVLVIDEISM--VSSFMFDL-LSHVMQIAR----------------HG--------D-----------------------Q-
--------------------------RPFGG-----AQVVVSGDFFQLPPVK-----------------PF---E--N-CYFCGRELQ----------IDGH-----
SGNLRLCPGTEF--EMKREVCRKP-------RFFDERK-----M-----------WAFCS------------------GAWQQ----CG--F-GCVELQTIHR----Q-R--------
---D--N-------------TLISILQKCRTG---YHL-E------Q--------------SE------------------------------ID--LL-CAP------------
RPHI------------------------------------------------------------------------------------------------------------------
V-D-----------A-------------------TQLLPKREDVLRENETRYN-NL----P---K--------------------------E-----------------------
SER-Q-Y--------------------Q---------------------SVD------SVEDC-----------------PG--------------------H---------------
----------LYK-S-------F----------------------------------------------------------S--------------SR-------------------
LE-------------H--------HKY----------LDCL-----KLRTGMVIVLRSNIS----------P----K--Q-----G-----LVNGSQGIVIGF

>QDP67633.1_Prokaryotic_dsDNA_virus_sp.

---------LQET-ALNIL------KN-------------------------------SK-----DNVFLTGAPGTGKSWLVDRYVEW---L-L--E--N------------------
-----G-EEPVITASTGIAALNI-------NG------KTLHSW-GGL---R--N------DH-P---I---D-------ER---D-------------------QDEI-----I---
-K----G---Y-------------SYENY--IS--TQTLIIDEVSM--VSAALLEN-INILAKRIR----------------G---------D-----------------------H-
--------------------------RFMGG-----IRVIVVGDFFQLPPVK-----------------GR-----------------------------------------------
--------------------------------------------FAFEC------------------EDWDE----AD--F-TVCYLHENKR----Q-S-----------E--P----
---------EFTDILQNIRGG---F-L-T------E--------------PQ------------------------------KE--VI-RS--KI---------IKDA-----SI---
---------------------------------------------------------------------------------------------VE---------D-P-----------
K-------------------IRLDTHNKKVDNINRMQLE-RL----P---G--------------------------F-----------------------P-Q-T-Y----------
----------K---------------------MQE------D----------------------G--------------------P--------------------------------
-----Y-----------------------------------------------------PDAI-E--------------K--------------------LK-----------K-N--
------CLS----------PEKL-----ILKVDTPVLFTRNDS---------------E--L-----R-----WVNGTQGVVREL

>C_Uhrbacteria

-----KISKEFKK-ALAIM------ED-------------------------------TK-----EHLFLTGNAGTGKSTLLQYFRKH---T-A------------------------
-------KKIVVLAPTGVAALNV-------KG------QTIHSF-FGF---N--P------SI-R---K-----------EN---VR------------------------K--A---
-S----V---D-------------KRRLF--ES--VETIVIDEISM--VRADLLDC-IDKSLRINR----------------N---------K----------------------PK-
--------------------------EAFGG-----VQMIFIGDLFQLPPVL-----------------TD---E--E-RYI----------------------------F-------
------EEE---------------YRS-----P-----------YFFSA------------------YVLGD----FD--I-NFIELKKVYR----Q-Q-----------D--N----
---------RFVELLNNLRNK---R-M-T------A--------------SD------------------------------VQ--IL-DT--CH---------DADF-----DP---
-----------------------------------------------------------------------GD---------------------D---------T-N-----------
F-------------------VHLTTTNKMAEQRNYSELQ-KL----E---G--------------------------E-----------------------E-Y-K-L----------
----------T---------------------GTK------N----------------------G--------------------D--------------------------------
-----F----------------------------------------------------------R------------------------------------A-----------S-R--
------LPS----------EDVL-----KLKVGARVMFTNNDS---------------Q--K-----R-----WVNGTLGTVTEI

>C_Aenigmarchaeota

-----EINDKFKE-SLGLM------ES-------------------------------TS-----KNIFITGKAGTGKSTLLNYFRSL---T-D------------------------
-------KKLAVLAPTGVAAINI-------DG------QTIHSF-FRF---K--P------DV-N---L-----------SS---IK------------------------K--Y---
-K----G---S-------------GGEIY--RR--IETIVIDEISM--VRADLLDC-IDKFLRING----------------K---------D----------------------SG-
--------------------------KPFGG-----VQMIFFGDLYQLPPVV-----------------RS---E--E-KEI----------------------------F-------
------KSH---------------YKS-----Q-----------YFFDA------------------KVFDS----LE--M-EFIELEKIYR----Q-K-----------D--E----
---------KFIRILNSIRNN---S-I-D------E--------------SQ------------------------------LK--LV-NE--RV---------KPDF-----KI---
-----------------------------------------------------------------------YL---------------------K---------D-I-----------
Y-------------------MQLTTTNKLSAEINEGELS-KI----R---S--------------------------P-----------------------L-L-S-Y----------
----------E---------------------GKI------K----------------------G--------------------N--------------------------------
-----F----------------------------------------------------------E------------------------------------K-----------H-Y--
------LPT----------EISL-----KLKVNSQIMLVNNDP---------------N--G-----R-----WVNGTVGKIIGI
```

```
>C_Pacearchaeota
-----EINEQFKR-ALELL------EN--------------------------------TS-----KNVLITGRAGTGKSTLLDYFVHH---T-Q-------------------------
-------KEVVVLAPTGVAAVNV-------GG-------QTIHSF-FGF---K--P------GI-T---V-----------DK---VK-------------------------K--A---
-Y----G---P-------------NSETY--KM--VDRIIIDEVSM--VRADLFDC-VDRFLRLNG------------------P---------D----------------------SS-
--------------------------KPFGG-----VQMAFIGDLYQLPPVV------------------KG---E--E-KGV-----------------------------F-------
------KTH---------------YKS------P-----------YFFDA------------------HLFQK----LK--V-EFIELEKIYR----Q-T------------D--Q----
---------EFIRLLNAVRNK---S-V-T-------E--------------ED------------------------------LK--KI-NK--RL----------APKF-----EP---
------------------------------------------------------------------------SS----------------------D---------E-F-----------
Y-------------------INLTTTNKLSEEINNKELG-KL-----S---S--------------------------K------------------------L-F-T-F----------
----------K---------------------GKI-------A----------------------G---------------------D--------------------------------
-----F-----------------------------------------------------------D-------------------------------------K-----------S-Y--
------LPT----------EEVL-----NIKEGSQIMLLNNDL----------------A--H-----R-----WVNGTVGKIIEI
>NMD11668.1_Acidobacteria_bacterium
-----EINDQFRQ-ALHWM------EE--------------------------------TA-----RPVFVTGKAGTGKSTLLEHFRET---T-L-------------------------
-------KKIAVLAPTGVAALNV-------RG-------QTIHSF-CGF---K--P------DI-T---L-----------AK---VR-------------------------K--
INAKKD----P---D-------------RAALL--RK--LDAVVIDEISM--VRADLLDC-VEKFLRLNG------------------P---------K---------------------
-PR---------------------------RPFGG-----LQLILIGDLYQLPPVV------------------AG---V--E-KTL-----------------------------F---
----------TLH---------------YET------P-----------YFFSA------------------HCLLR--DSFR--L-EFVELEKIYR----Q-T------------D--
A-------------GFIALLNAVRNR---S-A-G-------P--------------ED------------------------------LE--KL-HS--RY----------DPEF-----
VP---------------------------------------------------------------------------PE----------------------D---------D-F------
-----Y-------------------VTLTSTNDLAAARNREKLA-LL-----P---G--------------------------R------------------------L-Y-A-Y-----
---------------E---------------------AIV-------E----------------------G---------------------E---------------------------
----------F-----------------------------------------------------------E-------------------------------------R-----------
S-S--------LPT----------DEHL-----EIKAGAQVMLLNNDA----------------A--G-----R-----WVNGSIGRIAGV
>NPV00061.1_Brevinematales_bacterium
-----EINPEFAK-AMDFM------EN--------------------------------GK-----HHVFLTGKAGTGKSTLLSYFCEN---T-G-------------------------
-------LNHVILAPTGVAALNV-------GG-------QTIHSF-FGF---R--P------NI-T---K-----------DQ---IK-------------------------R------
-S----D---W-------------HVDFL--RN--LDVIIIDEVSM--LRADLLDY-IDEFLRINR------------------N--------------------------------PA-
--------------------------ETFGG-----IKMIFIGDLFQLPPVV------------------TG---N--E-EQI-----------------------------F-------
------KDY---------------YDS------P-----------YFFSA------------------HCLAG----VT--V-RYIELTKIYR----Q-N------------D--R----
---------RFIDILNNVRNN---N-I-T-------Y--------------RD------------------------------IS--EL-NR--RV----------DRGF-----EP---
------------------------------------------------------------------------SE----------------------D---------E-F-----------
Y-------------------IWLTPFNKTVMEINSYHLS-RL-----E---G--------------------------E------------------------A-H-R-F----------
----------T---------------------ADI-------R----------------------G---------------------D--------------------------------
-----F-----------------------------------------------------------E-------------------------------------E-----------K-Y--
------YPL----------EDPL-----ILKIGAQVMLLNNDI----------------E--G-----R-----WVNGSMGKITKI
>cd_WWE3
-----QLSNEFKN-AINLI------ET--------------------------------SG-----KNIFITGNAGTGKSTLLTYFTKV---T-D-------------------------
-------RNFAVLAPTGVAALNV-------SG-------QTIHSF-FGF---K--P------DI-T---L-----------NS---VK-------------------------K--V---
----------R-------------DASIF--EN--LEILIIDEISM--VRADLFDC-MEKALRINK------------------K---------S-----------------------N-
--------------------------LPFGG-----VQLVVIGDLNQLPPVV------------------TR---D--E-EHI-----------------------------FSG----
SA-----GSL---------------YDS------P-----------YFFSS------------------EAFKN----SS--F-EVVVLTKIYR----Q-S------------D--E---
----------NFLALLNSVRNN---T-L-S-------E--------------QD------------------------------IY--TI-NA--RV----------DPEY-----IP--
-------------------------------------------------------------------------DP----------------------E---------D-F----------
-T-------------------IHLTTTNKRAFELNEFQLS-KV-----F---G--------------------------E------------------------I-F-N-F---------
-----------K---------------------GST-------V----------------------G---------------------N-------------------------------
------F-----------------------------------------------------------D-------------------------------------I-----------R-Q-
-------LPV----------EETI-----VLKVGAQVMMLNNDR----------------E--K-----R-----WVNGSLGKITNI
>uncult_archaeon
-----EYSGEFRE-AFELM------EN--------------------------------TS-----ENAFITGRAGTGKSTFLKYFMGH---T-K-------------------------
-------KKSVVLAPTGVAAINV-------GG-------QTIHSF-FKI---P--P------RV-T---A-----------DE---AR-------------------------KEGLQ--
-R----K---K-------------RNGLY--RA--TELIIIDEISM--VRADLLDC-IDIFLRAGL------------------G---------S-----------------------E-
--------------------------KPFAG-----KQLAFIGDLYQLPPIV------------------MG---E--E-KEA-----------------------------F-------
------GQQ---------------YDS------E-----------YFFSA------------------KAMQK----TG--F-RRIEFTKIYR----Q-K------------D--Q----
---------EFIGILNRIRDK---T-A-T-------K--------------ED------------------------------IE--KI-NG--RF----------CEKI-----TD---
------------------------------------------------------------------------------------------------D---------R-G-----------
S-------------------IYVVMTNAMADEINMKKLG-EI-----I---G--------------------------V------------------------Q-Y-N-I----------
----------R---------------------GAI-------D----------------------G---------------------D--------------------------------
-----F-----------------------------------------------------------R-------------------------------------K-----------S-A--
------MPA----------DEIL-----HLKKGSQVMFLVNDP----------------Q--K-----R-----WVNGSLGEVTGI
>archaeon_CG07
-----LLNPEFKK-VLDLL------EN--------------------------------TN-----HSYFVTGKAGTGKSTLLKFFKDT---T-K-------------------------
-------KKAVVLAPTGLSALNV-------DG-------QTIHSF-FKF---P--P------RI-I---N---D-------KD---IK-------------------------K--V---
------------------------NSRIY--EE--LNCLIIDEVSM--VRADLMDG-IDKFLRKNR------------------N---------N-----------------------S-
--------------------------RPFGG-----VQVLFFGDLFQLPPVT------------------NE------E-TEI-----------------------------L-------
------NFT---------------YET------P-----------YFFSA------------------KAVLD----AD--L-KLVELENVYR----Q-Q------------E--K----
---------DFIQLLDNIRKG---N-N-V-------S--------------GS------------------------------LA--EI-NK--RV----------VNDF-----VP---
------------------------------------------------------------------------------------------------D---------N-D-----------
C-------------------VILTPTNYNADMINNQRLS-RL-----P---G--------------------------S------------------------E-K-N-Y----------
----------L---------------------ASA-------E----------------------G---------------------S--------------------------------
-----L-----------------------------------------------------------KN-----------------------------------QK-----------H-N--
------LPV----------NTVL-----KLKTGARVIFTRNDS----------------G--G-----S-----WVNGTLGTIIEL
>VVB74890.1_uncultured_archaeon
-----LLNDDFLK-AYNLL------EN--------------------------------NK-----SSFFITGRAGTGKSTLLRYFRDR---T-K-------------------------
-------KKIVVLAPTGLAALNV-------GG-------QTIHSF-FRL---P--P------RV-I---E---S-------HH---IK-------------------------K------
------V---E-------------DARLY--KE--LECIVIDEVSM--VRADLLDG-IDKFMRKNG------------------K---------D----------------------SD-
--------------------------KPFGG-----VQVILFGDLFQLAPVV------------------SS---T--E-TSL-------------------------------------
------SER---------------YMS------P-----------YFFSA------------------DVFDK----IK--L-SIIELEKVYR----Q-T------------D--K----
```

```
---------DFIAILDSIRTG---E-F-D------E--------------KT------------------------------LE--MI-NS--RV---------NPNF-----NA---
-----------------------------------------------------------------------ET---------------------D---------E-S-----------
F-------------------ITLTGTNEQANYLNMNKLG-SL----P---G--------------------------K-----------------------R-F-V-Y----------
----------N---------------------ASF------E----------------------G--------------------N--------------------------------
-----F----------------------------------------------------------D------------------------------------KNG---------K-N--
------FPV----------DPEL-----YLKVGSKVILTRNDP---------------S--G-----S-----YVNGSIGKVTDL
>MBI5066474.1_Candidatus_Woesearchaeota_archaeon
-----DFNDDFKK-AFNLI------EN-------------------------------TK-----RNLFITGKAGTGKSTLLKYFTAN---T-K------------------------
-------KNVVVVAPTGLAAVNV-------EG------QTLHSF-FKF---P--P------TL-I---T---K-------DD---IK------------------------K------
----------N-------------RGNVY--RW--IDTLIIDEISN--VRVDILDA-ADKVMRQNG----------------R---------N----------------------KN-
--------------------------EPFGG-----AQIVFFGDLYQLPPVV-----------------DR---A--A-SPF----------------------------I-------
------EDS---------------YGT-----P-----------YFFSL------------------NVIKE----LD--L-KIVELAKIYR----Q-K-----------D--Q----
---------EFIHILDKIRTG---K-L-S------E--------------ED------------------------------LN--KL-NE--RV---------TEEV-----SS---
--------------------------------------------------------------------------------------------------------D-E-----------
Y-------------------VNLVPTNYLAKIINCKKLD-AL----S---G--------------------------S-----------------------I-Y-T-Y----------
----------K---------------------AKL------E----------------------G--------------------E--------------------------------
-----F----------------------------------------------------------KP----------------------------------TT-----------N-N--
------LPA----------ELEL-----QLKEGARVIFVKNHP---------------H--E-----F-----WVNGTMGRVISL
>HHG53312.1_Spirochaetes_bacterium
-----ELNEDFKN-ALKLL------EN--------------------------------G-----ENIFLTGKAGTGKSTFLKYFIEH---S-N------------------------
-------KNMVVLAPTGVAALNV-------GG------QTVHSF-FSI---F--P------HE-D---M-----------SD---IE------------------KIIS--R--Q---
-T----K---T-------------KKNLY--KS--LDLIVIDEVSL--LRADLLDL-INEVLKTTL----------------N---------T-----------------------Q-
--------------------------KPFGG-----KQILFVGDLYQLPPVV-----------------TS---R--E-KEI----------------------------F-------
------SMF---------------YDS-----P-----------YFFSA------------------RCYPQ----LN--V-KIVEFEKIYR----Q-K-----------D--E----
---------NFIEILNKIRNG---E-V-S------Q--------------SD------------------------------ID--FL-NQ--RL---------ISNA-----KI---
-----------------------------------------------------------------------DN---------------------D---------V-M-----------
P-------------------VYLTPYNEMARKINEEKLS-EL----K---G--------------------------K-----------------------K-Y-E-F----------
----------P---------------------GKI------T----------------------G--------------------N--------------------------------
-----F----------------------------------------------------------N------------------------------------V-----------E-E--
------LPT----------DEIL-----VLKKNAQVMLLTNSK---------------D--G-----L-----WVNGTIGRVESF
>PIR99148.1_Candidatus_Collierbacteria_bacterium_CG10
-----DLNPQFAN-ALDIM------EN-------------------------------SP-----DSLFITGRAGTGKSTLLQYFKQT---T-L------------------------
-------KNIVVLAPTGVAAVNI-------GG------STIHSF-FQF---K--P------DV-T---L---E-------KA---WA------------------------K-GT---
-N----A---K-------------KPELY--RS--LDAIVIDEISM--VRADLLDC-VDAFMRRVC----------------A---------S-----------------------M-
--------------------------APFGG-----KRVIMFGDLYQLPPVV-----------------TQ---G--D-EEI----------------------------F-------
------RSR---------------YDS-----A-----------FFFSA------------------DVISR----TP--L-TFIELDHIYR----Q-T-----------D--D----
---------EFIKVLNAIREN---V-A-T------D--------------KD------------------------------LA--LL-NT--RV---------HEEY-----AP---
-----------------------------------------------------------------------PP---------------------E---------E-Y-----------
V-------------------IHLTGTNRDAQSYNTYQLH-SL----E---G--------------------------K-----------------------L-Y-S-F----------
----------K---------------------AES------T----------------------G--------------------A--------------------------------
-----F----------------------------------------------------------D------------------------------------R-----------R-T--
------EPA----------PREL-----ILKIGAQVMLTCNDR---------------E--K-----R-----FINGTVGRVEDI
>C_Micrarchaeota
-----EFGEDFNK-ALALL------ER-------------------------------PS-----GHVFVTGKAGTGKSTLLKYFRST---T-S------------------------
-------KKVAVLAPTGVAAVNV-------EG------QTIHSF-FGF---R--P------NT-----T---E-------SN---VR------------------------R--A---
-V----A---E-------------KQELF--KS--LDAVIIDEASM--VRADLLDC-IDKSLRLNR----------------S---------K-----------------------R-
--------------------------EPFGG-----VQMLFFGDLHQLPPVV-----------------TE---S--E-RDA----------------------------L-------
------QGA---------------YDS-----P-----------YFFDS------------------NALRK----TS--V-HVFELEKVYR----Q-K-----------D--A----
---------AFIELLNAIRTN---T-A-D------E--------------NH------------------------------LG--VL-NS--RV---------TRQL-----TG---
-----------------------------------------------------------------------QN---------------------G---------E-L-----------
C-------------------VTLTATNDVADSLNQQQLA-SI----R---R--------------------------P-----------------------P-Y-Y-F----------
----------E---------------------ATR------V----------------------G--------------------A--------------------------------
-----P----------------------------------------------------------D------------------------------------K-----------A-R--
------QPA----------PVRL-----ELKLGCQVMLLNNDS---------------A--G-----R-----WVNGTVGVVNGF
>Prevotella_sp
-----LQNPELQK-ALQII------QF-------------------------------TH-----NSLFLTGKAGTGKSTFLRYISST---T-K------------------------
-------KKHVILAPTGIAAINA-------GG------STLHSF-FKL---P--F------HP-L---V---P-------DD---SRYTPR-------HLR-----GTM--R--Y---
-N----G---D-------------KCKLL--RE--VELIIIDEISM--VRADIIDF-IDKVLRVYN----------------R---------N----------------------MR-
--------------------------EPFGG-----KQLLLVGDIYQLEPVV-----------------KE---D--D-RRL----------------------------L-------
------QPY---------------YAS-----N-----------YFFDA------------------KVFQD----YP--L-VSIELNKVYR----Q-N-----------D--S----
---------TFISILDHIRTN---Q-V-T------D--------------TD------------------------------FK--MI-NA--RV---------GASL-----EP---
-------------------------------------------------------------------QDK-DK---------------------E---------N-F-----------
T-------------------ITLSTKRDTVDWINNEGLD-RL----E---G--------------------------D-----------------------P-V-M-F----------
----------L---------------------GEI------K----------------------G--------------------E--------------------------------
-----F----------------------------------------------------------P------------------------------------E-----------S-S--
------LPT----------PMEL-----NLKVGAHVMFIKNDI---------------E--K-----Q-----WVNGTLGIIIGI
>B_ovatus
-----PQNHEQQL-AYELV------AN-------------------------------TN-----SSFFLTGRAGTGKTTFLHNVQKL---A-G------------------------
-------KQFITLAPTGVAAILA-------GG------DTIHSF-FGL---P--M------EV-C---T---P-------GT---CG------------------------K--M---
-N----E---T-------------KVLTL--LH--ADTIIIDEVSM--VRCDIMDA-IDYTMRKAL----------------R---------N-----------------------N-
--------------------------MPFGG-----KQIIFVGDMFQLPPVV-----------------KQ---GP-E-KDM----------------------------L-------
------KDL---------------YQTD---DF-----------FFYKS------------------NAIKR----MR--L-VKIEFRKVYR----Q-D-----------D--E----
---------HFLHILENVRLN---K-V-T------P--------------ED------------------------------IM--HL-NE--RV---------CTPT----------
-----------------------------------------------------------------------ED---------------------D---------G-A-----------
V-------------------ITLASINKTADKINLQHLE-EI----E---A--------------------------E-----------------------E-F-V-Y----------
----------E---------------------GTV------N----------------------G--------------------K--------------------------------
-----F----------------------------------------------------------E------------------------------------E-----------K-K--
------FPV----------DLEL-----RLKVGAQVMFTRNDQ---------------Q--K-----R-----WANGTLGKVTKL
```

```
>M_mazei
-----EADKDLQL-AFDFV------QH--------------------------------TN-----RSIFLTGKAGTGKTTFLKSLKLK---S-P------------------------
-------KRMIVVAPTGVAAINA--------GG------VTIHSF-FQL---P--F------HP-F---I---P-------SL---YLSEAGSTEKPERADR-----PGY--K--M---
-S----R---E--------------KINII--RS--LDLLVIDEISM--VRADTLDA-IDSTLRRYR----------------N---------R-----------------------F-
---------------------------IPFGG-----VQLLMIGDLQQLAPVV-----------------KD---D--D-REI----------------------------L-------
------GRY----------------YQS-----F-----------FFFES------------------KALEN----TD--F-VTIELKHIFR----Q-D-----------D--Q----
---------IFIDLLNRIRNN----D-V-N------Q--------------AV------------------------------LD--EL-NK--RY---------IPDF-----DP---
------------------------------------------------------------------------DS---------------------G---------G-G-----------
Y--------------------ITLTTHNHQARTINDSRLE-KL----P---G--------------------------K------------------------T-H-S-F---------
----------T----------------------AIV------K----------------------D--------------------E--------------------------------
-----F-----------------------------------------------------------P-------------------------------------E----------F-S--
------YPN-----------DTEL-----VLKTGAQVMFIKNDL----------S---GD--R-----L-----FFNGKIGKIISF
>Thermoplasmata
-----PPNHELEL-ANDFV------QY--------------------------------TG-----CNIFLTGKAGTGKTTFLHNLHKN---T-A------------------------
-------KRMIVTAPTGVAAINA--------GG------VTLHSF-FQL---P--F------GP-F---V---P-------GS---EA--------YERNKQ-----RRF--R--F---
-S----K---E--------------KKRII--QS--LDLLVIDEISM--VRADLLDA-VDAVLRGHR----------------R---------N-----------------------N-
---------------------------QPFGG-----VQLLLIGDLYQLSPVA-----------------KQ---D--E-WHL----------------------------L-------
------EQY----------------YES-----V-----------YFFSS------------------KALNL----TE--L-ITIELTHIFR----Q-S-----------D--A----
---------RFIKLLNRVRDN----R-L-D------E--------------SS------------------------------IA--DL-NL--RY---------IPNF-----TP---
------------------------------------------------------------------------GE---------------------D---------Q-G-----------
Y--------------------ITLTTHNRNAESINQTRLD-GL----P---K--------------------------K------------------------E-H-R-F---------
----------K----------------------AEV------S----------------------G--------------------D--------------------------------
-----F-----------------------------------------------------------P-------------------------------------E----------H-N--
------YPT-----------LATL-----LLKEGAQVMFVRNDL----------S---AE--K-----R-----YYNGKIGKITKI
>MBP7674859.1_Thermoanaerobaculia_bacterium
-----PENPDAER-AFELL------QG--------------------------------IE-----PCVFVTGRAGTGKSYLLRYFARK---T-K------------------------
-------KRIVLLAPTGLAALNV--------GG------QTIHSF-FMF---P--W------GL-M---N---R-------ED---VK------------------------Q-VW---
-D----S---N--------------KRQLI--RK--VDTFVIDEVSM--VNANLMDA-IDAFLRLNG----------------R---------D----------------------AR-
---------------------------KPFGG-----AQVVLFGDPYQLPPVL-----------------SR---E--DEAKF----------------------------M-------
------EYH----------------YRS-----P-----------FFWDA------------------KVFEQ----LP--I-TVVELRKNYR----Q-K-----------E--L----
---------EFMDVLNGIRLG----E-L-A------E--------------EH------------------------------QA--LL-NS--RC---------DPDF-----ET---
------------------------------------------------------------------------RP---------------------G---------E-I-----------
R--------------------PWLTTTNARAAQINAARLA-RL----P---G--------------------------P------------------------E-H-V-F---------
----------V----------------------ATF------S----------------------G--------------------K--------------------------------
----VF-----------------------------------------------------------D-------------------------------------G----------E-N--
------LPA-----------EAEL-----KLRPGAQVLFVKNDA---------------Q--D-----R-----WVNGTFGRVVTL
>C_collierbacteria
-----TVNPEKSD-IFDKI------EN--------------------------------SH-----KHFFITGKAGTGKSHLLKFLKTN---S-K------------------------
-------KQVVVCAPTGVAALNV--------SG------QTLHSF-FKI---P--P------HF-V---N---P-------EE---V-------------------------K--L---
-N----N---K--------------VAELL--RH--VEVLVLDEVSM--VRAEMIDI-IDHLLKQAR----------------E---------P-----------------------F-
---------------------------TPFGG-----VQLVMFGDPFQLPPIV-----------------AS---R--ELQEY----------------------------F-------
------SKN----------------HGG-----F-----------HFFNA------------------HVWED----VG--F-ETYELKEIFR----Q-K-----------D--D----
---------RFITLLNRVREG----D-V-D------D--------------DL------------------------------LA--QL-NR--RV---------EEFLE---------
-----------------------------------------------------------------------------------------------D---------S-P-----------
V--------------------IVLSTTNNKVNFINSNKLS-SI----P---S--------------------------K------------------------E-F-V-F---------
----------E----------------------AYI------S----------------------G--------------------A--------------------------------
-----L-----------------------------------------------------------D-------------------------------------E----------R-Q--
------YPA-----------DEIL-----KLKKGAQIMMLKNDP---------------D--D-----R-----WVNGSLGTVESL
>C_Pacebacteria
-----TLSQEQQE-VFNKL------ET--------------------------------TN-----GHFFITGKAGTGKSLLLQYFRTY---S-Q------------------------
-------KKLVVLAPTGVAALNV--------GG------QTIHSL-LRL---P--F------SA-I---T---L-------DS---F--------------------RRL--R--V---
-D----T---K--------------LKKLL--QS--LDCIVIDEISM--VRVDIMEA-IDYILKKAR----------------N---------S-----------------------Y-
---------------------------EPFGG-----VQMIMFGDLYQLPPVV-----------------TS---G--ELQQY----------------------------F-------
------DDT----------------YGG-----A-----------YCFNA------------------NSWRA----AK--P-EIITLSKIFR----Q-S-----------D--A----
---------TFIDLLNSLRDG----N-P-N------E--------------DF------------------------------LD--RL-NQ--RA---------SIA-----------
------------------------------------------------------------------------PP---------------------Q---------D-G-----------
A--------------------VTLATTNRTVSEINQRKLD-SL----R---A--------------------------D------------------------A-H-E-Y---------
----------E----------------------AEV------S----------------------G--------------------K--------------------------------
-----L-----------------------------------------------------------E-------------------------------------E----------S-A--
------FPT-----------DKVL-----QLKKGAQIMMLKNDR---------------D--K-----R-----WVNGSLGTIHSI
>Curtobacterium_sp
-----ELSDEQRA-VFEYI------EH--------------------------------TR-----DHVFITGRAGTGKSTLLNHLSWN---T-E------------------------
-------KQVVICAPTGVAALNV--------GG------QTIHSL-FRL---P--I------GL-I---A---D-------AE---LR------------------------Q------
-G----P---D--------------TRKLL--NT--IDTLVIDEVSM--VNADLLDG-MDRSLRKAR----------------G---------R----------------------QF-
---------------------------EPFGG-----VQVVMFGDPYQLPPVP-----------------GD--AD--E-RAY----------------------------F-------
------TDH----------------YRS-----M-----------WFFDA------------------KVWLE----AE--L-NIIELATVHR----Q-R-----------D--D----
---------AFAAMLTAVRHG----R-V-T------A--------------DI------------------------------AE--QL-NT--AG---------ARPA------P---
-----------------------------------------------------------------------------------------------D---------D-------------
A--------------------ITLATRNDTVARINKAALE-RL----P---G--------------------------K-----------------------V-K-T-A----------
----------K----------------------ADV------N----------------------G--------------------D--------------------------------
-----F-----------------------------------------------------------R-------------------------------------G----------R-N--
------FPA-----------DEAL-----ELKPGAHVMFLRNDA---------------D--Q-----R-----WVNGTLGIVTAI
>Clavibacter_sp
-----PLSPEQAA-VFQAI------EG--------------------------------TR-----DHIFVTGRAGTGKSTLLTHLSWN---T-E------------------------
-------KQIVICAPTGVAALNV--------GG------QTIHSL-FKL---P--I------GV-I---A---D-------EE---IE------------------------Q------
-T----G---E--------------LRKLL--NT--IDTLVIDEVSM--VNADLVDA-IDRSLRQAR----------------H---------K----------------------KD-
---------------------------VPFGG-----VQVVLFGDPYQLAPVP-----------------GD--GD--E-RAY----------------------------F-------
------ADR----------------YRS-----M-----------WFFDA------------------KVWEE----AQ--L-RIYELTEIHR----Q-H-----------E--E----
```

```
---------AFKEMLNAVRHG---R-V-T------A--------------EI--------------------------------AG--VL-NA--AG---------ARQA------P---
------------------------------------------------------------------------------------------------T---------D-G-----------
A-------------------ITLATRNDTVNRINAEALK-RL----P---G----------------------------K-----------------------S-L-T-A----------
----------T---------------------ADV------T------------------------G--------------------D--------------------------------
-----F------------------------------------------------------------G------------------------------------G-----------R-T--
------YPA----------DEKL-----DLKIGAQVMFLRNDA----------------D--Q-----R-----WVNGSVGVVTRI
>P_faecalis
-----QLTQEQQA-VYDAI------EQ---------------------------------TT-----EHLFVTGRAGTGKSTLLNHLSFH---S-E------------------------
-------KQLVICAPTGVAALNV--------GG------QTIHSL-FKL---P--I------GL-I---G---N-------QP---IE-------------------------Q------
-N----R---D-------------VKRLL--RK--IDTLVIDEISM--VSADLLDA-MDRSLRQAR-----------------E---------R-----------------------AA-
--------------------------EPFGG-----VQVVMFGDPFQLAPVP------------------PSD-PT--E-RAW-----------------------------L-------
------RDN---------------YRS-----M-----------WFFDA-------------------HVWRD----VE--M-RIHTLREIHR----Q-H------------D--D----
---------EFRSLLTAVRYG---Q-V-T------A--------------DM--------------------------------AG--RL-NE--VG---------ARTA------P---
------------------------------------------------------------------------------------------------T---------D-G-----------
I-------------------ITLASKNATVTRINSRELE-RL----P---G----------------------------R-----------------------A-M-T-A----------
----------E---------------------AEV------H------------------------G--------------------D--------------------------------
-----F------------------------------------------------------------E-----------------------------------GA-----------R-T--
------FPA----------EKEL-----VLKEGAQVMFLRNDP----------------D--G-----R-----WVNGTVGEVSRI
>Leucobacter_sp
-----TLTAEQQA-VFNRI------ET---------------------------------TR-----EHLFITGRAGTGKSTLLNHLAQN---S-S------------------------
-------KTLAICAPTGVAALNV--------GG------QTIHSL-LKL---P--T------GV-I---A---D-------HE---LT-------------------------Q------
-T----R---E-------------LKKLL--QA--LTTLVIDEISM--VSADLMDG-IDRALRQAR-----------------K---------K-----------------------PF-
--------------------------DPFGG-----VQIVMFGDPYQLPPVQ------------------PRD-PH--E-IAY-----------------------------Y-------
------KDT---------------YPS-----L-----------WFFDA-------------------KVWHD----AP--V-SVVELTEILR----Q-R------------D--D----
---------RFKEILNAVRIG---Q-V-D------S--------------AM--------------------------------AA--EL-NA--AG---------ARPA------P---
------------------------------------------------------------------------------------------------V---------E-G-----------
T-------------------ITLATTNRAVKEINERELA-KL----P---G----------------------------K-----------------------E-L-K-A----------
----------Q---------------------AEV------T------------------------G--------------------S--------------------------------
-----F------------------------------------------------------------S------------------------------------E-----------N-S--
------YPA----------DETL-----RLKVGAQVMFLRNDP----------------E--G-----R-----WVNGTLGVVSRI
>QIG57888.1_Microbacterium_phage_PauloDiaboli
-----TLSAEQQA-VVDLI------NG---------------------------------TR-----DHIFITGRAGTGKTAVLKAFAKQ---T-K------------------------
-------KKYVIAASTGIAALNA--------GG------MTLHRL-AGV---G--T------AL---------P-------AD---MG--------------------VDL--N--K---
-V----A---S-------------KRRWL--KH--IDTIIIDEVSM--VSADLMDS-VDRNLQHIR-----------------Q---------N-----------------------HQ-
--------------------------EPFGG-----AQIIMFGDPYQLPPVV------------------SK---I--D-QKW-----------------------------YD------
------ANK---------------YRS-----A-----------WFFDA-------------------KVWRG----NE--F-KTVELQTIFR----Q-E------------D--D----
---------MYKDLLNGVRDG---S-L-D------K--------------DG--------------------------------LF--AL-NA--LG---------ARQGR---------
------------------------------------------------------------------------------------------------T---------E-Q-----------
S-------------------LLLGSRNDIVLHRNRRKMG-EL----R---G----------------------------R-----------------------T-H-V-Y----------
----------E---------------------ARV------N------------------------K--------------------G--------------------------------
-----F------------------------------------------------------------G------------------------------------------------R-G--
------EPA----------ERRL-----EVKVGSHVMMLNNDS----------------E--D-----R-----WVNGSRGEIVYC
>Parabacteroides
-----IVTGEMKH-FLDLV------EN---------------------------------TS-----QCIFLTGKAGTGKSSLLRMLLAR---T-S------------------------
-------KQIVVAAPTGVAAVNV--------GG------VTLHSL-FQL---P--F------GP-F---I---P-------NI---DLLGY------THDAL------PAY--K--F---
-S----S---E-------------KAEVL--QN--MEVLVIDEVSM--LRADVLDA-INDVLCHVR-----------------N---------N------------------------P-
--------------------------LPFGG-----VQVVFIGDLYQLPPVV------------------NK---T--E-WKL-----------------------------L-------
------SSV---------------YQT-----P-----------FFFSS-------------------KALVL----TE--L-RLVCLTHIFR----Q-T------------D--D----
---------NFISLLNDVRCG---K-L-S------D--------------SS--------------------------------RR--LL-NA--LY---------KPGI-----DA---
-------------------------------------------------------------------------SE--------------------LE---------D-G-----------
F-------------------VMLTTHNAKADKVNQDKLD-EL----S---T----------------------------Q-----------------------Q-Q-V-F----------
----------T---------------------ASV------K------------------------G--------------------N--------------------------------
-----F------------------------------------------------------------S------------------------------------A-----------S-A--
------MPA----------EFEL-----CLKIGAKVMFLANDN----------------E-AH-----L-----YHNGSTGVVVSF
>DAU76505.1_Myoviridae_sp.
-----IINSAMKE-AIDLV------LN---------------------------------TN-----TNVYLTGRAGTGKTTLLRYILGV---C-K------------------------
-------KNTIIAAPTGVAAINA--------GG------VTLHSL-LKL---P--F------SP-Y---K---P-------AF---VR-------GKTLHVL------GSY--K--L---
-N----D---K-------------QIETI--QK--LELLVIDEISM--VRADLLDA-VNDALCFYR-----------------N---------T------------------------K-
--------------------------EPFGG-----VQLLLIGDLYQLPPVT------------------IK---E--E-WGL-----------------------------V-------
------EKY---------------YDS-----P-----------YFFCS-------------------KALKT----AG--F-KTVNLSHVFR----Q-S------------D--E----
---------EFLHLLNEVRNG---N-L-S------A--------------ES--------------------------------RK--KL-LE--LY---------DKRY-----IG---
-------------------------------------------------------------------------NK---------------------E---------S-G-----------
Y-------------------ITLCATNKSAQNINMDSLA-RL----E---G----------------------------E-----------------------I-Y-R-Y----------
----------D---------------------AIL------S------------------------G--------------------D--------------------------------
-----F------------------------------------------------------------P------------------------------------E-----------N-A--
------APC----------EPQL-----NLKVGAQVMFCANDQ-----------APM-EQ--R-----K-----FYNGMLGVVEEI
>AFB75491.1_Bacteriophage_sp.
-----ILTEEMQK-IMNLI------QD---------------------------------DE-----NNVFVTGKAGSGKTTFLKYLIEK---S-G------------------------
-------KNCIVAAPTGIAAINA--------GG------VTLHSL-FGI---P--F------GP-I---T---P-------YD---R---------------------LEN--K--F---
-S----E---Y-------------KVELL--LK--MELLIIDEISM--VRPDILDT-IDRKLRWVY-----------------E---------S------------------------D-
--------------------------EPFGG-----VQVVMFGDLFQLPPVT------------------KK---Q--E-REI-----------------------------L-------
------SDF---------------YDG-----F-----------FFFNA-------------------LVFKR----TG--F-HIVELTKIFR----Q-T------------E--P----
---------EFINVLNNIRNY---Q-V-T------S--------------DE--------------------------------LD--LL-SE--LK---------DRKI-----SS---
-------------------------------------------------------------------------SY---------------------D---------N-E-----------
Y-------------------IHICTHKADVERINADKLG-EQ-------------------------------------E-----------------------I-R-N-Y----------
----------D---------------------IVI------K------------------------D--------------------K--------------------------------
-----F------------------------------------------------------------P------------------------------------E-----------S-S--
------IPC----------DLHL-----KLRVGARVMSLVNDS----------------L--K-----G-----YYNGMLGIVTAL
```

```
>DAD87486.1_Siphoviridae_sp._ctAUQ2
-----ILTDEMSR-AMELV------QK-------------------------------TN-----HHVFITGKAGTGKTTFLKYLIKN---C-K------------------------
-------KNCVVAAPTGIAAINA-------GG------VTLHSL-FGI---P--F------KP-I---S---P-------VE---R--------------------LEY--K--F---
-T----E---Y-------------KTAML--LK--LDLLIIDEVSM--VRPDIMDT-VDRKLRWVR----------------E---------S-----------------------D-
--------------------------EPFGG-----VQVVMFGDLFQLPPVV-----------------KS---D--E-EEI----------------------------L-------
------GRF---------------YDD-----Y-----------FFFNA------------------QVWRQ----MG--F-HVIELNQVFR----Q-T-----------D--Q----
---------TFVNVLNNIRNY---K-V-S------D--------------EE------------------------------LD--IL-SE--IK---------DKNI-----SQ---
-----------------------------------------------------------------------SY---------------------T---------G-E-----------
Y-------------------IHICTHRKDVEKINTSLLG-E------------------------------------P-----------------------T-W-C-Y----------
----------K---------------------AVL------K----------------------D--------------------K--------------------------------
-----F----------------------------------------------------------T------------------------------------E-----------S-A--
------APC----------DMEL-----KLRVGARVMALCNNP---------------Q--Q-----G-----YYNGMLGFVVDL
>DAO03073.1_Siphoviridae_sp.
-----QLTNEMVE-AVDII------QN-------------------------------TN-----QSLYITGKAGTGKTTFLRYIVNN---I-K------------------------
-------KKFIVTASTGIAAVNA-------GG------VTLHSL-LNI---P--F------GV-L---T---E-------SE---N--------------------VHS--S--Y---
-K----P---E-------------KAMLL--RS--IDAIIIDEVSM--VRPDVIDY-VDRKLQMYR----------------G---------S-----------------------S-
--------------------------EPFGG-----VQIIMFGDLFQLPPVV-----------------KA---D--E-QHI----------------------------L-------
------SQF---------------YRG-----I-----------YFFHA------------------HVWRN----AG--F-KVIELTHIFR----Q-N-----------D--K----
---------RFIEILNNIREY---H-I-M------Q--------------ED------------------------------ID--DL-AA--LR---------NKNE-----SK---
-----------------------------------------------------------------------DF---------------------S---------N-S-----------
S-------------------IHICAYRKDVQKINTELLG-E------------------------------------P-----------------------T-H-V-Y----------
----------K---------------------AMV------T----------------------G--------------------D--------------------------------
-----F----------------------------------------------------------Q------------------------------------P-----------N-S--
------APC----------EQEL-----KLRVGARVMMLVNDP---------------A--H-----V-----YCNGSLGEVVNL
>DAQ71114.1_Podoviridae_sp.
-----EKNVPQGL-ALKEL------VE-------------------------------GK-----GHMFITGRAGSGKSTFLRRVFPF---L--------------------------
-------DNAVIVAPTGVASLNI-------GG------ATIHSC-FGL---P--I------DP-Y---C---P-------VV----N-------PSRTEFV-----NTC--K--F---
-N----P-----------------TAYKM--KK--VKMVIIDEISM--VRPDLLDC-LADVLRQIK----------------H---------N----------------------DS-
--------------------------DPFGG-----VRIIMFGDLSQLPPVT-----------------SN---D--D--P-----------------------------L-------
------YTY---------------YDS-----R-----------FFFSS------------------KALRA----SG--F-NVFNFDKVFR----Q-K-----------D--P----
---------TFLEVLDEVKSG---E-L-S------V--------------GS------------------------------EE--IL-NS--RV---------GTP-----------
-----------------------------------------------------------------------QN---------------------M---------D-N-----------
V-------------------VTICSTNMELQAINNENLM-RI----N---G--------------------------E-----------------------E-H-T-F----------
----------N---------------------AVI------K-------------------------------------------N--------------------------------
-----V----------------------------------------------------------K------------------------------------H-----------T-S--
------APC----------EEIL-----RLKVGAKVVITKNGL---------------P--------D-----YVNGSVGKVTGF
>DAJ22427.1_Podoviridae_sp._ctfN46
-----DKNVEQGR-ALKKI------FT-------------------------------TR-----ENLFITGRAGSGKSTFMRRIVKF---L-G------------------------
--------KCVIVAPTGVAALNA-------GG------QTIHSF-FSI---K--N------DP-Y---V---P-------GF----E---------HGMLS-----NKI--E--V---
-G----G---F-------------VKSKV--KR--LDTIIIDEVSM--VRPDLLDE-MADILRQSK----------------R---------S-----------------------K-
--------------------------NPFGG-----VRIIMFGDLSQLPPVV-----------------TE---D--D---I----------------------------I-------
------DRY---------------YDS-----H-----------FFFSS------------------KALRA----SG--F-SVIKFNRVFR----Q-N-----------D--N----
---------EILTVLEDIRNG---V-I-T------E--------------ES------------------------------KR--IM-ES--RV---------MVP-----------
-----------------------------------------------------------------------EN---------------------M---------D-D-----------
V-------------------VIVCSTNKEASVINNENLS-KL----S---G--------------------------E-----------------------S-Y-E-F----------
----------E---------------------AEV------V----------------------G--------------------D--------------------------------
-----------------------------------------------------------------------------------------------------R-----------P-N--
------APC----------EDKL-----VVKVGAKVLITRNGC------------------------G-----YVNGSTGIITSI
>DAL07837.1_Bacteriophage_sp.
-----DKNVEQGR-ALKKI------FT-------------------------------TR-----ENLFITGRAGSGKSTFMRRIVKF---L-G------------------------
--------KCVIVAPTGVAALNA-------GG------QTIHSF-FSI---K--N------DP-Y---I---P-------SI----E---------RGMLS-----NKV--D--V---
-S----P---F-------------MKKKI--RN--LDTIVIDEISM--VRPDLLDE-VADILRQCR----------------R---------S-----------------------K-
--------------------------EPFGG-----VRLIMFGDLSQLPPVV-----------------TA---D--D---F----------------------------I-------
------DKY---------------YES-----R-----------FFFSS------------------KALRA----SG--F-SVITFENVFR----Q-K-----------D--P----
---------QLLSVLEDIRCG---V-I-T------D--------------ES------------------------------RQ--IL-DS--RV---------KYP-----------
-----------------------------------------------------------------------DN---------------------M---------D-N-----------
T-------------------IIICSTNKEAYEINKTNLD-KI----N---N--------------------------K-----------------------V-F-K-F----------
----------D---------------------ATV------F----------------------G--------------------E--------------------------------
-----------------------------------------------------------------------------------------------------K-----------P-V--
------APC----------EDEL-----IVKVGAKVIITRNGN------------------------G-----YVNGSMGIITSI
>DAY30538.1_Bacteriophage_sp.
-----EGNVAQGK-AIKSI------CK-------------------------------SP-----KPLFITGKGGSGKTTFLKRIIPA---L--------------------------
-------KNAVVVAPTGVAAVNA-------GG------QTIHSF-FRI---G--M------QP-Y---I---P-------EI---RK----------GAFM---DNCEY--K--F---
-N----G---G-------------SEKIL--QN--IKYLIIDEISM--VRPDLLDN-VADILRHAR----------------G---------D-----------------------K-
--------------------------DPFGG-----VKLIMVGDLFQLPPVI-----------------KE------D---F----------------------------F-------
------REI---------------YDT-----S-----------YFFSS------------------KSLMA----SG--M-EMVSFEKIYR----Q-K-----------D--E----
---------KFISVLNKVREG---Q-M-D------D--------------DV------------------------------FD--TI-NS--RC---------IQS-----------
-----------------------------------------------------------------------DN---------------------N---------Q-G-----------
Y-------------------VEIVTTNSKATAINEMRIS-SL----P---G--------------------------S-----------------------L-R-K-L----------
----------E---------------------AVI------N----------------------G--------------------D--------------------------------
-----Y----------------------------------------------------------P------------------------------------------------K-D--
------APV----------EKTL-----FLKEGSRVMITRNGG------------------------E-----YFNGSLGTVLSI
>Hyphomonas_sp
--------TIYAK-PAEWV------SRG----------------------------AGAQ-----GNLFLTGRAGTGKTTLLRRFVEQ---A-G------------------------
-------DSAIVLAPTGVAAMNA-------GG------QTLHSF-FKL---P--P------RL-I---E---P-------QD---VK------------------------R--L---
----------R-------------TARIM--KA--AETIIIDEISM--VRADMLDA-IDRSLKLNR----------------G---------S-----------------------K-
--------------------------RPFGG-----VRMILSGDLHQLPPVV-----------------RG---D--E-DPI----------------------------L-------
------KER---------------YGG-----H-----------YFFNA------------------PAFKE----AE--F-ALLALKHVFR----Q-E-----------D--P----
```

```
---------RFLALLGAMRQG---R-L-T------P--------------AD------------------------------ES--VL-RS--VV---------SDRD-----AV---
-----------------------------------------------------------------------EA---------------------S---------E-T-----------
H-------------------IVLTPNNANAFRINQARLD-EL----P---G--------------------------P-----------------------E-K-V-F----------
----------E---------------------ARV------Q----------------------G--------------------D--------------------------------
-----F----------------------------------------------------------D------------------------------------E-----------K-T--
------YPT----------EADL-----ELKECARVMLIKNDP---------------D--G-----R-----WVNGSLATVSGW

>NQY15510.1_Henriciella_sp.

-----NPDTIYAK-PAEWV------ADG----------------------------AGAQ-----GNLFLTGRAGTGKTTLLRKFMAH---A-G------------------------
-------ESAIVLAPTGVAAMNA-------GG------QTLHSF-FKF---P--P------RL-I---E---P-------QD----VK-----------------------R--L---
----------R-------------TARLM--KA--AETIIIDEISM--VRADMLDA-IDRSLKLNR----------------G---------S-----------------------K-
--------------------------RPFGG-----VRMILSGDLHQLPPVV-----------------RG---D--E-DPI----------------------------L-------
------KER---------------YGG-----H-----------YFFNA------------------PAFKE----AE--F-ALLALKHVFR----Q-E-----------D--P----
---------RFLALLGAMRQG---R-L-T------P--------------AD------------------------------DS--VL-RG--LV---------SSRD-----AV---
-----------------------------------------------------------------------DA---------------------S---------E-T-----------
H-------------------IVLTPNNANAFRINQARLD-DL----P---G--------------------------P-----------------------E-K-V-F----------
----------E---------------------AKV------Q----------------------G--------------------T--------------------------------
-----F----------------------------------------------------------E------------------------------------E-----------K-S--
------YPT----------EADL-----ELKEGARVMLIKNDP---------------E--G-----R-----WVNGSLATVSGW

>Robiginitomaculum_sp

-----IDHDIYAP-VLEVL------EK-------------------------------SR-----DNVYLTGRAGTGKTTLLKAFVAR---N-A------------------------
-------ETTAVLAPTGIAAVNA-------GG------QTIHSF-FRL---P--P------RL-I---E---P-------GD----VK-----------------------R--I---
----------R-------------YARAL--RA--IETLVIDEVSM--IRSDVMAA-IDRSLRINR----------------D---------V-----------------------D-
--------------------------APFGG-----VQMVLVGDPYQLPPVI-----------------ER---G--L-EGY----------------------------L-------
------EET---------------HGG-----S-----------YFFSP------------------PAFRE----GG--F-QLIELTKVFR----Q-S-----------D--P----
---------VFLDILAGVRRG---D-M-D------R--------------DQ------------------------------ME--IL-SA--QV---------SSMD-----PV---
-----------------------------------------------------------------------AA---------------------S---------Q-T-----------
H-------------------VVLTGTNNAAFDINHRRLE-AL----P---G--------------------------K-----------------------A-Q-A-Y----------
----------A---------------------AQI------K----------------------G--------------------E--------------------------------
-----F----------------------------------------------------------D------------------------------------P-----------R-L--
------YPT----------EAPL-----YLKAGARVMMLKNDP---------------D--K-----N-----WVNGTLATVLST

>MBT3274541.1_Spirochaetales_bacterium

-----TPTVEMEDFVTEFH------SR-------------------------------KH-----GIFFVTGEAGTGKSTLLRQFYQQ---V-K------------------------
-------TRAVCVAPTGIAALNI-------SG------QTIHSF-FKF---P--P------SL-I---N---P-------AE----VK-----------------------K------
------Q---K-------------DPRIY--QK--IDFLIIDEVSM--LRPDLFDA-IDVSLKMNR----------------N---------C----------------------SE-
--------------------------LPFGG-----VTVILFGDLYQLPPVV-----------------ED---G--E-RNI----------------------------L-------
------KEM--------------GYRT-----R-----------YFFSA------------------MAFKENI--AQ--V-KMCRLTKVFR----Q-H-----------S--D----
---------SFINLLNQVRNC---E-L-T------D--------------EI------------------------------GR--LL-DS--RL---------IDEE-----DA---
-----------------------------------------------------------------------KL-------------------------------T-D-----------
A-------------------LVLTTTNKVANAYNQDFLD-EL----P---G--------------------------N-----------------------S-K-V-F----------
----------R---------------------AKV------T----------------------G--------------------D--------------------------------
-----F----------------------------------------------------------K------------------------------------K-----------Q-E--
------YPT----------EEYL-----ELKKDAKILFIKNDE---------------G--N-----R-----WINGSIGIVSGL

>A_illinoisensis

-----EILPEYLF-VKQLV------EQ-------------------------------QF-----PVIFLTGGAGTGKSTFIKWLCRE---Y-R------------------------
-------GEVLLGAPTAMAAINV-------GG------RTLHSM-FQL---P--P------AW-I---V---K-------QD----IK------------------------------
------P---G-------------KKREI--KK--AKLLIIDEISM--VTANLLDG-ISAYLRLNR----------------G---------I-----------------------D-
--------------------------KPFGG-----LTVVMVGDLFQLPPVI-----------------SE---K--T-RDL----------------------------F-------
------EQV---------------YGS-----P-----------KFYNA------------------RSLKT----TD--Y-CAIELTHTYR----Q-T-----------Q--Q----
---------DFVQLLCNIREG---Q-D-L------A--------------DS------------------------------IE--QL-NQ--RC---------LITK-----TP---
----------------------------------------------------------------------------------------------P---------Q-G-----------
A-------------------VWLSPRNAEVEHKNQAELA-RI----D---A--------------------------S-----------------------E-V-C-Y----------
----------S---------------------GKL------E----------------------G--------------------E--------------------------------
-----F----------------------------------------------------------K------------------------------------E-----------D-R--
------LPS----------PLHL-----RLKVGAQVMFTQNDP---------------Q--R-----R-----WLNGTVGQMTAL

>MBL6903300.1_SAR86_cluster_bacterium

-----SFDEIKDQ-VIHLL------DND------------------------------EQ-----EFIYLTGAAGTGKTTLLEVIKAD---L-D------------------------
-------KKMIVVAPTGIAALNI-------GG------TTINSA-FRI---G--F------DT-F---P---E-------IT----KS--------------------------K---
-D----P---R-------------FNKLL--KK--LEVLIIDEVSM--VRAPMLDA-ISQTLKIHR----------------G---------N-----------------------D-
--------------------------EPFGG-----VSVLACGDLFQLPPVV-----------------KE---Y--E-EKI----------------------------I-------
------FDK---------------YDS-----I-----------YFFSA------------------HSFQE-F--TQ--P-KFFELTKSFR----QED-----------D--N----
---------DFYDLLNNIRLG---E-D-L------E--------------NT------------------------------IN--SF-NR--SC------------F-----SPE--
-----------------------------------------------------------------------SE---------------------T---------E-S-----------
S-------------------MIITSRKNRAEHINEEMLN-RI----E---G--------------------------T-----------------------Q-V-S-T----------
----------K---------------------SKE------Y----------------------G--------------------D--------------------------------
-----L----------------------------------------------------------N------------------------------------E-----------N-D--
------LPA----------PREL-----KLKVDAKVMFIKNDS---------------A--G-----R-----WVNGTVGIVTQC

>MBR3410882.1_Candidatus_Methanomethylophilaceae_archaeon

-------NVDQTL-AREAI------LN-------------------------------SS-----LNLLIVGKAGTGKTTFLREVVGQ---C-K------------------------
-------KKLAVVAPSGIAAIEA-------EG------RTIHSF-FGF---N--T------AA-F---A---P-------GS----K------------------DGSL-KR--L---
-T----Q---G-------------QREWI--NR--LELLIIDEISM--VRADLLDH-IDSRLRTIR----------------H---------I-----------------------E-
--------------------------RPFGG-----VQVVMIGDLKQLPPVI-----------------DR---R--D-GEI----------------------------L-------
------DDF---------------YET-----G-----------YFFES------------------QALKA----SD--Y-VFIEFKTVYR----Q-D-----------D--K----
---------AFVSLLNRVRDN---L-V-T------D--------------KD------------------------------IA--EI-NK--RF---------REQC----------
----------------------------------------------------------------------------------------------N---------E-D-----------
Y-------------------VHLVTHRRQARRINESRME-AL----P---G--------------------------R-----------------------S-Y-S-F----------
----------H---------------------GSV------E----------------------G--------------------F--------------------------------
-----F----------------------------------------------------------Y------------------------------------K-----------K-D--
------YPA----------PEEL-----VLKKGAKVMFVRNDD---------------P--H-----G-----YVNGTFGIVESV
```

```
>MBR5312660.1_Clostridia_bacterium
-----LLDKEQAF-ACSEM------EH-------------------------------TQ-----DNFFITGKAGTGKSFLLDVFRNT---T-E------------------------
-------KNHIVLAPTGIAALNV-------GG------VTLHSV-FGY--YN--L------EN-L---S---I-------DM---LS------------------SATL--R--L---
-K----S---E-------------IYSIL--QR--VSTIIIDEISM--VRVDIFEK-VDRILKIIN----------------N---------N-----------------------D-
--------------------------LPFGG-----KQLLLFGDLFQLPPVA-----------------KS---K--E-REY----------------------------L-------
------LDQ---------------YGG-----V-----------HFFFS------------------NAYKT----GT--F-RFLELTINHR----QKD-----------D--A----
---------EYFSLLNRIREG---K-V-T------P--------------ED------------------------------IV--TL-NT--RV---------SNDI-----SV---
----------------------------------------------------------------------------------------------Y---------D-R-----------
F-------------------TTLLPKKADVEHINQYRIA-QL----D---S--------------------------V-----------------------G-Y-T-Y----------
----------E---------------------AKI------V----------------------L--------------------D--------------------------------
----KY----------------------------------------------------------PDKN----------HN--------------------LE-----------S-L--
------FPV----------AHSL-----CLKKGALIMMVANDP---------------E--R-----R-----WVNGTLGIVNNL
>MBU4069976.1_Nanoarchaeota_archaeon
-----VDKFDKEG-LFTFL------ER-------------------------------TN-----TNILITGPGGTGKSTILKKFKEK---T-K------------------------
-------KNCVILSPTGIAANNV-------GG------QTIHSF-FKL---D--I------GV-Q---T---P-------ET---MN------------------------K------
------E---R-------------WSPLY--EK--IDLIIIDEISM--VRKDVFEY-MDKIMRKYK----------------D---------S-----------------------A-
--------------------------KPFGG-----VKLILFGDLYQLPPVI-----------------TM---E--A-KNH----------------------------L-------
------RNI---------------YDND---LN-----------YFFDS------------------EIYSK----LD--L-LILNLNEIYR----Q-K----------ED--K----
---------PYAKLLDKMRRN---E-I-D------N--------------ED------------------------------LD--IL-NQ--NV---------TSNE----------
----------------------------------------------------------------------------------------------P---------D-K-----------
E-------------------PILSTKNDLVESYNRKKLS-SL----P---G--------------------------D-----------------------I-K-I-Y----------
----------D---------------------SKVVPPPWLKY------------------------------------------N--------------------------------
-----F---------------------------------------------------------------------------------------------VLK-----------K-Y--
------CNA----------EEKL-----ELKIGARIMVLINDA----------G---EN--K-----R-----YFNGSLGTVKEL
>MBQ7366747.1_Spirochaetaceae_bacterium
-----HDNPELKK-AYEDI------KS-------------------------------NV-----PAIFLTGGAGTGKSTFIKYLQNK---L-K--E--E------------------
----TG-KNCIIIAPTGIAAVNV-------RG------QTIHSF-FKF---P--I------GP-F---E---E-------KD---IK-------------------------------
-K----Q---N-------------KNPVV--DH--TDLIIVDEISM--VSSWLLDR-MDYALRLWC----------------N---------S-----------------------E-
--------------------------KPFGG-----KQVLLIGDCFQLPPVN-----------------NS---N--D-KDVQ---------------------------KF-------
------LKQ---------------WDN-----I-----------FFFAA------------------KVFEN----IE--V-EPIQLTKIYR----Q-E----------AD--K----
---------PFINILNSIRTC---T-K-G-----VA--------------DA------------------------------IN--FL-NE--KC-------LIEKRL----GTP---
-----------------------------------------------------------------------NV---------------------P---------S-N-----------
C-------------------LLLTTTNSDANKFNIERMN-NL----KY-KG-------------------------KE-----------------------S-M-T-F----------
----------K---------------------ASK------S----------------------G--------------------V--------------------------------
-----F----------------------------------------------------------D------------------------------------A-----------D-D--
------FLT----------PETL-----ELCIDATVMVTKNTS------------------S-----G-----LINGNMGRVVSF
>CAB4198187.1_uncultured_Caudovirales_phage
------------D-HLKFL------LD-------------------------------TE-----GNIFLTGKAGTGKSTLINQFCEK---F-G-----K------------------
-----T-NKIVKLAPTGIAAYNI-------GG------QTIHSF-FKF---K--I------TT-T---I---D-------SV---Y----------------------------F---
-E----E---E-------------LAKIC--KS--VKVIIIDEVSM--LRPDLLDC-IDQSLRLHT----------------G---------K----------------------SK-
--------------------------SPFGD-----IKMIFVGDLYQLEPVV-----------------KS---G--E-L------------------------------L------
D------------------------YET-----K-----------YFFSA------------------KVFKY----SK--L-KIKELDKIHR----Q-N-----------D--P---
----------VFIDFLNKVRLG---N-L-S------Y--------------LE------------------------------LN--QL-NH--LL---------STSL---------
-----------------------------------------------------------------------------------------------D---------R-E----------
-L-------------------ITLTTTNYKSLIINNENLL-KN----K---H--------------------------P-----------------------E-E-F-S---------
-----------S---------------------AVI------D----------------------G--------------------E-------------------------------
------F----------------------------------------------------------N------------------------------------S-----------N-N-
-------ILA----------EEEL-----VLKYDCKVMILANGT----------CFDDPN--K-----A-----YFNGSIGLFRGF
>DAM57115.1_Myoviridae_sp.
-----TYEIGTDA-ALAAV------L--------------------------------CG-----ENVYISGPGGTGKTHLLQDIQSL---L-G------------------------
-------ESCMVVAPTGVAALNA-------GG------VTAHRA-FDL---S--A------GV-T---V---P-------ED---FT------------------------E--I---
-R----S---K-------------TAKPLKSKA--LRTLVIDEVSM--VRADKFVE-MDKKLQHLR----------------K---------T-----------------------S-
--------------------------EPFGG-----LQVIMFGDFYQAQPVI-----------------ST---Q--E-RED----------------------------Y-------
------YKY---------------WDT-----D-----------LCFYT------------------QSWKD----LN--L-KCVALVEQFR----Q-E-----------S--I----
---------RFATMLNCVREG---R-R-T------G--------------DV------------------------------VK--EL-NS--RC---------YHGG-----Q---
----------------------------------------------------------------------------------------------A---------S-D-----------
A-------------------IILCSTNKRVEEINREFYD-RI----D---G--------------------------E-----------------------E-R-M-Y----------
----------K---------------------GTL------K----------------------G--------------------K--------------------------------
-----F----------------------------------------------------------------------------------------------PP-----------N-Q--
------LPV----------EDLM-----CLKVGMKVMIVANDL---------------N--P-----NHKVPCYVNGSRGTILKF
>YP_007006388.1_Escherichia_phage_FV3
-----QYDVGTDA-ALVAI------M--------------------------------SG-----ENVFVSGPGGTGKTYLINMIQSM---Y-G------------------------
-------DSCITVAPTGVAALNV-------NG------ATAHRT-FDL---A--A------GV-S---M---E-------SD---WT------------------------A--I---
-R----A---K-------------TAKPLKSKA--FTILIIDEISM--IRADKFIE-MDRKLRFLR----------------K---------N-----------------------D-
--------------------------KPFGG-----IQVLLFGDFYQAPPVV-----------------SS---M--E-KEA----------------------------Y-------
------FNF---------------YHT-----D-----------LCCYT------------------ESWED----LN--L-HNIALVDQFR----Q-E-----------S--V----
---------RFATMLNCVREG---R-R-I------K--------------EV------------------------------VA--EL-NT--RC---------YHGG-----V---
----------------------------------------------------------------------------------------------P---------T-D-----------
A-------------------LTICATNKQAEEVNRRFYD-AI----K---A--------------------------P-----------------------E-K-T-Y----------
----------I---------------------GKM------K----------------------G--------------------K--------------------------------
-----F-----------------------------------------------------------------------------------------------P-----------S-T--
------LPV----------EQEM-----RLKIGMKVMITSNDV---------------D--P-----THKVPYYVNGTRATVVKF
>AXC42890.1_Escherichia_phage_LL12
-----QYDVGTDA-ALIAI------M--------------------------------SG-----ENVFVSGPAGCGKTYLINMIQSM---Y-G------------------------
-------DSCITVAPTGVAALNV-------NG------ATAHRT-FDL---A--A------GV-S---M---E-------SD---WT------------------------A--I---
-R----A---K-------------TAKPLKSKA--FTILIIDEISM--IRADKFIE-MDRKLRFLR----------------K---------N-----------------------D-
--------------------------KPFGG-----IQVILFGDFYQAPPVV-----------------SS---M--E-KEA----------------------------Y-------
------FNF---------------YHT-----D-----------LCCYT------------------ESWKE----LD--L-HNIALVDQFR----Q-E-----------S--I----
```

```
---------RFATMLNCVREG---R-R-I------K--------------EV------------------------------VA--EL-NA--RC---------YKGG-----V----
----------------------------------------------------------------------------------------------P---------T-D-----------
A-------------------LTICATNKQAEEVNRRFYD-AI----N---A--------------------------P-----------------------E-K-V-Y----------
----------T---------------------GEM------K----------------------G--------------------K--------------------------------
-----F-----------------------------------------------------------------------------------------------P-----------S-T--
------LPV----------EQEM-----KLKIGMKVMITANDV---------------D--P-----THKVPYYVNGTRATIVKF
>YP_009851551.1_Erwinia_phage_pEp_SNUABM_01
-----SFGVGTDA-AVKAV------L--------------------------------GG-----DNVFVTGPGGTGKTHTIKKIQAL---Y-P------------------------
-------DSTLTVAPTGVAALNV-------EG------MTAHRA-FGL---S--M------GV-S---T---D-------EC---VS------------------------D--I---
-K----K---R-------------HEKLMKSRD--LERIIIDEISM--IRADKLWE-IDEKLKLVR----------------K---------N-----------------------P-
--------------------------KPFGG-----LQMIMFGDFFQNLPVL-----------------TN---A--E-EDL----------------------------Y-------
------RGL---------------FNT-----E-----------LSCWS------------------DTWKN----AQ--M-YPVLLEKMYR----Q-Q-----------S--D----
---------NFARMLNCLRRG---E-R-L------D--------------DV------------------------------VA--YI-ND--NC---------YKPL-----N----
----------------------------------------------------------------------------------------------N---------P-Q-----------
A-------------------ITLTSTNAQAERINKKFFD-EI----K---S--------------------------P-----------------------V-K-V-F----------
----------K---------------------SKV------E----------------------G--------------------D--------------------------------
-----F-----------------------------------------------------------------------------------------------------------K-S--
------RPG----------PDEL-----ELKEGLKVMITANQM----------CQPNED--P-----A-----YVNGSIGFIKKM
>YP_009854417.1_Erwinia_phage_Hena1
-----AFGVGTDD-AIKAV------M--------------------------------GG-----GNVFVTGPGGTGKTHTIKKIQAL---F-P------------------------
-------DTTLTVAPTGVAALNV-------EG------MTAHRA-FGL---S--M------GV-S---S---D-------ED---VM------------------------N--I---
-K----R---R-------------HEKLMKSKD--LERIIIDEISM--IRADKLWE-IDQKLRLVR----------------K---------K----------------------PN-
--------------------------EPFGG-----IQVIKFGDFFQNLPVL-----------------TS---T--E-EDL----------------------------Y-------
------RSH---------------FNT-----E-----------LCCWS------------------DTWRD----AQ--P-YPVMLEKMYR----Q-Q-----------S--D----
---------NFARMLNCLRKG---E-R-L------D--------------DV------------------------------VD--YI-ND--NC---------YKPL-----D----
----------------------------------------------------------------------------------------------N---------P-Q-----------
A-------------------ITLTSTNAQTERINKKFFD-DI----Q---S--------------------------P-----------------------V-K-I-Y----------
----------K---------------------SKV------E----------------------G--------------------D--------------------------------
-----F-----------------------------------------------------------------------------------------------------------K-S--
------KPG----------PDEL-----ALKVGLKVMITANQI----------SKPHED--P-----A-----YVNGSIGFIRKM
>YP_008857219.1_Escherichia_phage_4MG
-----DFGVGVEA-ALGAI------M--------------------------------SG-----DNVFVTGPGGSGKSYTIKTIQSL---Y-A------------------------
-------GSVLTVAPTGAAAINV-------DG------MTAHRA-FRL---S--M------GV-A---T---Q-------KD---AE------------------------E--L---
-K----P---K-------------VKRLLKSKA--LKIIIIDEISM--FRADKLWE-MDMKCRAAR----------------R---------Q----------------------PN-
--------------------------KPFGG-----LQICMFGDFFQNPPVL-----------------TE---T--E-KEM----------------------------Y-------
------FQF---------------HPT-----E-----------LCCFS------------------DTWQE----LN--P-YPVLLEKIYR----Q-N-----------S--R----
---------RFSDILNLLRRG---Q-R-I------P--------------EI------------------------------VR--EL-NG--VC---------YRGG-----E----
----------------------------------------------------------------------------------------------A---------IPD-----------
A-------------------ITITSTNAAAEKVNRKRFE-EI----P---G--------------------------L-----------------------P-V-L-Y----------
----------T---------------------AKK------N----------------------G--------------------D--------------------------------
-----F-----------------------------------------------------------------------------------------------------------T-Q--
------KPV----------PEEL-----YLKEGAKVMITVNDP----------K-GFEE--P-----E-----YVNGSRGEVIEL
>QPI14547.1_Salmonella_phage_GEC_vB_MG
-----DFGVGVEA-ALGAI------M--------------------------------SG-----DNVFVTGPGGSGKSYTIKTIQSL---Y-A------------------------
-------GSVLTVAPTGAAAINV-------DG------MTAHRA-FGL---T--M------GV-A---T---K-------KD---TE------------------------E--I---
-K----P---K-------------VKRLLKSKA--LKIIIIDEISM--FRADKLWE-MDMKCRLAR----------------R---------Q----------------------PN-
--------------------------KPFGG-----LQICMFGDFFQNPPVL-----------------TE---A--E-KEM----------------------------Y-------
------FQF---------------HNT-----E-----------LCCFS------------------DTWQE----LN--P-YPVILEKVYR----Q-N-----------S--V----
---------HFSTMLNCLRRG---Q-R-I------P--------------EI------------------------------VQ--FM-NT--HC---------FDNG-----K----
----------------------------------------------------------------------------------------------P---------L-D-----------
A-------------------ITITSTNAAADKVNKKRFE-EV----P---G--------------------------M-----------------------P-T-L-Y----------
----------A---------------------AKK------T----------------------G--------------------D--------------------------------
-----F-----------------------------------------------------------------------------------------------------------S-Q--
------KPV----------PEEI-----YLKEGAKVMITVNDP----------K-GFDE--P-----E-----YVNGSRGVIIEL
>YP_009835918.1_Raoultella_phage_Ro1
-----DFGVGTEE-ALGAI------I--------------------------------EG-----KNVFITGPGGSGKSHLIKTIQSL---Y-S------------------------
-------SSTLTVAPTGVASLNV-------DG------MTTHRA-FGL---S--M------GI-A---T---E-------DD---GK------------------------T--V---
-K----T---K-------------PKKLLKSKS--LERIIIDEISM--VRADKLWE-MDQKLRVAR----------------R---------E----------------------PK-
--------------------------KAFGG-----LQVIMFGDFFQNPPVL-----------------TD---S--E-ENA----------------------------Y-------
------FEL---------------HST-----E-----------LSCFS------------------DTWRE----IN--P-YPVLLDKIYR----Q-N-----------S--V----
---------HFSSLLNHMRKG---E-R-I------D--------------EI------------------------------VK--FL-NN--QC---------YSKG-----A----
----------------------------------------------------------------------------------------------A---------L-N-----------
A-------------------ITLTSTNAAAERINKKHYD-QI----Q---G--------------------------E-----------------------E-V-I-Y----------
----------K---------------------ASK------T----------------------G--------------------D--------------------------------
-----F-----------------------------------------------------------------------------------------------------------A-Q--
------RPV----------AESL-----HLKVGTRVMITVNDQ----------NPDEDG--P-----K-----FVNGTRGIIKAL
>AXN57909.1_Acinetobacter_phage_ABPH49
-----SYEIGSES-AIKSI------M--------------------------------SG-----KNTFITGPGGSGKSQIIHTVQDM---L-G------------------------
-------ESSLSLAPTGIAALNI-------NG------MTAHRA-MGL---S--M------GV-T---M---D-------ED---IT------------------------K--V---
-R---SN---K-------------QAKLLSSPA--IKRIILDEVSM--IRADKLYE-MDHKFRHFR----------------K---------N-----------------------S-
--------------------------KPFGG-----LQVVAFGDGFQISPVL-----------------TQ---R--E-AMD----------------------------F-------
------RNL---------------YGS-----E-----------IPFDS------------------QTWSE----AG--F-HNILLDKVWR----Q-E-----------D--K----
---------EFSGALNNLRVG---R-D-I------D--------------AA------------------------------IA--FI-NN--RC---------ANKG-----I----
----------------------------------------------------------------------------------------------H---------S-D-----------
A-------------------VTLTSTNKLADEINLREFN-AL----P---G--------------------------K-----------------------K-S-T-H----------
----------R---------------------ASI------L----------------------G--------------------D--------------------------------
-----F-----------------------------------------------------------------------------------------------------------K-D--
------RPV----------AEVL-----ELKEGLKVMITANDQ----------A---VP--S-----R-----YVNGTVGIVRRM
```

```
>YP_009042324.1_Cronobacter_phage_CR8
-----KTVIGNEE-AMRAI------M--------------------------------AG-----DNVFITGPGGSGKSLMIATLREF---F-A------------------------
-------DSFLFVAPTGIAALNI-------SG------ITAHKA-FGL---T--F------GV-T---T---K-------ED---YK------------------------A--K---
-S----K---K-------------PAMLMASNA--LDAIVFDEISM--IRSDKLRE-IDMKLRYHR----------------K---------V-----------------------N-
--------------------------KPFGG-----LQVIMFGDGFQIKPVL-----------------KR---E--E-TAM----------------------------F-------
------REL---------------HGN-----E-----------IPFGS------------------DIWNQ----LD--F-TNAYLPKVHR----Q-S-----------D--P----
---------VFAQHLNNIRVG---N-N-V------G--------------AA------------------------------VD--YF-NQ--HC---------FGPAL---------
--------------------------------------------------------------------------------------------------------P-G-----------
A-------------------VTLTTTNKLAEEINQREFE-KI----K---A--------------------------Q-----------------------P-H-V-F----------
----------E---------------------AKI------S----------------------G--------------------E--------------------------------
-----F-----------------------------------------------------------------------------------------------------------P-E--
------RPV----------NEVL-----NLKEGLKVMIVVNDN----------DQKKKE--P-----D-----YVNGTVGIIKKI
>QEG12160.1_Klebsiella_phage_vB_KaeM_KaOmega
-----KTVIGNEE-AMRAI------M--------------------------------AG-----ENVFITGPGGSGKSLMIATLREF---F-A------------------------
-------DSFLFVAPTGIAALNI-------NG------ITTHKA-FGL---T--F------GV-T---T---K-------ED---YK------------------------A--K---
-S----K---K-------------PAMLMASDA--LDAIVFDEISM--TRSDKLRE-IDMKLRYHR----------------K---------V-----------------------N-
--------------------------KPFGG-----LQVIMFGDGFQIKPVL-----------------KR---E--E-TAM----------------------------F-------
------REL---------------HGN-----E-----------IPFGC------------------DIWNE----LN--F-TNAYLPKVHR----Q-T-----------D--P----
---------VFAEHLNNIRVG---N-N-V------G--------------AA------------------------------VD--YF-NE--KC---------FGPAL---------
--------------------------------------------------------------------------------------------------------P-G-----------
A-------------------VTLTTTNKLAEQINQREFE-KI----K---A--------------------------E-----------------------P-H-V-F----------
----------E---------------------AKI------S----------------------G--------------------E--------------------------------
-----F-----------------------------------------------------------------------------------------------------------P-E--
------RPV----------NEVL-----NLKEGLKVMIVVNDN----------DQKKKE--P-----D-----YVNGTVGIIKKI
>QUL77343.1_Escherichia_phage_UPEC06
-----KTVIGNEA-AMQAI------M--------------------------------NG-----ENVFITGPGGSGKSMMIAALRDF---F-A------------------------
-------DSFLFVGPTGVASLNI-------RG------VTTHKA-FGL---T--F------GV-T---T---S-------ED---YK------------------------A--K---
-S----K---K-------------AAMLMASDA--LDGIVFDEIGM--TRSDKLKE-IDMKLRHHR----------------K---------S-----------------------D-
--------------------------KPFGG-----LQIIMFGDGFQIKPVL-----------------KK---E--E-VPL----------------------------F-------
------REL---------------HGK-----E-----------IPFGC------------------ETWES----LN--L-TYAYLPKVHR----Q-S-----------D--P----
---------VFAEHLNNIRIG---K-N-I------P--------------DA------------------------------VN--FF-NQ--RC---------FGRPL---------
--------------------------------------------------------------------------------------------------------D-G-----------
A-------------------VTLTTTNKLAEEINNKEYA-KI----N---A--------------------------E-----------------------Q-H-E-F----------
----------N---------------------ARI------T----------------------G--------------------E--------------------------------
-----F-----------------------------------------------------------------------------------------------------------V-E--
------RPV----------NEKL-----FLKEGLKVMIVVNDN----------DQKKKV--P-----D-----YVNGTVGIVRRI
>YP_009595586.1_Pseudomonas_phage_pf16
-----QLNKKQTY-AFEQI------M--------------------------------RG-----HNVYVSGPGGVGKSVLISKIRDL---C-E------------------------
-------DDTIFLAPTGIAALNI-------KG------ATIHRT-FKL---Q--L------GY-L---D---P-------AQ---RS------------------------R--V---
-N----E---K-------------VRELFSDDS--IKRIVIDEISM--VRGDIFTA-VDKALRLAK----------------R---------R-----------------------N-
--------------------------KPFGG-----LQVIVVGDFFQLSPVL-----------------NER-SQ--E-GEL----------------------------Y-------
------LKE---------------FSS-----P-----------FCFDT------------------DAWRE----AG--F-QTIELDEIMR----Q-S-----------D--A----
---------KFIGALNSIRTR---A-D-D-----FE--------------TA------------------------------LD--FL-NR--IG---------MEKE-----DV---
----------------------------------------------------------------------------------------------P---------D-D-----------
T-------------------LFLCSTNKEADAVNKHNYD-DV----M---G--------------------------E-----------------------E-R-L-Y----------
----------Y---------------------GKK------K----------------------G--------------------P--------------------------------
-----F-----------------------------------------------------------------------------------------------------------R-D--
------LPV----------PEVL-----SLKVGVKVLICANAE---------------D--G-----S-----YYNGMTGYVEKM
>QDP60500.1_Prokaryotic_dsDNA_virus_sp.
--------DLQQE-AIRRI------N--------------------------------GG-----ESIFLTGSGGVGKSWVIDQVNDE------------------------------
--------NTLLCAPSGIAALNI-------GG------ITCHRA-FAL---P--L------GI-P---T---D-------ED---FY------------------------K--I---
-P----R---Y-------------MWDTFSGNA--VKRIIIDEVGM--LRTDYFVL-ISRRLQRIR----------------G---------N-----------------------D-
--------------------------LPFGG-----IQVVLVGDFYQLEPIL-----------------KH---S--E-QEY----------------------------FD------
------------------------FAS-----K-----------FCFGS------------------KLWD---------F-PTIELTDVKR----Q-S-----------N--K----
---------RQVLMLNSIRRK---D-K-H-----YK--------------KA------------------------------LE--YI-QK--EC---------KPYE-----P----
----------------------------------------------------------------------------------------------D---------D-N-----------
T-------------------LHLCCYNKDADYINQLYYS-KM----E---G--------------------------E-----------------------E-R-C-F----------
----------Y---------------------AKI------P----------------------P--------------------N--------------------------------
-----W----------------------------------------------------------G------------------------------------------------K-E--
------RPV----------DEVV-----RLKVGTRVLIAQNCP---------------Q--G-----T-----YVNGDRGVVVGF
>QDP64781.1_Prokaryotic_dsDNA_virus_sp.
-----NNSVEQDL-ALKYI------L--------------------------------SG-----ENVLITGSGGVGKSHIIKQILDP------------------------------
--------NTLLCAPTGIAALNI-------GG------ATCHRT-FGL---P--L------GV-P---T---I-------QD---FM-------------------TAS--R--K---
-V----Q---D-------------LFNPF--SP--IKRIVIDECSM--LRMDQLEL-INSKLQMIR----------------G---------N-----------------------K-
--------------------------KPYGG-----LQLVLVGDYFQLDSVI-----------------TS---Y--E-EKA----------------------------Y-------
------YSQ---------------YSS-----P-----------FNFKS------------------DIFD---------F-KVVELTTVFR----Q-E-----------D--K----
---------RQVDMLNSIRRK---D-K-Y-----YK--------------YA------------------------------LD--AI-VA--EA---------LPYE-----P----
----------------------------------------------------------------------------------------------S---------P-D-----------
V-------------------TVMCCYKADVRKYNRRYFK-ML----D---T--------------------------P-----------------------I-F-E-F----------
----------N---------------------AKI------E----------------------N--------------------V--------------------------------
-----L----------------------------------------------------------TE----------------------------------DK----------WN-D--
------SAV----------PHKI-----ELREGCKVMFKANDL---------------H--G-----E-----YVNGEKGTVSYV
>AUR91792.1_Vibrio_phage_1.164
-----DLKLKQQE-ALGLM------K--------------------------------TG-----ANVFLTGKAGTGKSFVTDLFTAW---A-E--E--Q------------------
-----D-KNILICAPTGIAALNI-------GG------ATIHRT-FKL---P--I------NY-V---S---D-------SS--HL-------------------------Y--S---
-S----N---E-------------GRALI--EA--ADIVLIDEVSM--LRADTFSH-VEYKMRESV----------------L---------S-----------------------G-
--------------------------SAFGG-----KQIIVVGDFYQLPPVI-----------------KN---E--E-RAD----------------------------L-------
------KAH---------------FGG-----H-----------YAFEC------------------QAWKD----AK--F-KMVELDEVVR----Q-S-----------D--L----
```

```
---------EFVDALNSIREK---N-VHS------N--------------KS------------------------------LG--YV-NH--NA---------NGDL-----M----
----------------------------------------------------------------------------------------------E---------N-D-----------
V-------------------VTLCFTNKVAEQINKKELA-KI----E---T--------------------------L-----------------------P-V-E-F----------
----------I---------------------ASV------S----------------------G--------------------T--------------------------------
-----V----------------------------------------------------------K------------------------------------E-----------S-E--
------KPV----------PEHL-----ELKVGAKVIFCVNDQ---------------D--G-----R-----FVNGTTGYVTGF
>AUR89562.1_Vibrio_phage_1.124
-----DLKLKQQE-ALGLM------K--------------------------------TG-----ANVFLTGKAGTGKSFVTDLFTEW---A-E--E--Q------------------
-----D-KNILICAPTGIAALNI-------GG------ATIHRT-FKL---P--I------NY-V---S---D-------SS--HL-------------------------Y--S---
-S----N---E-------------GRALI--EA--ADIVLIDEVSM--LRADTFSH-VEYKMRESV----------------L---------S-----------------------G-
--------------------------SAFGG-----KQIIVVGDFYQLPPVI-----------------KN---E--E-RAD----------------------------L-------
------KAH---------------FGG-----H-----------YAFEC------------------QAWKD----AK--F-KMVELDEVVR----Q-S-----------D--I----
---------EFVDALNSIREK---N-VHS------N--------------KS------------------------------LG--YI-NH--NA---------SGDL-----M----
----------------------------------------------------------------------------------------------E---------N-D-----------
V-------------------VTLCFTNKVAEQINKKELA-KI----E---A--------------------------L-----------------------P-V-E-F----------
----------I---------------------ASV------S----------------------G--------------------T--------------------------------
-----V----------------------------------------------------------K------------------------------------E-----------S-E--
------KPV----------PEHL-----EIKVGAKVIFCVNDQ---------------D--G-----R-----FVNGTTGYVTGF
>DAE75004.1_Bacteriophage_sp.
-----KLNKKQRY-ALDTM------L--------------------------------SG-----SNVFLTGDAGTGKTTVIQTFIDE---A-E--K--A------------------
-----G-KSVLVSATTGIAADNIGY-----GA------TTVHRA-LNI---S--I------KF-----------------ED---YK------------------------K--K---
-V----K---S-------------RAELL--KE--ADILIIDEISM--CRFDLFNM-IAKTIITEN----------------E---------E--------------------RAVD-
--------------------------RLLSGEDKEDVQLIVIGDFYQLPPVI-----------------TT---D--D-RKI----------------------------L-------
------CRMYGSDY-----GKGGKYEH-----G-----------YAFMS------------------EYWKE----IG--F-EYIKLDDVCR----Q-N-----------D--E----
---------GFKYVLNDIKYG---N-N-I------R--------------KS------------------------------IA--YL-EN--NE---------SDKVI---------
--------------------------------------------------------------------------------------------------------P-E-----------
A-------------------PFLVGTNAEADRINNTFLG-KL----D---K--------------------------KT----------------------E-K-V-F----------
----------H---------------------AAV------D----------------------G--------------------D--------------------------------
-----L----------------------------------------------------------TS------------AD--------------------IK-----------N-I--
------AFA----------REDL-----ILNIGAKVMITVNDL---------------S--G-----N-----YVNGTIGIIQKI
>DAF60248.1_Siphoviridae_sp._ctqK313
-----KLNKKQRY-ALDTM------L--------------------------------SG-----SNVFLTGDAGTGKTTVIQTFIDE---A-E--K--A------------------
-----G-KSVLVSATTGIAADNIGY-----GA------TTVHRA-LNI---S--I------KF-----------------ED---YK------------------------K--K---
-V----K---S-------------RAELL--EE--ADILIIDEISM--CRFDLFNM-IAKTIITEN----------------E---------E--------------------RAVD-
--------------------------RLLSGEDKEDAQLIVIGDFYQLPPVI-----------------TT---D--D-RKI----------------------------L-------
------CRMYGSDY-----GKGGKYEH-----G-----------YAFMS------------------EYWKE----MG--F-EYIKLDEVCR----Q-N-----------D--E----
---------GFKYVLNDIKYG---N-N-I------R--------------KS------------------------------IA--YL-EN--NE---------SDKVI---------
--------------------------------------------------------------------------------------------------------P-E-----------
A-------------------PFLVGTNAEADRINNTFLG-KL----D---K--------------------------KT----------------------E-K-V-F----------
----------H---------------------AAV------D----------------------G--------------------E--------------------------------
-----L----------------------------------------------------------TS------------AD--------------------IK-----------N-I--
------AFA----------REDL-----ILNIGAKVMITINDL---------------S--G-----N-----YVNGTIGIIQKI
>DAP73423.1_Bacteriophage_sp.
-----DLNKKQRY-ALDTM------L--------------------------------SG-----SNVFLTGDAGTGKTTVIQTFIDE---A-E--K--A------------------
-----G-KNILVSATTGIAADNIGY-----GA------TTVHRA-LNI---S--I------KF-----------------ED---YK------------------------K--K---
-V----K---S-------------RAELL--KE--ADVLIIDEISM--CRFDLFNM-IAKTIITEN----------------E---------E--------------------RAVD-
--------------------------RLLIGEDKEDIQLIVIGDFYQLPPVI-----------------TT---D--D-RKI----------------------------L-------
------CRMYGSDY-----GKGGKYEH-----G-----------YAFMS------------------EYWKE----MG--F-EYIKLDEVCR----Q-N-----------D--E----
---------GFKYVLNDIKYG---N-N-I------R--------------KS------------------------------IA--YL-EN--NE---------SDKVI---------
--------------------------------------------------------------------------------------------------------P-E-----------
A-------------------PFLVGTNAEADRINNTFLG-KL----D---K--------------------------KT----------------------E-K-V-F----------
----------H---------------------AAV------D----------------------G--------------------E--------------------------------
-----L----------------------------------------------------------TS------------AD--------------------IK-----------N-I--
------AFA----------REDL-----ILNIGAKVMITVNDL---------------S--G-----N-----YVNGTIGIIQKI
>DAE16492.1_Siphoviridae_sp._ctqBH20
-----KLNKKQRY-ALDTM------L--------------------------------SG-----SNVFLTGDAGTGKTTVIQTFIQE---A-E--E--M------------------
-----G-KNVLVSATTGIAADNIGY-----GA------MTVHRA-LNI---S--V------RF-----------------EE---YR------------------------K--K---
-V----K---S-------------RVDLL--KE--ADVLIIDEISM--CRFDLFNV-IAKTIFLEN----------------E---------E--------------------RAVE-
--------------------------RLLKGEDKEDLQLIVIGDFYQLPPVI-----------------TQ---S--D-RNI----------------------------L-------
------CRMYGSEY-----GKGGKYEQ-----G-----------YAFMS------------------AYWKD----MG--F-EYIKLDEVCR----Q-N-----------D--E----
---------GFKYVLNDIKYG---N-N-I------R--------------KS------------------------------IS--YL-EK--NE---------AGKVI---------
--------------------------------------------------------------------------------------------------------P-E-----------
A-------------------PFLVGTNAEADRINQTFLN-KL----K---K--------------------------ET----------------------E-R-V-F----------
----------H---------------------AQV------S----------------------G--------------------E--------------------------------
-----L----------------------------------------------------------ES------------AD--------------------IR-----------N-I--
------NFA----------KEEL-----ILNIGAKIMITVNDL---------------S--G-----D-----YVNGTIGIIQKI
>DAG97916.1_Ackermannviridae_sp.
-----LLTEEFKK-AYDLL------EH-------------------------------TK-----EFVFLTGDAGSGKTTFLKWWLSN---T-S------------------------
-------KKTVVLSPTGMGAVNLL----PIRA------STIHKF-FKF---G--N------KP-L---F---T-------SN---I--------------------PRLSSK--K---
-Y----K---E-------------NRQLY--LN--VDTIIIDECSM--VSSMMMQA-IDDFYRINF----------------D---------S-----------------------D-
--------------------------EPFGG-----KQIVLVGDMAQLPPVI-----------------GS--DA--E-RQY----------------------------T-------
------KDR---------------FGG-----K-----------YFFDA------------------TIFKE----VN--I-KFVEFTEIFR----Q-N-----------D--P----
---------EFIGYLNKIRTG---T-I-T------Q--------------SD------------------------------II--KL-ND--IF--------TSNKV----------
----------------------------------------------------------------------------------------------S---------D-D-----------
A-------------------MVISFRNDVVDMINDYKLN-EI----K---A--------------------------E-----------------------D-V-F-L----------
----------Y---------------------SSI------N----------------------G--------------------F--------------------------------
-----F----------------------------------------------------------NP-----------------------------------------------K-S--
------CPV----------KEIT-----RVRPGCRIMCRNNDK---------------D--E-----R-----WVNGTIAKFVKK
```

```
>DAX71650.1_Myoviridae_sp.
-----QINEKFLL-AEKGI------D---------------------------------AG-----HNLLILGVAGTGKSTFLYYMNKK---F-E--S--Q-------------------
-----G-KKVVYLAPTGIASINMAQ--RTGSA------QTLHSY-FKI---P--I------GG-E---L---S-------AN----S---------------------VKV-----L---
-K----E---E-------------EAKLF--KE--VDIIVVDEISM--CRSDVLNY-IDLFLKYNT------------------E---------N-----------------------F-
--------------------------EPFGG-----KQMVFLGDVLQLAPVV------------------AT---I--E-EKLY----------------------------L-------
------KHT---------------FGG-----D------------WFWNT------------------PGFKA----GK--F-KLVQFTKKYR----Q-A-----------ED--S----
---------KFAIWLDKIRTG---E-I-T------S---------------DE-------------------------------LS--EL-NQ--II---------VSPP----------
------------------------------------------------------------------------------------------------N---------P-Q-----------
A-------------------ITLCTTNATADRINTVALE-NI----D---S---------------------------P------------------------L-Y-E-H----------
----------L---------------------GKI------NNI--------------------TG---------------------D--------------------------------
-----------------------------------------------------------------K------------------------------------IE----------WS-A--
------FPV----------DYKF-----KYKLGCKVMIRKNGE-------------------------G-----YSNGSIGTIVKI
>ARR75030.1_Mimivirus_AB566O17
-----TLSESQER-VLELV------R---------------------------------KG-----NNVLILGSAGCGKSTVIKEIKSE---F-K-------------------------
-----N-KKVYITSTTGISAYNI-------QG------VTLHSF-MGF---G--T------GE-G---P---L-------HT---L-----------------L----SRI--R--R---
-R----K---G-------------YTQRL--IE--TEILIVDEISM--MSAELFEK-VDTILREIR------------------R---------I-----------------------Q-
--------------------------LPFGG-----IQMIFSGDLLQLKPVI------------------KE---T--E-WNP------------------------------------
D------------------------PDQ-----R------------LIFES-----------------SRFSE----Y---F-ETVVLTTNFR----Q-Q------------HD--V---
----------IYQGLLTNIRRN---T-L-T------S---------------GD-------------------------------LQ--LL-TD--CL---------GKKP---------
-------------------------------------------------------------------------------------------------P---------K-G----------
-V-------------------PFLVPTNKAAGEINKRETL-KL----K---T---------------------------P------------------------K-F-T-Y---------
-----------T----------------------TVF-----Q-----------------------K---------------------E-I------HTSSHDETLS---------D-M-
------Y---------------------------------------------------------LAE------------------------------------LK-----------N-Q-
-------FKQK--------DLDEL-----VLRAGSRVMLTRNLD-----------V----S--S-----G-----LVNGALGTISSA
>QHN71346.1_Mollivirus_kamchatka
-----KLSDGQKE-ALEVA------K---------------------------------RG-----DNLSVSGSGGTGKSLAAKMLIGT---L-M-LE--K-------------------
-----R-KTVRVVASTGAAALLI-------GG------DTAHSA-LGD---G--I------NP-D---D---PV------KS-------------------------AVRLLSK--N---
-K----G---K-------------KADYW--CD--TEVLFFDEVGM--IEPVFFEW-MAMTVGHIR------------------A---------R----------------------
RKVK--GSF----------FTGADGKRVVRPFGG-----IQVIVFGDFLQLQPIV-------------------KG---I--P-RGY----------------------------I----
---------PTV---------------TDG---LLE------------FPFQL------------------DVWRE----LD--F-HCIELTHVFR----Q-S------------D--R-
------------PFVAALNDIRFG---R-V-T------P---------------HA-------------------------------AR--MF-ES--CV---------GRRF-----
ED-----------------------------------------------------------------------AED-DE----------------------S---------K-R------
-----P-------------------TRIYTKRDKVNEYNKMMME-KL----P---K----------------------------P-----------------------E-K-R-Y-----
---------------R----------------------GTI------QYDYE------------------PG--------------------A-------LADYETKS----------
RFG-R-----HAF-----------------------------------------------------------N------------------------------------LK---------
--K-H--------CRV----------PEEL-----ALRKGALVMLTRNLR-----------Y-------G-----G-----LTNGSVGVVVGF
>YP_009165351.1_Mollivirus_sibericum
-----KLSDGQKE-ALEVA------K---------------------------------RG-----DNLSVSGSGGTGKSLAAKMLIGT---L-M-LE--K-------------------
-----R-KTVRVVASTGGAAQLI-------GG------DTAHSA-LGD---G--I------NP-D---D---PV------KS-------------------------AVRLLSK--N---
-K----G---K-------------KADYW--CD--TEVLFFDEVGM--IEPVFFEW-MAMTVGHIR------------------A---------R----------------------
RKVK--GSF----------FTGADGKRVVRPFGG-----IQVIVFGDFLQLQPIV-------------------KG---I--P-RGY----------------------------I----
---------PTV---------------TDG---LLE------------FPFQL------------------DVWRE----LD--F-HCIELTHVFR----Q-S------------D--R-
------------PFVAALNDIRFG---R-V-T------P---------------HA-------------------------------AR--MF-ES--CV---------GRRF-----
ED-----------------------------------------------------------------------AED-DE----------------------S---------K-R------
-----P-------------------TRIYTKRDKVNEYNKMMME-KL----P---K----------------------------P-----------------------E-K-R-Y-----
---------------R----------------------GTI------QYDYE------------------PG--------------------A-------LADYETKS----------
RFG-R-----HAF-----------------------------------------------------------N------------------------------------LK---------
--K-H--------CRV----------PEEL-----ALRKGALVMLTRNLR-----------Y-------G-----G-----LTNGSVGVVVGF
>C_sorokiniana
-----PSAKNSSE-AVKAL------AL--------------------------------QG-----RNIFLTGCGGSGKSYWIKHMIAH---W-E--R--E-------------------
-----G-KEVALAAMTGCAAELI-------GG------RTLHSC-LQL---G--L------VT-K---V-----------GD---V----------------V-----NVS--S--K---
-K----K---R-------------LVEKL--AF--LDVLICDEVSM--LSAELFQF-IVEQISLAR------------------A---------THFRQELQRLGAGSAGAERLRRLLA-
--------------------------QPLSG-----LQLILVGDFLQLPPVD------------------KG---P--E-DCQRAAELA----------VQNLEEK------L-------
---IKNKQVTGAQ----VRSNSAVCNR-----G------------LCFQS------------------EAWRR----LD--M-HVAVLKQVHR----Q-S------------E--R----
---------EFISILHAIRDG---S-A-T------R---------------PQ-------------------------------LD--RL-WQ--LC---------SRPL-----P----
----------------------------------------------------------------------Q-ND----------------------G---------I-V-----------
P-------------------TTLYCKNINANERNAAELA-KL----P---T---------------------------K------------------------Q-F-E-F----------
----------H----------------------AAH------RVV-----PKVR----------KG--------------------ESPS----CQAVQRREA----------RLR-E--
-----M----------------------------------------------------------MPD-------------V---------------------LK-----------E-D--
------RKV----------RDLI-----LLKEGAQVMCTANIM----------------S--G-----T-----LVNGSRGVVVGF
>ARX71979.1_Erinnyis_ello_granulovirus
-----TLNEKQQK-LFDYL------TQT----K--------------------------SF-----APVFVSGSAGTGKSALLVALREH---W-L--K--Q-------------------
-----D-KIVFVAAYTHLAARNI-------NG------KTCHSL-FRF---D--F------EL-N---L---L-------R---------------------------------------
----------A-------------Q---I--GV--PHYLIIDEISM--VPEKMLDG-IDSRLRQTS------------------G---------K----------------------FS-
--------------------------LPFGG-----VNVVVFGDLYQIPPVD------------------KH------------------------------------------------
-------------------------HL-----L------------PPYKA------------------DIWYY--------F-ELYELTENMR----Q-S------------E--P----
---------EFIANLNMLRVG---D-V------------------------KC-------------------------------LS--YF-NR--FV---------VNAE----------
-------------------------------------------------------------------TQNI-QD----------------------C---------V-N-----------
C-------------------TSLVSTHKEANDLNKKCYA-HI----V---G---------------------------D----------------------GEE-M-V-C----------
----------T----------------------VKE------TKG------------------------------------------R--------WNR----------------D-M--
---VVF-----------------------------------------------------------N--------------E---------------------EQ-----------A-Q--
------LIF----------GESI-----KVCVGARVMITHTTD-------------------------T-----FCNGDLGVVRSF
>YP_009506054.1_Clostera_anastomosis_granulovirus_B
-----TLNTRQQK-LFDYL------TQT----K--------------------------SF-----SPVFISGSAGTGKSALLVALREH---W-L--A--E-------------------
-----D-KIVFVTAYTHLAARNI-------NG------KTCHSL-FRF---D--F------DL-N---L---L-------R---------------------------------------
----------A-------------Q---I--GV--PHYVIIDEISM--VPEKMLDG-IDSRLRQNS------------------G---------K----------------------FS-
--------------------------LPFGG-----VNVVVFGDLYQIPPVD------------------KH------------------------------------------------
------------------------RQL-----L------------PPYKS------------------DIWHS--------F-ELYELTENMR----Q-S------------E--P----
```

```
---------EFIANLNMLRVG---N-V-----------------------KC------------------------------IP--YF-NG--FV---------VDAA----------
------------------------------------------------------------------TQNI-ED---------------------S---------V-G-----------
C-------------------TSLVSTHKEANDLNLKCYT-YI----G---A--------------------------AADDDIE-----------GGEKKEE-L-V-C----------
----------V---------------------VKR------SHG-----------------------------------------R--------WTK----------------D-M--
---IVF----------------------------------------------------------N--------------E--------------------EQ-----------A-Q--
------LIF----------GESI-----RVCVGARMMITHTTD------------------------S-----FCNGDLGVVQSF
>YP_654526.1_Choristoneura_fumiferana_granulovirus
-----KLNKKQQQ-IFDIL------TEK----E------------------I------YF-----KPVFVSGSAGTGKSALLTTLREH---W-Q--G--L------------------
-----Q-KIVYVAAYTHLAARNV-------SG------KTCHSL-FGF---D--F------DL-N---L---V-------K-------------------------------------
----------T-------------Y---V--GL--PNYLIIDEISM--IPEKMLDK-IDSRLRQNS-----------------G---------N---------------------RY-
--------------------------TPFGG-----VNVIVFGDLYQLPPVT-----------------KT-----------------------------------------------
------------------------SDY-----L-----------PPYKA------------------DVWQC--------F-RLFELTENMR----Q-S-----------E--T----
---------DFINNLNLLRIG---D-N-----------------------TC------------------------------LS--YF-NN--MV---------LKTP----------
-------------------------------------------------------------------QSL-EE---------------------K---------L-L-----------
Y-------------------TSLVSTHSEANALNNQCYN-YN----K---S--------------------------N----------------------EE-F-L-C----------
----------D---------------------IST------HTT-----------------------------------------K--------WRR----------------N-M--
---LCF----------------------------------------------------------N--------------V--------------------DQ-----------E-N--
------LIF----------PQNL-----KVRKGTRVMITHTND------------------------S-----FCNGDLGIVESF
>NP_663278.1_Phthorimaea_operculella_granulovirus
-----KLNASQQY-LFDRL------ARA----Q-------------------------KF-----DPIFVSGSAGTGKSALLIALRDH---W-L--S--Q------------------
-----G-KCVSVAAYTHLAARNI-------GG------RTCHSL-FGF---D--F------DL-N---L---I-------D-------------------------------------
----------R-------------C---I--SI--PHYLILDEISM--IPEKMLDG-IDARLRTTT-----------------R---------K---------------------YD-
--------------------------QPFGG-----VNIIAFGDMYQLPPID-----------------TN-----------------------------------------------
--------------------------------E-----------PIYMS------------------DVWNT--------F-RLYELTENMR----Q-S-----------E--H----
---------EFITNLNLLRVG---D-L-----------------------NC------------------------------LP--YF-NT--LV---------MKQK----------
-------------------------------------------------------------------PKI-ED---------------------K---------L-R-----------
C-------------------TSLVSTHREADEINDQCYE-AI----A---D--------------------------K---------------------ESE-T-V-M----------
----------E---------------------STH------EMV-----------------------------------------P--------WSY----------------K-A--
---TVF----------------------------------------------------------N--------------K--------------------DQ-----------E-K--
------VVF----------KDKL-----KVCIGTRVMITHSTQ------------------------G-----FCNGDMGTIKYI
>AIU36910.1_Cydia_pomonella_granulovirus
-----KLNREQQL-MFDRV------ANA----R-------------------------RF-----EPLFVSGSAGTGKSALLVALRNH---W-R--E--R------------------
-----G-KIVYVGAYTHLASRNI-------DG------RTCHSL-FGF---D--F------DL-N---L---T-------E-------------------------------------
----------K-------------D---V--GV--PNYIILDEISM--IPDKMLDG-IDSRMRQNT-----------------R---------N---------------------PH-
--------------------------TPFGG-----VNVIVFGDLYQLPPVD-----------------KN---NY------------------------------------------
------KKR---------------EKV-----L-----------PPYEA------------------DVWTE--------F-KIYELGENMR----Q-T-----------E--Q----
---------EYIHNLNLLRLG---D-F-----------------------SC------------------------------LP--YF-NT--LV---------MDFA----------
-------------------------------------------------------------------PEI-EE---------------------K---------V-A-----------
H-------------------TSLVSTHDEANTINNECYN-FV----V---N------------------------------------------------EAE-T-T-L----------
----------K---------------------CTT------KLV-----------------------------------------P--------WSY----------------K-V--
---NVF----------------------------------------------------------N--------------A--------------------QQ-----------E-R--
------LIF----------KQEL-----QVCPGTRVMVTHTTQ------------------------H-----FCNGDTGIIEYI
>ADO85536.1_Pieris_rapae_granulovirus
-----SLNAKQQH-LFNFL------VNS----D-------------------------YF-----EPVFVSGSAGTGKSALLITLRNY---W-R--E--Q------------------
-----G-KIVFVTAFTHLAARNI-------DG------KTCHSL-FGF---D--F------DM-N---I---T-------D-------------------------------------
----------K-------------R---V--GL--PDYIIIDEISM--IPEKMLDG-IDLRMRQNS-----------------R---------N---------------------FE-
--------------------------LPFGG-----VNVVAFGDLYQLPPVN-----------------NR---------------------------------------I-------
------------------------NYT-----L-----------PPYES------------------DVWNL--------F-KLYELTENMR----H-T-----------E--P----
---------EYIKNLNLLRVG---D-L-----------------------RC------------------------------LN--YF-DS--LV---------SNRI----------
-------------------------------------------------------------------PTV-EE---------------------K---------V-A-----------
F-------------------TSLVSTHKEADNINLQCYT-YI----A---D--------------------------R---------------------EQE-I-V-H----------
----------T---------------------CET------KLL-----------------------------------------P--------WSH----------------K-I--
---TVY----------------------------------------------------------N--------------A--------------------DQ-----------E-R--
------LIF----------KPNI-----KICPNTRIMVTHTTQ------------------------H-----FCNGDMGVIEYI
>NP_891963.1_Cryptophlebia_leucotreta_granulovirus
-----TLNKEQKY-LFDKV------ADT----H-------------------------NF-----SPIFVTGSAGTGKSALLMTLRNY---W-R--N--Q------------------
-----G-KTVFVAAYTHLASRNI-------DG------KTCHSL-FGF---D--F------KL-N---L---I-------DK------------------------------------
----------K-------------N---I--GI--PDYIILDEISM--IPDKMLDG-IDSRMRQVT-----------------R---------E---------------------PQ-
--------------------------KPFGG-----VNTIVFGDLYQLPPIE-----------------DK-----------------------------------------------
--------R---------------DMT-----L-----------PPYSA------------------DIWSV--------F-KLYELKHNMR----Q-T-----------E--A----
---------EYIKNLNLMRSG---E-I-----------------------SC------------------------------LK--FF-NT--LV---------TKFS----------
-------------------------------------------------------------------IGI-ED---------------------L---------L-V-----------
H-------------------TSLVSTHREADDINMQCYI-YN----S---E--------------------------E---------------------KEE-I-I-L----------
----------K---------------------STS------SLV-----------------------------------------S--------WNF----------------Y-L--
---NVF----------------------------------------------------------N--------------T--------------------EQ-----------E-K--
------LIF----------RDSL-----KVCKGTRVMITHTTG------------------------D-----FCNGDLGIIDNI
>QOD40078.1_Matsumuraeses_phaseoli_granulovirus
-----NLNEKQQK-LFDYL------VNV----D-------------------------NF-----EPVFVTGSAGTGKSALLSALRDH---W-Q--S--Q------------------
-----N-KSVYICAYTHLAARNI-------KG------KTCHSQ-FGF---D--F------KL-N---L---M-------G-------------------------------------
----------R-------------W---A--GL--PQYLILDEVSM--IPDKMLDG-IDTRLRRSS-----------------R---------D---------------------YN-
--------------------------LPFGG-----VNVVAFGDLYQLPPVE-----------------DR-----------------------------------------------
------LKK---------------DKV-----L-----------PPFES------------------DVWNT--------F-RLYELTENMR----Q-S-----------E--V----
---------EFIKNLNLLRVG---D-N-----------------------SC------------------------------VQ--YF-DT--LV---------FRKC----------
-------------------------------------------------------------------TNI-EE---------------------K---------S-N-----------
C-------------------TSLVSTHKEADSVNVECYD-YI----S---R--------------------------N---------------------KAQ-K-T-L----------
----------T---------------------LTQ------KIV-----------------------------------------P--------FGY----------------K-M--
---IVF----------------------------------------------------------N--------------A--------------------DQ-----------E-M--
------LIF----------KKDL-----KVCVGTRVMVTHTTT------------------------H-----FCNGDTGVIERI
```

```
>YP_009182312.1_Diatraea_saccharalis_granulovirus
-----KLNARQQA-LFDAV------ANV----D-------------------------YF-----PPIFVSGSAGTGKSALLVALRNY---W-Q-RE--Q------------------
-----D-KNVCVTAFTHLAARNI-------EG------KTCHSV-FGF---D--F------KM-N---L---D-------N-------------------------------------
----------R-------------P---I--TL--PDYLIIDEISM--LPSKMLDD-IDLVLRKNS----------------K---------N----------------------YH-
--------------------------TPFGG-----VNVIVFGDLYQLQPVG-----------------AR-----------------------------------------------
--------------------------------------------PPYEA------------------DVWSV--------F-SLYELTENMR----Q-S-----------E--S----
---------EYMQNLNLIRVG---N-I-----------------------KG------------------------------LD--YF-DK--LV---------MKPY----------
-------------------------------------------------------------------PTI-KD---------------------S---------I-A-----------
Y-------------------TSLVSTHKECDNINEKCYK-FL----K---G--------------------------T---------------------KDD-E-V-M----------
----------W---------------------CKL------ESV-----------------------------------------K--------RTH---------------QH-T--
---TVF----------------------------------------------------------N--------------A--------------------GQ-----------K-T--
------IIF----------KPII-----RLFPGARVMITHTTD------------------------W-----FCNGDAGIVERI
>YP_003517787.1_Lymantria_xylina_nucleopolyhedrovirus
-----LLNAKQQY-IFDYF------TQR----D-------------------------SF-----APVFVSGSAGTGKSALLMALHEF---W-R--R--R------------------
-----N-EIVLVAAYTNLAARNV-------KG------KTCHSL-FGF---D--F------NL-N---A---K-------C-------------------------------------
----------T-------------P---L--PV-KPRCVIIDEISM--IPAKMLDG-IDRKLQQTT----------------G---------E----------------------HD-
--------------------------KPFGG-----VNVIVFGDLYQLPPVN-----------------KT-----------------------------------------------
------------------------SDA-----K-----------PVYAA------------------DAWNA--------F-RLYELTENMR----Q-S-----------E--S----
---------VFIDNLNLLRVG---D-F-----------------------KC------------------------------LK--YF-NS--LK---------LKTP----------
-------------------------------------------------------------------PRI-ED---------------------Q---------L-K-----------
S-------------------TSLVSTHKEADAINRQCYE-AV----S---A--------------------------NA--------------------QSR-V-V-V----------
----------S---------------------VTE------NAV-----------------------------------------R--------REHM-------------ERD-A--
---QIF----------------------------------------------------------N--------------S--------------------EQ-----------E-K--
------LIF----------KPQL-----TLCAGARVMITHTTA------------------------E-----FCNGDLGTVESV
>QWO71653.1_Orgyia_pseudotsugata_nuclopolyhedrovirus
-----KLNIKQQL-LFDFL------TQA----T-------------------------EF-----RPLFVSGCAGTGKSALLRALRNF---W-T--R--Q------------------
-----N-ETVYVAAYTNLAARNV-------DG------KTCHSL-FGF---D--F------KL-N---V---K-------RP---F--------------------------------
---------------------------SL--KV--PHCLILDEISM--IPGQMLDK-IDEILKRAC----------------K---------N----------------------DE-
--------------------------KPFGG-----VNLVVFGDLYQLPPVD-----------------KN-----------------------------------------------
------------------------DTM-----K-----------PVYEA------------------KVWPQ--------F-TLYELTENMR----Q-S-----------E--A----
---------LFIDNLNMLRTG---D-A-----------------------KC------------------------------VE--YF-NT--LT---------LKTP----------
-------------------------------------------------------------------PTV-EN---------------------Q---------L-N-----------
N-------------------TCLVSTHNESNSINVNCYN-AI----T---I--------------------------D---------------------QVE-T-V-V----------
----------R---------------------LNK------RVL-----------------------------------------N--------RKSV-------------KRD-T--
---QIF----------------------------------------------------------N--------------V--------------------EQ-----------E-N--
------MIF----------KSNL-----KLCPGTRIMVTHTTN------------------------N-----FCNGDFGIVESV
>YP_001651017.1_Orgyia_leucostigma_nucleopolyhedrovirus
-----QLNAKQQS-IFNYL------TEK----D-------------------------TF-----EPIFVSGCAGTGKSALLKALRKF---W-F--K--E------------------
-----K-KTVVVAAYTNLAARNV-------QG------KTCHSA-FGF---D--F------KL-N---I---R-------R-------------------------------------
----------I-------------P---L--SS-KPDYVIIDEISM--IPAQMLDK-IDTKLKYSS----------------G---------A----------------------TS-
--------------------------EPFGG-----VGVVVFGDLYQLPPVD-----------------KN-----------------------------------------------
------------------------VTT-----K-----------PVYEA------------------NVWPS--------F-KLFELTENMR----Q-S-----------E--A----
---------LFIDNLNSLRTG---N-T-----------------------SC------------------------------VD--FF-ST--LT---------LKQP----------
-------------------------------------------------------------------PTV-EN---------------------Q---------L-N-----------
S-------------------TCLVSTHNEANIINANCYE-SI----A---A--------------------------D---------------------QLE-I-V-I----------
----------Q---------------------LKE------RLV-----------------------------------------S--------RKEV-------------RGD-T--
---QIY----------------------------------------------------------N--------------V--------------------EQ-----------E-N--
------LIF----------KRDL-----KLCPKTRIMITHTTK------------------------N-----FCNGDFCVVEKV
>YP_006908627.1_Epinotia_aporema_granulovirus
-----TLNQSQQK-LFDYV------VSR----Q-------------------------EF-----EPIFVSGSAGTGKSALLLALQKR---W-E--D--D------------------
-----K-KIVMTVAYTHMAARNV-------NG------TTCHSA-FGF---D--F------NL-N---L---K-------SY------------------------------------
----------------------------ICNPV--PNYLIIDEISM--IPDKMLNG-IDEKLRYNT----------------G---------V-----------------------D-
--------------------------KPFGG-----VNVIVFGDLYQLPPIN-----------------DE---K--K----------------------------------------
------------------------NNF-----K-----------PPFYS------------------RVWNS--------L-SLYELTENMR----Q-T-----------E--A----
---------EFIANLNMLRVG---D-I-----------------------RC------------------------------KK--FF-DK--LV---------TKP-----------
-------------------------------------------------------------------PLI-SE---------------------S---------V-K-----------
C-------------------TTLVPLNYKADIVNINCYK-YI----R---G-----------------------LNKK---------------------AEE-Y-T-V----------
----------K---------------------IEQ------HIL-----------------------------------------R--------KTF--------------ENS-R--
---ILF----------------------------------------------------------T--------------K--------------------KQ-----------E-E--
------MIF----------QPGM-----KFCVGTRIMATQNIN------------------------G-----FCNGDVGIVTEV
>YP_009513161.1_Agrotis_segetum_granulovirus
-----KLNEEQQR-IYDYV------TRV----K-------------------------KF-----APIFVSGSAGTGKSALLIAIRDW---C-R--A--E------------------
-----K-KVVWIVSYTNLAARNI-------EG------KTIHSM-FKF---D--F------NL-N---I-----------SN------------------------------------
----------Y----------------RI--NA--PEFLIIDEISM--VPAKMLNG-IDAQLKRST----------------G---------E-----------------------D-
--------------------------AAFGG-----VNTIVFGDLYQLPPVE-----------------NR-FRQ-------------------------------------------
------------------------NFT-----L-----------PPYHS------------------HAWSD--------F-RLFNLTINMR----Q-S-----------E--E----
---------LFIKALNLLRKG---D-A-----------------------SC------------------------------QD--FF-NS--KV---------IDQE----------
-------------------------------------------------------------------PCL-EE---------------------K---------I-N-----------
C-------------------TSLVSTHLEANHLNNICYE-YV----K---S--------------------------KSK------------------DKKE-Y-Q-V----------
----------K---------------------LIK------TLE-----------------------------------------K--------RHT----------------T-S--
---MPY----------------------------------------------------------N--------------K--------------------SQ-----------E-E--
------MIF----------KDNI-----KYCVGTRVMITLNVR----------DFV-GE--N-----S-----FCNGDIGTIVQV
>YP_001257069.1_Spodoptera_litura_granulovirus
-----TLNKQQQQ-LFDYV------TLT----K-------------------------EF-----GPIFVSGSAGTGKSALLRALQSH---W-K------------------------
-----N-KTIWVTTYTNLAARNV-------NG------TTLHKQ-FKF---N--F------KG-E---M---N-------TN------------------------------------
----------A----------------CV--GV--PNYFIIDEISM--VSSKMLQQ-IHECLQNNT----------------Q---------V-----------------------D-
--------------------------LPFGG-----VNTIVFGDLYQLPPIS-----------------TA---K-------------------------------------------
------------------------DKS-----L-----------PPYHA------------------DVWKE--------F-KLFELTENMR----Q-N-----------E--K----
```

```
---------DFIDALNMLRIG---D-S-----------------------RC------------------------------QK--FF-DD--KV---------LQKS----------
-------------------------------------------------------------------PSV-EE---------------------K---------L-N-----------
T-------------------TSLVSTHNEANAINEQCYK-RI----C---I--------------------------D---------------------KEE-H-T-V----------
----------E---------------------LSV------EKT-----------------------------------------N--------RGR----------------D-M--
---IVF----------------------------------------------------------N--------------Q--------------------NQ-----------I-D--
------MIF----------KDKM-----KYCVGTRIMVTHNVA------------------G-----V-----FCNGDVGEIVGI
>AXS01146.1_Spodoptera_frugiperda_granulovirus
-----TLNEEQQK-LFDYV------TGL----E-------------------------EF-----APIFVSGSAGTGKSALLKALKKF---W-T--E--Q------------------
-----D-KLVWVVSYTNLAARNV-------EG------STIHKQ-FGF---D--F------QC-Q---L---R-------NN------------------------------------
------D---R----------------NL--GA--PNYLILDELSM--VPAMMLDG-ISERLKQST----------------R---------L-----------------------D-
--------------------------MPFGG-----VNTIMFGDLYQLPPIS-----------------NA---H-------------------------------------------
------------------------SVQ-----L-----------PPYHA------------------KVWPS--------L-RLYELTTNMR----Q-S-----------E--S----
---------DFIEALNLLRVG---N-N-----------------------AC------------------------------LN--FF-DK--QV---------VTSP----------
-------------------------------------------------------------------ISI-ED---------------------Q---------V-K-----------
C-------------------TSLVATHREADLINAKCYA-HV----K---K--------------------------T---------------------QGD-A-V-P----------
----------E-----------------YLLQLNY------KPE-----------------------------------------S--------RNL----------------H-Q--
---VVY----------------------------------------------------------A--------------S--------------------SQ-----------E-G--
------LVF----------KDGL-----KYCVGTRVMITHNLK---------------G--V-----G-----FCNGDIGTVMSV
>YP_009249960.1_Mocis_latipes_granulovirus
-----LLNEQQQC-IFDYV------TNR----V-------------------------SF-----SPIFVSGSAGTGKSALLKALRHY---F-V--T--N------------------
-----Q-KIVWVVSYTNLAARNV-------DG------LTIHKQ-FGF---D--L------KC-N---L---N-------RY------------------------------------
------N---K----------------NA--GA--PNYLIIDEVSM--VPAKMLDN-IDIFLKNNT----------------K---------I-----------------------D-
--------------------------LPFGG-----VNTIIFGDLYQLPPIT-----------------DQ---Q-------------------------------------------
------------------------CSQ-----L-----------PPYRA------------------NIWKS--------L-QLYHLTINMR----Q-S-----------E--X----
---------DFIDALNMLRKG---D-K-----------------------RC------------------------------LE--FF-DQ--KV---------TDHE----------
-------------------------------------------------------------------ITI-ED---------------------Q---------S-Q-----------
C-------------------TSLVPTHREADYINSKCYA-YI----K---T--------------------------L---------------------SED-T-L-L----------
----------E-----------------YLLKINS------RHE-----------------------------------------S--------RQL---------------HNGS--
---IVY----------------------------------------------------------G--------------A--------------------KQ-----------E-W--
------SIF----------RDGL-----KYCVGTRVMITHNLK---------------G--L-----S-----FCNGDIGTVVDI
>NP_059294.1_Xestia_cnigrum_granulovirus
-----KLNEQQQQ-IFDYV------TQR----D-------------------------SF-----EPIFVSGSAGTGKSALLKSLRTH---W-I--D--R------------------
-----K-KVVWVVSFTNLAARNI-------DG------QTIHKQ-FGF---D--F------KC-N---L---N-------AN------------------------------------
------N---K----------------NV--GT--PNYFILDEVSM--VPAKMLQN-IHTYFQQNT----------------R---------M-----------------------D-
--------------------------LPFGG-----VNTIIFGDLYQLPPIS-----------------NQ---Q-------------------------------------------
------------------------CYQ-----L-----------PPYCA------------------DIWKS--------L-RLYHLTINMR----Q-S-----------E--S----
---------DFIDALNLLRVG---D-K-----------------------KC------------------------------LE--FF-NQ--KV---------MNHS----------
-------------------------------------------------------------------ITV-QD---------------------Q---------F-E-----------
C-------------------TSLVPTHREADYINSKCYA-HI----K---S--------------------------I---------------------SEE-P-V-V----------
----------E-----------------YILQLSV------RRE-----------------------------------------S--------RRL---------------HS-M--
---MVY----------------------------------------------------------A--------------S--------------------GQ-----------E-E--
------LIF----------RDKL-----KYCVGTRVMITHNLK---------------G--L-----A-----FCNGDIGTVIAI
>QNH90674.1_Mamestra_configurata_nucleopolyhedrovirus_B
-----ILNEQQQK-LFDYV------VNR----D-------------------------QF-----EPIFVSGSAGTGKSALLKTLKAY---W-Q--D--L------------------
-----G-KHVWVVSYTHLAARNV-------DG------QTIHRQ-FGF---D--L------KG-N---L---R-------DS------------------------------------
------------------------ASS-FQRTV--PDYLIVDEISM--VSAKMLEG-MNIRLQRMT----------------D---------E-----------------------I-
--------------------------VPFGG-----VNTLIFGDLYQLPPIS-----------------NK---RYGK----------------------------------------
------------------------DDT-----L-----------PPFKA------------------PVWSS--------L-RLYELTINMR----Q-S-----------E--T----
---------DFIEALNMLRVG---N-V-----------------------QC------------------------------LN--FF-NQ--KA---------LEQT----------
-------------------------------------------------------------------PSM-DV---------------------Q---------M-S-----------
C-------------------TSLVSTHAEANAINARCYK-HL----Q---N--------------------------T---------------------NKN-E-Q-F----------
----------E-------------------LQITQ------KNK-----------------------------------------S--------RKL----------------F-S--
---MVY----------------------------------------------------------N--------------K--------------------DQ-----------E-Q--
------LIF----------KDKM-----MYCVGTRVMVTFNLK---------------N--S-----P-----FCNGHIGVIVSI
>QKV50030.1_Plutella_xylostella_granulovirus
-----TLNEQQQK-IYNYL------TSV----D-------------------------CF-----EPIFVSGSAGTGKSALLVTLTKA---W-T--M--K------------------
-----N-MRVDVGTYTNLAARNV-------NG------KTLHKL-FGF---D--L------KM-E---L---R-------SN------------------------------------
----------F-------------C---F--NA--PDYLIIDEISM--VPDKMLAG-IDERLQQAG----------------L---------N-----------------------G-
--------------------------IPFGG-----VNVVVFGDLFQLPPIS-----------------ND---K-------------------------------------------
------------------------DAA-----K-----------PPYYA------------------SVWSS--------F-KLYELTINMR----Q-S-----------E--Q----
---------EFIDALNKLRVG---D-L-----------------------TC------------------------------QK--FF-NK--QV---------LKKP----------
-------------------------------------------------------------------PSI-AE---------------------K---------L-Q-----------
C-------------------TSLVSTHKEADFNNNLCYN-HI----K---K--------------------------N---------------------KDE-K-T-I----------
----------E---------------------LKE------QYA-----------------------------------------Q--------RFK----------------H-D--
---IVY----------------------------------------------------------N--------------A--------------------NQ-----------E-K--
------IIF----------KDGM-----KYCVGTRVMITQTVP---------------T--T-----T-----LCNGDIGEIVSI
>YP_009186748.1_Sucra_jujuba_nucleopolyhedrovirus
-----ILNEEQLK-FIKML------KIR----R-ANG---------------------QC-----DPIFVSGNAGTGKTFLLKYLFQE---M-Y-LK--E------------------
-----Q-IKVKKIAFTALAARNI-------DG------TTMHKL-FRF---S--F------TG-E---F---K-------SN------------------------------------
----------N-------------INKDL--YH--IEMLIIDEISM--IHATYLDK-MDEILRLTK----------------H---------Q----------------------PD-
--------------------------LPFGG-----VQVVAFGDLYQLPPVV-----------------ES---C--N-Q------------------------------F-------
------KDQ---------------KID-----E-----------RCYFA------------------GVWKH--------F-ILFTLTETMR----Q-N-----------E--L----
---------DFITALNQLRIG---D-E-----------------------RG------------------------------IG--YF-NR--LR---------ATQT----------
---------------------------------------------------------------------Q-LN---------------------P---------M-Q-----------
S-------------------TTLVTTVAGACKINDINNK-KV----C---E--------------------------QS--------------------NVV-Y-D-I----------
----------E---------------------SRS------KIR-----------------------------------------T--------AKN----------------S-E--
---LLY--------------------------------------------------------VHS-----------------------------------AD-----------S-L--
------GVI----------PEKI-----TLAIGSRILVTSNCV---------------N--S-----H-----CINGDIGVIVDF
```

```
>YP_010086327.1_Hyposidra_talaca_NPV
-----PLNEDQQK-FMCIF------EDM----L-TRG---------------------DT-----TPIFVSGNAGTGKTFLLKHLFKE---L-T-IN--R------------------
-----K-FYVEKIAFSALAARNI-------DG------STMHKL-FRF---T--F------TG-E---Y---DM------NR---I--------------------------------
------------------------DKYAL--RC--LRVLIIDEISM--IHATYLDN-IDAILRLVH----------------E---------K----------------------PD-
--------------------------VPFGG-----VYVVAFGDLYQLPPVI-----------------ES---R--N-Q------------------------------F-------
------KNQ---------------KAN-----E-----------KCHGA------------------DVWKE--------F-QLFILTQMMR----Q-N-----------E--P----
---------EFIEALNQLRVG---N-L-----------------------RG------------------------------IA--FF-NR--LR---------AVQKP---------
-----------------------------------------------------------------------FD---------------------P---------M-E-----------
A-------------------TTLTSTIADAERINDTNNK-KI----L---A--------------------------DA--------------------VTS-Y-K-I----------
----------T---------------------CES------KKR-------------------------------------------------LIKS--------------EER-K--
---YLY----------------------------------------------------------P-----------------------------------AN-----------N-T--
------SMI----------PDNL-----TLAQGSRIMVIANCK---------------E--S-----K-----CINGDLGVVEEC
>YP_009133285.1_Lambdina_fiscellaria_nucleopolyhedrovirus
-----SLNEDQQQ-FLNSY------LAI----I-DAG---------------------DV-----QPVFVSGNAGTGKTFLLKRLYEE---L-S-ER--R------------------
-----D-MNTEKIAMSAIAARNI-------DG------ITLHRL-FMF---G--F------NG-E---Y---N-------TR------------------------------------
----------Y-------------TSKIV--RQ--MEALIVDEVSM--INAMYLDN-VDKILKTVK----------------C---------Q----------------------PN-
--------------------------IPFGG-----VHVIVFGDLYQLPPVM-----------------ND---N--K----------------------------------------
------------------------LGN-----Q-----------QCFLA------------------NVWKH--------F-KLFTLNKMMR----Q-N-----------E--K----
---------DFIEALNCLRVG---D-D-----------------------NG------------------------------IA--FF-NR--LR---------VTQH----------
---------------------------------------------------------------------Q-FD---------------------P---------M-E-----------
A-------------------STLVSTNKAAATLNDRNNV-KI----L---------------------------L--AT--------------------DKK-H-T-I----------
----------E---------------------STT------K---------------------NG--------------------Y--------IKD----------------T-R--
---YLY----------------------------------------------------------S--------------A--------------------DN-----------I-Y--
------QII----------PKSI-----TLGVGSRIIVTHNCN---------------K--S-----S-----CINGDLGVVEDF
>YP_009049868.1_Peridroma_alphabaculovirus
-----ALNDDQSA-FMRIC------DAT----L-DQR---------------------QQ-----LIAFVTGNAGTGKTFLLKHLNTH---L-A--D--R------------------
-----N-LLVERIAFSALAAQNI-------NG------KTMHKL-FKF---N--L------RG-H---Y---KL------TD------------------------------------
----------F-------------LIQDL--MH--IDVLIIDEVSM--IHGSYLDK-IDEILRVVM----------------G---------R-----------------------N-
--------------------------VPFGG-----VHVIAFGDLYQLPPVV-----------------DR---------------------------------E-W---V-------
------NPT---------------ASS-----E-----------KCYSA------------------AVWKE--------F-RLYTLRQMMR----Q-S-----------E--P----
---------DFIRALNQLRVG---D-E-----------------------KG------------------------------IQ--YF-NE--LR---------DRQK----------
---------------------------------------------------------------------A-ID---------------------S---------M-E-----------
A-------------------TTLVSTVAAAHAINTKNNK-TL----L---E--------------------------NA--------------------DET-H-E-L----------
----------V---------------------STS------KVM-----------------------------------------P--------AVD----------------P-D--
---FLY----------------------------------------------------------P--------------H--------------------KN-----------M-H--
------QVV----------PDKL-----TLCVGSRIIVTVNCK---------------D--S-----E-----CVNGDLGVVEKF
>YP_008004327.1_Choristoneura_biennis_entomopoxvirus
-----DCNIEQKN-FIDYLDKNIINDNI----T-------------------------NL-----YPIFITGSAGSGKSYLLRCIIDK---F-K--D--Y------------------
-----N-INPDIAAFTAIVSKSI-------GG------RTIHSL-FKF---D--F------FG-K---C---L-------K-------------------------------------
----------P-------------NVSLL--KN--MKVLIIDEISM--VSAKYLDS-INDMLMKYK----------------K---------N----------------------TN-
--------------------------V-FGG-----VFVIVFGDLYQLEPIS-----------------ND-----------------------------------------------
--------E---------------NDE-----L-----------PVYKS------------------IVWQN--------F-LKYQLYENMR----Q-N-----------E--K----
---------EFINALNMIRIG---K-L-----------------------DS------------------------------LD--YF-NN--IY---------KKSKS---------
------------------------------------------------------------------NNLE-EK---------------------E---------I-N-----------
S-------------------TTIVSTNDEAYIINTRIFD-KI----K---Q--------------------------N---------------------NSE-I-Y-Y----------
----------LN-----------------NANYKT------KYI-----------------------------------------N--------YDP----------------D-VY-
--DYSY----------------------------------------------------------D--------------K--------------------NN-----------I-N--
------KIF----------P-NI-----YICKGTKIMITANCV---------------E--N-----S-----CKNSDMGYIDNI
>YP_803305.1_Trichoplusia_ni_ascovirus_2c
-----MMTPCQLR-AYNIL------IEN----M-NKNP-----------LDP------TR-----LPIFISGGGGTGKSYVLKKFKDY---V-V--N--V------------------
-----N-KKIAVVATQAIAATLI-------DG------KTIHSV-FNI---R--G------GN-A---Q---T-------PD------------------------------------
------Q---R----------C-----TL--TSFPYDVLIIDEISM--LNGELLDL-IENTLVTVK----------------R---------S-----------------------A-
--------------------------MPFGG-----VYVCVLGDLLQLPPVN-----------------KS-----------------------------------------------
--------------------------------------------PVYKA------------------NCWKW--------F-RLISLVTNVR----H-K----------GD--D----
---------EFSNIMARVRIG---D-R-S------A----------------------------------------------ID--IL-NE--KC---------LKTI----------
-------------------------------------------------------------------PEI-EHL------------------RFE---------E-G-----------
I-------------------ITIVATNRQVQKINNAATK-KF----S---D--------------------------N-----------------------G-T-L-I----------
----------KTIH------------------SKD------ESTFMR--SDTYT---------IG--------------------N--------------------------------
---IMY----------------------------------------------------------N--------------H--------------------DDI------------D--
------RIV----------PKSI-----DIFIGAIVMITANDI----------N----GN-G-----R-----WCNGDTCKIVNI
>AYD68236.1_Heliothis_virescens_ascovirus_3h
-----IWTDSQKS-AYDGI------ITT----F-KRNV-----------VDV------QR-----CPIFVTGRGGTGKSFLLHRLREY---F-E--N--H------------------
-----G-VRVVVTATQAVAAQLV-------SG------KTLHST-FKI---R--R------IR-S---V---D-------SA--AF--------------V-----CDI---------
---------------------------DI--FP--YDVLIIDEVSM--LSDTLLDT-IEQKLTTIR----------------D---------C-----------------------R-
--------------------------APFGG-----VFVVGFGDLLQLSPVQ-----------------DR-----------------------------------------------
--------------------------------------------EVYMA------------------KSWKY--------F-KLVALTTSVR----HGN-----------D--K----
---------KYDNLMSRLRLG---D-K-T--------------------------V--------------------------VN--AI-NE--YC---------VKTS----------
-------------------------------------------------------------------CEI-DKD------------------LLE---------N-N-----------
T-------------------TVVVAKNIFAERNNLSIAK-RL----V---R--------------------------DN--------------------DLS-N-S-Y----------
-----RTL--KRHESTVDC-------------AND-----------------------------S--------------------D-----------------------------
KDY-----Y----------------------------------------------------------T-----------------------------------LQ-----------
EVE--------RIV----------PKEL-----SVFPGATLMFTANGL----------S----G--G-----P-----WCNGDICKVVSL
>P_oligandrum
-----SFTSEQQL-FIDLA------VV-------------------------------TR-----RNLLLVAPAGYGKSFVIKEIVER---F-R--H--ELTRID-------------
----EQ-PVYALCASTGKAASLI-------GG------RTLHSY-LGI---G--L------AQ-G---T---P-------DE---W--------------V-----MCL--R--V---
-N-SSMR---P-------------KLEAL--KA--VQVILVDEVSM--VSAEFLDK-ISTYLQLLR----------------H---------N-----------------------Y-
--------------------------RPFGG-----VQMILIGDLCQLPPVK-----------------ES-----------------------------------------------
--------------------------------------------FIFRS------------------KEYKR----GY--F-HPFQFTRCFR----Q-N-----------N--R----
```

```
---------EFVALLNEVRFG---D-C-A------D--------------RE------------------------------FA--TL-QQ--RT---------SIDP----------
--------------------------------------------------------------------QY-SN--------------------G---------L-T-----------
P------------------MRIVSTNEEVDKINADEME-AL----L---R-------------------------K----------------------T-G-V-----------
-------------------------------AKE------RYT-----------------CYLS-------------------NP------------------------KHA-E--
-----Y-------------------------------------------------------A-K-------------K--------------------CR-----------E-D--
------ARI----------PEFV-----DVAVGCQLVVTHNLT-----------------D-----K-----IVNGTQGRVIAT
>YP_008437119.1_Pandoravirus_salinus
-----RWSPAQAH-AARLA------E--------------------------------AG-----HNLLLSGSGGAGKSFLLRYMIAA---K-R--A--Q-----------------
-----G-KTVQVTGSTGMAAVNV-------GG------TTLHRV-LGC---G--L------GA-E---P---L-------PA---L-------------Q-----AAL--G--T---
-R----P---K-------------VVARW--RA--MDVLVVDEISM--VDAEFFHK-CDQLARWMR----------------G---------R----------------------MD-
-------------------------RAFGG-----IQVILVGDFAQLPAIV----------------DR---T--P-APG-------------------------------------
---------------------GAERP-----Q-----------FCFEL-----------------PLWTDRA--LA--L-QVVDLRTVFR----Q-R-----------D--D----
---------TLVGALNRMRFA---E-Q-T------P---------------ED------------------------------EA--LF-AA--RV---------GAVL-----A----
----------------------------------------------------------------------T-DD---------------------G---------V-E-----------
P-------------------TRLCPLVAQVAAINAERLA-GL----T---G-------------------------E----------------------S-D-T-F----------
----------A--------------------SKC------PWR-----LD-----------DG-V-----------------K--------MTPKIEAT-----------LN-A--
-----H-------------------------------------------------------R-A---------------A----------------------LE-----------K-N--
------APA----------APYV-----NLKVGAQVVLLANLD----------V----E--H-----G-----LVNGARGVVRRF
>BBI30459.1_Acanthamoeba_castellanii_medusavirus
-----TLNEDQAA-FIRWL------EDL-----------------------KN------SR-----KSAFVTGPAGTGKSALIEEATAV---L-E--R--Q-----------------
-----E-RNFVRTASTGAAAFNI-------GG------TTTHSA-FSV---G--A------CL-L---D---I-------EK---L---------------C-----KRM-DK--M---
-P----S---E-------------FTERW--LE--MDDVIIDELPM--ISADAFEK-IIAVAKHLR----------------P----------N----------------------R-
-------------------------PP--------LRFLFFGDFFQLPPVL----------------ER---G--E-RERR--------------------------------------
------EAK---------------NLP-----I-----------YCFQT-----------------DAWRR----LR--P-QVFYLSKIER----Q-E-----------D--R----
---------AFAEILSRVRTG---D-K-T------P---------------AD------------------------------DA--FF-AN--RV---------RATH-----
AARASA------------------------------------------------------------LPKGAPAPQL-NQITRD---------------PEYE---------R-D------
-----L-------------------TRIRTNNPDVDTINLAAFD-VT----R---N-------------------------P----------------------SGTSTPY-----
--G-F----EQTF--K-----------------VY---SVPK-----SVPKSRKLQDFA----------EA--------------------SPTQ-------------------
ISKHK-------F------------------------------------------------------ELEAMQKL-----------------------------------LI---------
--S-Q--------CTA----------DRPL-----RLRVGVEVVIVCNID----------T----S--S-----G-----LVNGARGTVVGF
>OSX74557.1_Porphyra_umbilicalis
-----AASQPAAS-IVSQL------S--------------------------------RS-----RSVFLTGAAGTGKTTLLKEVVPQ---L-R--V--L-----------------
-----D-RSMGVCATTGMAASLI-------GG------VTLHSW-AGL---R--P------AT-T---A---ALVGGTSASE---L--------------------VSL-----F---
-P----P---R-------------ARERL--SS--ARFLVLDEISM--LNAALLDG-IDRVCRLLR----------------R---------Q----------------------PN-
-------------------------TPLGG-----LVVLFCGDFVQLPPVS----------------GQ---G--M-------------------------------------------
-----------------------FSG-----T-----------YAFRA-----------------AVWPA----LF--ADQGVLLRVNFR----QGA-----------D--S----
---------RFLGLLHRMRRA---E-L-S------Q---------------DD------------------------------VQ--ML-NS--RV---------GRST----------
------------------------------------------------------------------------------------------------P---------A-D-----------
V-------------------VTLFSKNEQAHEHNAERLD-QL----K---T-------------------------P----------------------A-V-E-Y----------
----------Y--------------------AVD------DYKQLD--KE-----------QG-----------------------------------------------------
---VAL--------------------------------------------------------------------------------------------LA-----------A-V--
------TAA----------QLVV-----TLRVGAVVVLLSNQY----------F----HV-H-----Q-----LCAGSRGVVVGF
>OSX70336.1_Porphyra_umbilicalis
-----TIVGAERK-VCRLL------T--------------------------------GN-----ECVFLTGPPGCGKTHLVNDVVKT---L-R--A--V-----------------
-----G-LSVSVCGSSGVAAALV-------GG------TTVHAW-AGF------V-----NGD-A---D---V-------AT--PL--------------------ETVLTK-VI---
-P----P---A-------------AKYRM--RS--AMALVIDEVGT--LSAALITR-LDVVLRDVW----------------R---------C----------------------A-
-------------------------LPFGG-----LVVLFSGDFLQLAPPI----------------GN----------------------------------------------
-----------------------------------------FAFLS-----------------GAWRE----AF-DN-RAIVLDTHWR----HIN-----------D--R----
---------QLLDVLLRMRVG---L-H-T------T---------------ED------------------------------IQ--LL-AT--RR---------SAKP----------
------------------------------------------------------------------------------------------------P---------P-N-----------
A-------------------IWLFCHTIPAKDKNEEELR-QL----P---G-------------------------P----------------------N-V-T-Y----------
----------H--------------------AQD------KVKV----------------------------------------------------------------------E--
-----YL-----------------------------------------------------TLDS---------------ARTL-------------------LD-----------E-G--
------LKF----------VRVL-----KLRVGAVVLVPSNCL----------A----G--D-----G-----VPAGSRGWCFVF
>AYV86632.1_Sylvanvirus_sp.
-----HLSQSQRI-AAYLI------IH-------------------------------GR-----HNMLLTGGAGVGKSLLADFTSKA---F-K--K--M-----------------
-----N-VLVRFTSTTGNAALQLP------NG------TTFHSW-LGL---G--L------AK-E---N---P-------RT---L---------------A-----VSA--A--N---
-K----K---H-------------VKETM--FN--TECLWIDEVSM--FPARLLEL-LNLLGQQVR----------------K---------N----------------------N-
-------------------------RPYGG-----IQVILSGDFAQSMPIP----------------DK---L--N-QFQRRSDSM-----------------EEAAYI-------
------EKL---------------EKL-----E-----------FCFQH-----------------PEWDM----LV--D-YTIYLQEVFR----Q-T-----------D--R----
---------AFVEMLNRIRLG---R-H-T------V---------------ED------------------------------TK--KL-NSINKF--------TTPQQEE--
NEDQLMASGSSKT---------------------TS----------------TSTSSS------------SSISSF-DDPYNQ---------------FIPVD--VEHRFS-D-----
------Y-------------------VHMYAINDEVDRKNEKEMA-KL----A---H---------------------------P----------------------S---T-Y----
-----------KQFKFK--------------------TQT---------------------------FKS--------------------TP------------------------
FN-K-----IAF--------------------------------------------------DSWLKD------------------------------------QR----------
-D-K--------SLV----------KPQI-----ELTLGARVMLMVNQS----------I----Q--D-----G-----LVNGSVGIIIDF
>CAE7237458.1_Symbiodinium_microadriatiucum
-----PETKHQRH-AMEHI------IQE----V-LSRP------------NTK-----DGSNPERLHMLLHGPGGCGKSVVIRAAAHM---L-R--Q--G-----------------
-----G-VGVVIAAPTGVAAWNI-------NG------VTLHAC-CLL---P--V------VN-K---SYGKP-------GD---LP---------------PPSGPLL---------
------A---T-------------LPSMW--RL--VSALFVDEMSF--ISSFMLER-LDQHLRLAR----------------D---------T----------------------PN-
-------------------------LPFGG-----VHIVFAGDLYQLPPPG----------------GH----------------------------------------------
-----------------------------------------PLFKS-----------------QLW------LL--F-RLCELRGNQR----AAK-----------D--P----
---------EWAALLARVRVG---K-C-T------E---------------KD------------------------------IK--EL-RD--MV---------VKPS-----SS---
----------------------------------------------------------------K-QP--------------------A---------P-K-----------
A-------------------VHLYATRRAVAESNRTYFE-EHVSRTN---A-------------------------D--------------------------I-Y----------
----------E--------------------SPALD-----VNVR------------------TG--------------------A------------------------PLSPE--
-----V-------------------------------------------------------VWP--------------D--------------------PE-----------N-T--
------GGL----------EALV-----RVAVGVRVMLRHNID----------V----Q--D-----G-----LVNGACGFVEQV
```

```
>Perkinsela_sp
-----APSADQEK-ALKLA------L--------------------------------ER-----KCLYLGGPAGTGKTHVLHRIYRH---L-T--S--K------------------
-----R-QVVLVTASTGIAAQSM-------NG------RTFQHF-FGI---R----------------------------GD---C--------------------------------
------------------------QVKLI--ED--VDCILLDEVSM--MFPTILEN-FDATARLMR----------------G---------S-----------------------A-
--------------------------EPFGG-----IRMILCGDFLQLPPVA-----------------KD---K--A----------------------------------------
-------------------------------AQ-----------VIFEH------------------PLFRD-----N--F-YLTALHVVHRV--YA-------------E-YS----
---------SFREGLAKLRYG---Q-L-T------S--------------EM------------------------------YS--LI-RS--RA---------GTPD----------
-----------------------------------------------------------------------DEL------------------------------S-D-----------
A-------------------TYLFLTRHQAAHRNLLELE-KI----P---G--------------------------E-----------------------S-L-A-FPRVLTEP---
----------KL--------------------SST---WTSSVAFTC--SEKIPP----ALRSTG-----------------NLRSALNYV--LSGKLFARGAPSDALVLEVLS-
ETADVATYFTHANTFWF-----------RRFAADAEDKAATLRRNLIRAIEYIGGDIVHGQGDAILRK-------------VSPA------------------ID-----------
RCTAR------DRV----------GETI-----HLRVGARVMLTRNLT------------------P-----A-----LVNGSIGVVEDF
>QFR57062.1_Klebsiella_phage_AmPh_EK29
-----MLNKGQKK-AFDYI------ISR----I-KAG---------------------KG-----NHITLNGPAGTGKTTMTKFIVDY---L-I--S--Q------------------
-----GVSGVVLAAPTHAAKKVLS----KLSG---IEARTIHSL-LKI---N--P------TT-Y---E---D-------SV---T----------------------------F---
-E----Q---K-------------GDVDV--SE--LRVVVCDEASM--YDRKLFQI-LMATIP-------------------------------------------------------
-----------------------------RY-----CLVIAIGDKAQIRPVE-----------------PG---S--T-VPA----------------------------L-------
--------------------------S-----------------PFFSH------------------K---D--------F-DQLELDEVMR----S-------------N--A----
---------PIIKVATDIRNG---K--------WIYDHQR--DDHGVHGFTS---------------------------TTALKD-FM-MK--YF-----------------------
----------------------------------------------------------------EIVKDP-EDM------------------------------F-E-----------
N-------------------KMFAFTNKSVDKLNSIIRR-RI-----------------------------------------------------------------L----------
-------------------------------------------------------------------------------------------------------------------E--
-----------------------------------------------------------------------------------------------------------------T-E--
-----D----------------------AFITGEVIVMQEPLIKELEFEGKRFN----D--L-----K-----FNNGQYVRIVSA
>YP_010093920.1_Enterobacter_phage_myPSH1140
------LNEDQKD-TFNRV------VER----I-KAG---------------------RG-----GHITINGPAGTGKTTMTKFIINY---L-I--S--T------------------
-----GVSGVMLAAPTHGAKRVLS----KLAG---VAANTIHSI-LKI---N--P------TT-Y---E---E-------NM---L----------------------------F---
-E----Q---K-------------EVPDM--AK--CRVLICDEASM--YDRKLFQI-IMATIP-------------------------------------------------------
-----------------------------SW-----CLIIAIGDKSQIRPVE-----------------PG---S--T-VPA----------------------------L-------
--------------------------S-----------------PFFTH------------------K---D--------F-EQLYLTEVMR----S-------------N--A----
---------PIIKVATDIRNG---E--------WIYEHLV--DGEGVHGFTS---------------------------QTALRD-FM-MT--YF-----------------------
----------------------------------------------------------------ENVKTM-EDM------------------------------F-E-----------
N-------------------RMLAFTNKSVDKLNSIIRR-RI-----------------------------------------------------------------F----------
-------------------------------------------------------------------------------------------------------------------Q--
-----------------------------------------------------------------------------------------------------------------T-E--
-----E----------------------PFIVGEVVVMQEPLIKELEYDGKKFS----E--V-----I-----FNNGQYVRILSC
>YP_009190175.1_Edwardsiella_phage_PEi20
-----SLNKGQRE-AFDII------TSA----I-QRR---------------------NG-----ERLTLNGPAGTGKTTLTKFIIQH---I-V--R--N------------------
-----GVLGVVLAAPTHQAKKVLA----KMSG---MEANTIHRV-LKI---N--P------MT-Y---E---D-------QD---V----------------------------F---
-E----Q---R-------------EMPDM--SK--CNVLVCDEASM--LDGKIFKI-ILNSIP-------------------------------------------------------
-----------------------------PW-----CVLIGIGDREQIQPVE-----------------PG---S--DGTPQ----------------------------I-------
--------------------------S-----------------PFFTH------------------P---S--------F-KQVHLTEVMR----S-------------N--A----
---------PIIDVATDIRTG---G--------WLRHHII--DGHGVHEFAS---------------------------TTALKD-FM-MQ--YF-----------------------
----------------------------------------------------------------DVVKTP-EDL------------------------------F-E-----------
T-------------------RMLAFTNKSVEKLNNIIRR-KL-----------------------------------------------------------------Y----------
-------------------------------------------------------------------------------------------------------------------E--
-----------------------------------------------------------------------------------------------------------------T-E--
-----V----------------------PFINEEVIVMQEPFIKELEFDGKKFS----E--I-----V-----FNNGEMVRIKDC
>YP_003934641.1_Shigella_phage_SP18
-----DLNTGQKE-AFDYI------TEA----I-QRR---------------------SG-----ECITLNGPAGTGKTTLTKFVIDH---L-V--R--N------------------
-----GVMGIVLAAPTHQAKKVLS----KLSG---QTANTIHSI-LKI---N--P------TT-Y---E---D-------QN---I----------------------------F---
-E----Q---R-------------EMPDM--SK--CNVLVCDEASM--YDGSLFKI-ICNSVP-------------------------------------------------------
-----------------------------EW-----CTILGIGDMHQLQPVD-----------------PG---S--T-QQK----------------------------I-------
--------------------------S-----------------PFFTH------------------P---K--------F-KQIHLTEVMR----S-------------N--A----
---------PIIEVATEIRNG---G--------WFRDCMY--DGHGVQGFTS---------------------------QTALKD-FM-VN--YF-----------------------
----------------------------------------------------------------GIVKDA-DML------------------------------M-E-----------
N-------------------RMYAYTNKSVEKLNNIIRR-KL-----------------------------------------------------------------Y----------
-------------------------------------------------------------------------------------------------------------------E--
-----------------------------------------------------------------------------------------------------------------T-D--
-----K----------------------AFLPYEVLVMQEPHMKELEFEGKKFS----E--T-----I-----FNNGQLVRIKDC
>DAF95488.1_Myoviridae_sp._ctCo31
-----QLNEDQRA-ALVTS------INI----L-------------------Y----NTR-----DSVCISGPAGTGKTFLTKVLLKI---L-E--S--L------------------
--Y-DS-SKIALSAPTHQAKKVLA----NSSG---RDAFTVHSL-FRI---L--P------NL-E---E---D-------RT-----------------------------E--F---
-T----Q---R-----------GDDLPKL--QD--ILFLVIDEVSM--IDEKLFKI-IYEKLP-------------------------------------------------------
-----------------------------MN-----TRIIALGDPYQLAPVN-----------------SE---S--I----------------------------------------
--------------------------S-----------------LFFTH------------------K---D--------F-TQIKLTKIMR----Q-S-----------S-GS----
---------PIIEQGDNIRRRVQNNLV-T------S----NDGKNGIFGFNT------------------------------EQE-FL-DK--YL-----------------------
----------------------------------------------------------------SIVKSA-DDA------------------------------I-D-----------
N-------------------RIIAYTNNKVNELNNIIRR-VI-----------------------------------------------------------------Y----------
----------K---------------------TDD-----------------------------------------------------------------------------------
----------------------------------------------------------------------------------------------------------------------
----------------------------QIVKGELLVLQQAVMND--------N----E--S-----V-----FDNGEILKVLNI
>YP_009006117.1_Vibrio_phage_VH7D
-----GLTNCQQG-AMNAF------LD-------------------------------SD-----GHMTISGPAGSGKTFLMKSILAA---L-D--A--K------------------
-----G-KNVAMVAPTHQAKNVLH----KMTG---RDVSTIHSL-LKI---H--P------DT-Y---E---D-------QK-----------------------------H--F---
-K----Q---A-----------G-DVEGL--DE--IDVLVVEEASM--IDNELYDI-MGKTMPR------------------------------------------------------
--------------------------K---------CRILGVGDKYQLQPVK-----------------HE-----------------------------------------------
--------------------------------------------PGIIS------------------PMFTK--------F-NTYEMTEVVR----Q-A----------KD--N----
```

```
---------PLIQVATEVRQG---EWLRT---NWSK-----ELRQGVLHVPN------------------------------VNK-ML-DT--YL-----------------------
----------------------------------------------------------------SKINTP-EDL------------------------------L-D-----------
Y-------------------RILAYTNDCVDTFNGIIRE-HI-----------------------------------------------------------------Y----------
----------N---------------------TSEP----------------------------------------------------------------------------------
-----F----------------------------------------------------------------------------------------------IP-----------N-E--
----------------------------YLVTQMPVMQSNGKY----------PVC-----------V-----IDNGEIVKILDV

>Hel_A_tha
-----MLTPEQRG-VYNEI------TEA----V-FNN---------------------LG-----GVFFVYGFGGTGKTFIWKTLSAT---I-R--Y--R------------------
-----D-QIVLNVASSGIASLLLE------GG------RTAHSR-FGI---P--L------NP-D---E---F-------SV---C---------------------KI--K--P---
-K----S---D-------------LANLV--KK--ASLVIWDEAPM--MSRFCFEA-LDKSFSDII----------------K---------N-----------------TD----N-
--------------------------TVFGG-----KVVVFGGDFRQVFPVI-----------------NG---A--G-RAE----------------------------I-------
------------------------VMS-----------------SLNAS------------------YLWDN--------C-KVLKLTKNTRL---L-ANNLSE-TEAKEI--Q----
---------EFSDWLLAVGDG---R-I-NE----SN-----DGVAIIDIPED----LLIT----------------N-ADKPIES-IT-NE--IY----G----DPKIL----H----
----------------------------------------------------------------E-ITDP-KFF------------------------------Q-G-----------
R-------------------AILASKNEDVNTINEYLLD-QL----H---A--------------------------E-----------------------E-R-I-Y----------
----------L---------------------SAD------SIDPT-----------------DS-----------------DS-LSNP----------------------VITPD--
-----F--------------------------------------------------------LNS-----------------------------------IK-----------L-P--
------GLP----------NHSL-----RLKVGAPVLLLRNLD----------P----K--G-----G-----LCNGTRLQITQL

>Hel_B_nap
-----KMTCEQRK-IYEEI------LSA----V-NKG---------------------DG-----GMFFVSGFGGTGKTFLWKLLSAA---I-R--S--R------------------
-----G-DIVLNVASSGIASLLLP------GG------RTAHSR-FGI---P--L------NP-D---E---F-------SS---C---------------------TM--K--H---
-G----S---D-------------QANLV--KA--SSLIIWDEAPM--MSKHCFEA-LDKSLSDIV----------------G---------K-----------------HD----T-
--------------------------QPFGG-----KVIVFGGDFRQVLPVI-----------------NG---A--G-RAE----------------------------I-------
------------------------VMA-----------------SLNSS------------------YLWTH--------C-KVLKLSKNMRL---L-SAGLSP-AEAKDL--Q----
---------EFSEWILKVGDG---K-L-SE----PN-----DGEAEIEIPSE----FLIT----------------D-CNDPIEA-IS-EE--IY----G----TTTSL----H----
----------------------------------------------------------------E-KKDA-NFF------------------------------Q-E-----------
R-------------------AILCPTNEDVNTVNEYMLD-KL----E---G--------------------------E-----------------------E-K-I-Y----------
----------N---------------------SAD------SIDPS-----------------DT-----------------CA-VNNE----------------------ALSAD--
-----F--------------------------------------------------------LNT-----------------------------------IK-----------V-P--
------GLP----------NHSL-----RLKVGCPVMVLRNIA----------P----T--D-----G-----LMNGTRLQITQL

>XP_033143622.1_Brassica_rapa
-----QLTDEQRS-VYEEI------MAS----V-NTS---------------------SG-----GVYFVYGYGGTGKTFIWNLLSAA---I-R--S--R------------------
-----G-DIVLNVASSGIAALLLP------GG------RTAHSR-FSI---P--I------NP-D---E---F-------ST---C---------------------KI--Q--P---
-G----S---D-------------QAELI--SK--ASLIIWDEAPM--MSKHCFEA-LDRSLCDIM----------------Q---------T-----------------TD----E-
--------------------------TPFGG-----KVVVFGGDFRQILPVI-----------------PK---G--N-RAD----------------------------I-------
------------------------VMA-----------------SLNSS------------------YLWKY--------C-KVLQLTKNMRL---F-S-EQDN-SAAEEI--A----
---------EFSKWILDVGDG---K-I-NE----PN-----SGETMIDIPED----ILIT----------------Q-CDDPIEA-IV-SE--VY----G------TTF----K----
----------------------------------------------------------------D-SKDP-IFF------------------------------R-E-----------
R-------------------AILSPTNEDVDVINNYMLD-HL----T---G--------------------------E-----------------------E-R-I-Y----------
----------L---------------------SSD------SIDPA-----------------DT-----------------KS-KDDS----------------------VFTPE--
-----F--------------------------------------------------------LNS-----------------------------------IK-----------T-S--
------GLP----------NHSL-----RLRIGTPVMLLRNLD----------T----T--E-----G-----LCNGTRLQITQV

>XP_006279329.2_Capsella_rubella
-----MLTEEQRR-VYQDI------IYS----V-NQN---------------------KG-----GMFFVHGFGGTGKIFLWSILGAD---I-R--S--K------------------
-----S-NIVLNVASSGIAALLLE------GG------RTAHSR-FCM---P--I------NI-N---E---Y-------SM---C---------------------SI--D--A---
-E----S---D-------------LAELI--RE--AKLIIWDEAPM--MNKHCFET-LDRTLQDIM----------------K---------C-----------------------N-
--------------------------RIFGG-----KVVVLGGNFRQILPVI-----------------PE---G--G-RVA----------------------------T-------
------------------------VLA-----------------SIKSS------------------LLWPS--------C-KVLKLTENMRL---R-K-GVNN-VQSDAL--A----
---------EFSKWLLDIGDG---K-I-NE----PN-----DGEVEIEIPED----LLTA------------------SEDPIHA-IV-HE--IY----G------KSF----A----
----------------------------------------------------------------K-ENDP-KFV------------------------------K-R-----------
R-------------------AILSPRNEDVDKINQYMLS-QL----P---G--------------------------E-----------------------E-R-R-Y----------
----------L---------------------SLD------SIETS-----------------DT-----------------SV-FDDM----------------------VYSQE--
-----F--------------------------------------------------------LNS-----------------------------------IN-----------V-S--
------GLP----------KHEL-----TLKKGAPIMLLRNID----------P----K--G-----G-----LCNGTRLIVTQM

>Hel_B_vul
-----CLTCEQRS-VYDEI------MMA----V-SRG---------------------QG-----GVFFVYGYGGTGKTYVWKTLCAA---I-R--S--K------------------
-----G-EIVLPVASSGIASLLLP------RG------RTAHSR-FGI---P--L------NV-S---E---N-------ST---C--------------------VGI--K--P---
-G----S---D-------------LAALL--MK--TKLIIWDEAPM--MHKYCFEA-LDRSLKDIM----------------Q---------S-----------------VDPSNKY-
--------------------------KPFGG-----LVVVFGGDFRKILPVI-----------------PK---G--S-RQD----------------------------I-------
------------------------VFS-----------------ALSSS------------------YLWNS--------C-KVLKLTRNMRL---Q-T-GSSE-TSLKEV--K----
---------EFSEWILSVGDG---N-A-GG----PN-----EGEAEIKVSND----ILIE----------------G-ESDPIAA-IV-ES--TY----P------LLK----D----
----------------------------------------------------------------H-LWEP-KYF------------------------------Q-E-----------
R-------------------AILAPTYEIVEMVNDYVLS-HL----P---G--------------------------E-----------------------E-K-L-Y----------
----------L---------------------SSD------AISNV-----------------EG-----------------NL-GASE----------------------IYSTE--
-----F--------------------------------------------------------LNT-----------------------------------IR-----------C-S--
------GLP----------NHHL-----RLKVGAPVMLLRNID----------Q----T--S-----G-----LCNGTRLVIKHL

>CAD1820584.1_Ananas_comosus_var._bracteatus
-----SLTDEQKG-VYETI------ISV----V-SKN---------------------EG-----GVFFLYGYGGTGKTFIWRTLSAA---I-R--S--K------------------
-----G-QIVLNVASSGIASLLLP------GG------RTAHSR-FGI---P--L------AI-T---E---E-------ST---C---------------------NI--K--Q---
-G----S---D-------------LAELL--IH--TKLIIWDEAPM--AHRFCFEA-LDRTLRDIL----------------R---------F-----------------SNPSSCE-
--------------------------QPFGG-----KVVVFGGDFRQILPII-----------------PK---G--T-RQD----------------------------I-------
------------------------IFA-----------------TINSS------------------YLWSF--------V-KILTLTKNMRL---E-T-GSSN-YNLEEM--R----
---------EFSKWILSVGDG---D-A-GE----EN-----DGEEIIEIPDE----FLIK----------------E-SINPIVS-IV-DS--TY----P------SLL----S----
----------------------------------------------------------------N-IHDL-QYL------------------------------Q-E-----------
R-------------------AILAPTLEIVDAVNEYMLS-LI----H---G--------------------------E-----------------------E-K-V-Y----------
----------L---------------------SSD------SVCKT-----------------DV-----------------GLDGPED----------------------VYTPD--
-----F--------------------------------------------------------LNS-----------------------------------IK-----------C-S--
------GVP----------NHML-----KLKQGAPVMLLRNID----------K----S--S-----G-----LCNGTRLVITQL
```

```
>XP_012840144.1_Erythranthe_guttata
-----SITDEQRK-VYDVI------MDA----V-TND---------------------SG-----GMFFLYGHGGTGKTFLWKTLSAA---V-R--S--K------------------
-----G-KIVINVASSGIASLLLP------GG------RTTHSR-FGL---P--I------DV-H---E---S-------ST---C---------------------SI--S--Q---
-Q----S---P-------------HAELL--IR--AKLIIWDEAPM--MHRYCFEA-LDKTMKSIL----------------Q---------T-----------------------D-
--------------------------KPFGG-----KVVILGGDFRQILPVV-----------------LK---A--S-RQD----------------------------I-------
------------------------VHA-----------------TINSS------------------PLWNF--------C-RVMKLTKNMRL---Q-S-CCSP-SNVDEI--K----
---------EFGDWILNVGNG---D-V-GE----DN-----DGEASIEIPDD----MLIG----------------D-SEAPFRD-LL-EF--VY----P------DLL----S----
----------------------------------------------------------------N-MYDR-DYF------------------------------Q-G-----------
R-------------------AILAPTNECVESVNDHLMS-LL----P---G--------------------------E-----------------------E-K-V-Y----------
----------L---------------------SSD------SMCRD-----------------EH-----------------TTEDNAE----------------------IYSTE--
-----I--------------------------------------------------------LNT-----------------------------------IR-----------C-S--
------GVP----------SHAL-----RIKVSAPVMLIRNID----------Q----A--R-----G-----LCNGTRLQIIRT
>Hel_H_ann
-----LLTEEQRS-VFQQI------INA----V-EGN---------------------KG-----GVFFVYGYGGTGKTFLWKTLSAA---I-R--S--K------------------
-----G-QIVLNVASSGIASLLLS------GG------RTAHSR-FRI---P--L------NL-T---E---D-------SV---C---------------------HI--K--P---
-N----G---D-------------VARLL--HE--TNLIIWDEAPM--VHKHAFEA-LDRTMNDIF----------------N---------I-----------------ETSNRSN-
--------------------------IRFGG-----KVIVLGGDFRQILPVV-----------------PN---G--G-RQE----------------------------I-------
------------------------VNA-----------------SISSS------------------YLWNT--------C-KLMRLTKNMRL---T-V-GSSA-SDAEEI--K----
---------QFAKWLLDIGEG---N-V-GG----PN-----DGEASIEIPSD----LLIT----------------D-TSDPIST-LI-DF--VY----P------SIL----E----
----------------------------------------------------------------N-FNNQ-NYF------------------------------S-E-----------
R-------------------AILAPKNEVVHEINDRLLS-LF----P---G--------------------------E-----------------------E-R-E-Y----------
----------L---------------------SSD------SLCQS-----------------ED-----------------PNATQQK----------------------LYSPD--
-----V--------------------------------------------------------LNG-----------------------------------LK-----------V-S--
------GLP----------NHRL-----ALKVGVPVMLLRNID----------Q----Q--N-----G-----LCNGTRLQVKKM
>XP_022031972.1_Helianthus_annuus
-----LLTDEQRN-VFDQI------MES----V-RTN---------------------KG-----GVFFVYGYGGTGKTFLWKTLSAA---I-R--S--K------------------
-----S-EIVLNVASSGIASLLLS------GG------RTAHSR-FSI---P--L------NL-N---E---D-------SL---C---------------------RM--N--P---
-G----S---E-------------LACLL--KK--TQLIIWDEAPM--IHKHAFEA-LDRTLKDIL----------------M---------P-----------------DCSNSEA-
--------------------------LPFGG-----KVIVFGGDFRQILPVV-----------------PN---G--S-RQD----------------------------I-------
------------------------VNA-----------------SLSSS------------------YIWNK--------C-KLLRLTKNMRL---T-V-GMNH-GDIDKT--K----
---------EFAKWLLDIGEG---K-L-GG----RN-----DGEALIDIPQE----LLIT----------------E-STNPIGN-LI-NF--VY----P------SIL----E----
----------------------------------------------------------------S-FNDP-NYF------------------------------Q-E-----------
R-------------------AILAPKNDVVHEINDTLLA-MF----P---G--------------------------D-----------------------H-K-E-Y----------
----------L---------------------SSD------SICQS-----------------EN----------------VTDHIRHN----------------------VYPPD--
-----V--------------------------------------------------------LNG-----------------------------------LK-----------V-S--
------GMP----------NHKL-----VLKVGVPIMLLRNLD----------Q----K--N-----G-----LCNGTRLQVVKL
>ABA95557.1_Oryza_sativa_Japonica_Group
-----SLNTDQKK-AFDAI------MES----I-NGG---------------------QG-----KQIFVEGYGGTGKTYLWKALTTK---L-R--S--E------------------
-----G-KIVLAVASCGIAALLLQ------GG------RTAHSR-FRI---P--I------KI-T---E---E-------ST---C---------------------EI--K--Q---
-G----T---H-------------LAELL--KR--TSLILWDEAPM--ANKHCFEA-LDKSLRDIM----------------R---------F-----------------TNENSSE-
--------------------------RPFGG-----MTVVLGGDFRQILPVI-----------------PK---G--R-REN----------------------------I-------
------------------------VNA-----------------SIKRS------------------YLWNH--------F-EIIKLTENMRL---S-C-MSNEPLEKQKV--A----
---------EFAKWILHIGDG---A-S-A-----SD-----EGEEWVKIPSD----ILLQ----------------K-GQDPKET-IV-KS--IY----P------NLL----D----
----------------------------------------------------------------N-YRER-EFL------------------------------E-E-----------
R-------------------AILCPRNETVQEINEYIMN-QI----Q---R--------------------------E-----------------------E-M-T-Y----------
----------L---------------------SCD------TVCKA-----------------MT-----------------NNSSMEH----------------------MYPTE--
-----F--------------------------------------------------------LNT-----------------------------------LK-----------F-P--
------GIP----------NHEL-----KLKVGLPVMLLRNIN----------Q----T--A-----G-----LCNGTRMTITQL
>RCV07316.1_Setaria_italica
-----KLNLDQRK-AFDAI------TQS----V-NSK---------------------LG-----KLIFVNGYGGTGKTFLWKAITKS---L-R--S--E------------------
-----G-KIVLAVASSGIAALLLP------GG------RTAHSR-FHI---P--L------NI-N---N---E-------ST---C---------------------DI--K--Q---
-G----S---L-------------LAELL--NK--TSLILWDEAPM--TNKHCFEA-LDKSLRDIL----------------R---------F-----------------TDENSKD-
--------------------------KPFGG-----MTIVMGGDFRQTLPVI-----------------PK---G--R-RTH----------------------------I-------
------------------------IDA-----------------SLKRS------------------YLWKH--------F-EEIKLTTNMRL---T-A-VTNSTEEKKKI--Q----
---------EFADWILSIGDG---L-A-G-----DK-----DDEAWITIPQD----LILQ----------------K-GEDELET-IV-NN--TY----P------DLS----R----
----------------------------------------------------------------N-YSNR-TYL------------------------------E-E-----------
R-------------------AILCPRNEMVDNINSYIMS-QI----P---G--------------------------E-----------------------E-T-T-Y----------
----------L---------------------SSD------TVCKA-----------------IS-----------------TKESEDQ----------------------LYPTE--
-----F--------------------------------------------------------LNS-----------------------------------LK-----------F-P--
------GIP----------NHKL-----QLKVGLPIMLLRNIN----------Q----S--A-----G-----LCNGTRLTITQL
>XP_039834415.1_Panicum_virgatum
-----QLNYEQRH-IYDVV------IQS----V-YGK---------------------TG-----RCFFVYGYGGTGKTFLWNAIISR---L-R--S--E------------------
-----K-HIVLAVASSGVAALLLL------GG------RTAHSR-FKI---P--I------VI-D---E---S-------SM---C---------------------DI--K--R---
-G----T---F-------------LADLI--VQ--SSLVIWDEAPM--THRHCFES-LDRSMRDIL----------------G---------Q-----------------LDSSNSD-
--------------------------RMFGG-----KTMLLGGDFRQVLPVV-----------------EG---G--N-RLD----------------------------T-------
------------------------IDA-----------------SITNS------------------YLWDH--------V-KILKLTTNMRI---L-G-MGRSGLAAKEV--K----
---------DFSDWVLSVGDG---T-T-KGTAEIDD-----GDSELIEIPSD----ILVP----------------R-LDSAIDD-II-SS--TY----P------NLG----A----
----------------------------------------------------------------S-YSDP-TYL------------------------------R-E-----------
R-------------------AIIAPKNDTIDEINSRILS-LV----P---G--------------------------N-----------------------E-K-V-Y----------
----------L---------------------SSD------TLVES-----------------SK-----------------ENGNLDL----------------------LYPVE--
-----F--------------------------------------------------------LNS-----------------------------------LQ-----------F-K--
------GIP----------HHKL-----MLKVGSPVMLLHNLN----------Q----S--A-----G-----LCNGTRLIITQL
>XP_026386115.1_Papaver_somniferum
-----NLNEEHTR-VFEAV------MRS----V-EDS---------------------KG-----GLFFVYGSGGTGKTYLWRTIITA---L-R--A--Q------------------
-----S-KIVLAVASSGIASLLLP------GG------RTAHSQ-FKI---P--F------KL-Y---D---N-------ST---C---------------------TV--N--K---
-K----S---D-------------LAELI--CK--ADLIIWDEAPM--INKHALEA-LERTVTDIM----------------T---------K-----------------DDTVSPK-
--------------------------PIFGG-----KTLLLGGDFRQILPVI-----------------QK---G--S-REM----------------------------I-------
------------------------VDS-----------------SISRS------------------KLWKH--------F-KIFKLSTNMRL---M-N-ADSDDAQQQEI--A----
```

```
---------DFGKWVLDVGDG---K-I-PIS--ETK-----DDSTWIQIPDD----LLVKC---------------D-NGDYINT-IV-ES--TY----P------SLL----E----
----------------------------------------------------------------R-YVDY-RSL------------------------------E-E-----------
R-------------------CILAPTNESADQINEHMIS-LI----P---G--------------------------E-----------------------D-H-V-F----------
----------R---------------------SAD------SISPE-----------------TS-----------------DFQSKEV----------------------FYTNE--
-----F--------------------------------------------------------LNS-----------------------------------LT-----------F-S--
------GFP----------NHEI-----YLKVGIPIMLLRNLK----------Q----S--E-----G-----LCNGTRLIVTQI
```

```
>PIA60703.1_Aquilegia_coerulea
-----NLNEEQRF-VFDKV------VHA----V-ENA---------------------KG-----GMYFVYGSGGTGKTFLWKTIISS---L-R--S--K------------------
-----G-KIVLVVASSGIASLLLP------GG------RTAHSR-FKI---P--L------EV-D---D---Y-------ST---C---------------------FI--S--Q---
-K----S---D-------------LAQLI--KH--ADLVIWDEAPM--NHRNIFEA-VDKTFQDLM----------------R---------K-----------------EIGDSDG-
--------------------------QIFGG-----KTILLGGDFRQTLPVV-----------------PK---G--S-RED----------------------------V-------
------------------------VTS-----------------SISRS------------------YLWSK--------C-QVFVLKTNMRL---R-G-NDLNSEMAKEI--E----
---------EFSEWVLQLGEG---K-L-PTKTMNTY-----DEPNWIQIPDD----LLLR----------------N-N---------------------------------------
----------------------------------------------------------------------------------------------------------G-----------
R-------------------CILTPTNDCADKVNKEVLS-RI----Y---T--------------------------T-----------------------S-R-T-Y----------
----------A---------------------SAD------TISPM-----------------SE-----------------LVNEQD--------------------------LE--
-----Y--------------------------------------------------------LNH-----------------------------------LE-----------V-S--
------GVP----------NHLL-----ELKVGIPVMLVRNIN----------P----S--R-----G-----LCNGTRLVVTSL
```

```
>XP_039793773.1_Panicum_virgatum
-----SLNKEQRA-AYDEI------LSY----I-DSK---------------------DG-----GLFFLDGPGGTGKTFLYRALLAK---V-R--S--Q------------------
-----N-KIAVATATSGVAASIMP------GG------RTTHSC-FKI---P--L------TI-E---S---G-------GY---C---------------------SF--T--K---
-Q----S---G-------------TATLL--HT--ASLIIWDEVSM--IKKQAVEA-LDNSMRDIM----------------D---------R----------------------PD-
--------------------------LPFGG-----KTIVFGGDFRQVLLVV-----------------GK---G--S-RAQ----------------------------I-------
------------------------VDA-----------------SLRRS------------------YLWGY--------M-RHLKLVRNMRA---H-S-----------D--P----
---------WFAEYLLRIGNG---T-E-------ES-----NADGEVCLPDE----ICVPYT--------------G-DDNDLDR-LI-QC--IF----P------NLN----E----
----------------------------------------------------------------N-MVDK-DYI------------------------------T-S-----------
R-------------------AILSTRNDWVDSINMKMIG-YF----Q---G--------------------------G-----------------------E-V-E-Y----------
----------Y---------------------SFD------SAVDD----------------------------------------PHN----------------------YYPSE--
-----F--------------------------------------------------------LNT-----------------------------------LT-----------P-N--
------GLP----------PHVL-----KLKVGCPIILLRNID----------P----A--N-----G-----LCNGTRLVVRGF
```

```
>TVU37829.1_Eragrostis_curvula
-----ILNAEQRA-GFDEI------MDH----V-TSE---------------------KG-----QVFFVDGPGGTGKTYLYKALIAT---V-R--S--M------------------
-----G-FIAVATATSGIAASIMP------GG------RTAHSR-FKI---P--I------KI-G---D---E-------SM---C---------------------NF--T--K---
-Q----S---G-------------TAELL--RS--ARLLIWDEVAM--TKRQSIEC-LDRSLQDIM----------------G---------C-----------------------D-
--------------------------EPFGG-----KIMVFGGDFRQVLPVV-----------------PR---G--T-RAQ----------------------------I-------
------------------------TNA-----------------TLQRS------------------YIWDR--------I-RKIRLTQNMRA---Q-S-----------D--P----
---------LFSQYLLRVGDG---V-E-------ES-----VGDDYIRLPEE----IVIDYD--------------E--EKGIEK-LV-ED--IF----P------DLL----A----
----------------------------------------------------------------N-VSDA-VYM------------------------------S-S-----------
R-------------------AILSTKNEYVDQLNSKMIE-TF----P---G--------------------------P-----------------------S-K-V-F----------
----------Y---------------------SFD------SVEDD----------------------------------------QTN----------------------NYPID--
-----F--------------------------------------------------------LNS-----------------------------------LT-----------P-N--
------GLP----------PHEL-----KIKVNCPLILLRNLD----------P----H--N-----G-----LCNGTRLVVRGF
```

```
>XP_020197274.1_Aegilops_tauschii_subsp._strangulata
-----KLNSEQRL-AFDEI------MTH----V-LHQ---------------------KS-----MVFFIDGPGGTGKTYLYKALLAK---V-R--S--M------------------
-----G-LIAIATATSGIAASIMP------GG------RTAHSR-FKI---P--I------NI-Q---D---D-------SM---C---------------------NF--S--K---
-Q----S---G-------------TAELL--RR--SSLIIWDEVAM--KKRQAVEA-LDRSLQDIT----------------G---------C-----------------------G-
--------------------------SPFGG-----KVVVFGGDFRQVLLVV-----------------RH---G--T-RAQ----------------------------I-------
------------------------TDA-----------------TLKKS------------------YLWPD--------I-RHIKLWRNMRA---L-F-----------D--P----
---------WFSDFLLRIGNG---T-E-------ES-----IGQDYVRLPEE----IVIGYT--------------D-VKASVGK-LI-DE--IF----P------SMD----K----
----------------------------------------------------------------N-GNSP-SYI------------------------------S-A-----------
R-------------------AILSTKNEYVDELNEMLID-RF----P---G--------------------------E-----------------------E-K-V-Y----------
----------Y---------------------SFD------SVVDD----------------------------------------PHN----------------------HYQPE--
-----F--------------------------------------------------------LIS-----------------------------------LT-----------P-N--
------GLP----------PHIL-----RLKINCPVILIRNLD----------P----S--N-----G-----LCNGTRLIIKAF
```

```
>KAF5187279.1_Thalictrum_thalictroides
-----KLNNEQKH-AFDMI------MDA----V-HHK---------------------TS-----SVFFIDGPAGTGKTFLYRSLLAA---I-R--H--E------------------
-----G-HIALATATSGIASIMMP------GG------RTAHSR-FKI---P--I------PT-L---P---T-------ST---C---------------------RI--S--K---
-Q----S---D-------------EGILL--HE--TTLIIWDEATM--AHRYTIEA-LDKTLRDLF----------------H---------N-----------------------D-
--------------------------QPFGG-----KIVVLGGDFRQVLPVV-----------------PR---G--T-RSQ----------------------------A-------
------------------------IDA-----------------CITYS------------------SLWDH--------V-KLFHLTQNMRA---R-T-----------D--S----
---------LYSDMLMRIGNG---S-E-------PY-----VVDDLIRMPDE----IVIPWE--------------G--EQSILQ-LI-NA--VF----P------KMS----D----
----------------------------------------------------------------N-AYDR-NYI------------------------------M-E-----------
R-------------------AIITPKNNYVDQLNHQVLQ-LF----P---G--------------------------N-----------------------E-I-I-F----------
----------H---------------------SFD------SAEND----------------------------------------PRN----------------------LYQLE--
-----L--------------------------------------------------------LNS-----------------------------------IS-----------T-S--
------QLP----------PHKL-----TVKIGCPMIVLRNLD----------P----K--N-----G-----VCNGTRVLLRGI
```

```
>XP_026391420.1_Papaver_somniferum
-----KLNEDQSR-AYKTI------MEA----I-ERK---------------------ES-----KVFFIDGPGGTGKTYLCRAILAT---V-R--K--N------------------
-----G-GIALATTTSGIAATMLP------GG------RTAHSR-FQL---P--M------TP-T---S---T-------ST---C---------------------RT--K--K---
-Q----T---E-------------EAKLL--RH--GIVLMWDEATM--AHPYSLEA-FDRTMRDIT----------------G---------I-----------------------E-
--------------------------EPFGG-----KILIMGGDFRKVLPVI-----------------PR---S--T-RGQ----------------------------T-------
------------------------VDA-----------------CLSRS------------------HLWEN--------V-HVLHLKKNMRA---A-E-----------D--A----
---------SYSEFLIRVGDG---D-E-------PC-----IANERIKVPEE----MVIPWV--------------S--DASLAQ-LI-DV--TF----P------NLV----E----
----------------------------------------------------------------N-ARDV-DYM------------------------------V-N-----------
S-------------------ALITPLNECVEKLTDRVFS-IF----P---G--------------------------E-----------------------E-V-L-F----------
----------Y---------------------SFD------PVDDD----------------------------------------THG----------------------LYQQE--
-----Y--------------------------------------------------------LNN-----------------------------------ID-----------P-G--
------GLP----------SHIL-----KLKIGAPIMLLRNVD----------A----K--N-----G-----LCNGTRLIIKEF
```

```
>KAG0566608.1_Ceratodon_purpureus
-----QLNGEQRY-CYDAI------LSS----I-EDR---------------------SG-----VVFFVNGPAGTGKTFLYNIVTAN---V-R--S--R------------------
-----G-KIVLCVASSGIAALLLH------GG------RTAHST-FKI---P--F------EV-D---E---F-------SM---C---------------------TI--N--K---
-N----S---E-------------YADVF--RE--ASLIIWDEVPM--QHRHCAEA-VDRSLRDIR----------------D---------S-----------------------N-
--------------------------SPFGG-----VTVVFGGDFRQILPVI-----------------PR---G--S-RPQ----------------------------I-------
------------------------VGA-----------------CLRRS------------------TIWQH--------V-RIMNLSINMRL---Q-N-ASLA------N--R----
---------EFAQWLLQVGDG---S-N------FDD-----ANCNMIQLHNW----INI--------------------VSSIRC-LI-DN--IY----N------NID----DM---
----------------------------------------------------------------S-LHED-QYF------------------------------R-D-----------
R-------------------TILSARNTDVDLINKEILQ-SF----P---G--------------------------N-----------------------L-E-T-F----------
----------R---------------------SAD------SNTVE-----------------AG-------------------ADNHA----------------------AYPSE--
-----Y--------------------------------------------------------LNS-----------------------------------LD-----------L-S--
------GIP----------LSKL-----DLKIGCPIILLRNLA----------P----K--Q-----G-----LCNGARMVLTRF
>Hel_A_ara
-----QLNDEQRM-VYETV------TAA----I-DRQLATAA----SQANAG------DQ-----RLFFLDGPGGTGKSFLVEKILAH---V-R--R--C------------------
-----G-EIALATAASGIAALLLT------GG------KTVHST-FKL---P--L------DL-N---N---H-------ST---C---------------------SI--T--V---
-Q----S---K-------------RAEML--RQ--TALIVWDEASM--SSRFALEA-VDRTLQDIT----------------G---------V-----------------------Q-
--------------------------LPFGG-----KVVLLSGDFRQILPIV-----------------PK---G--T-DAQ----------------------------I-------
------------------------INE-----------------CIKKS------------------TLWPL--------F-RSLQLRDNMRV---R-T-APNA-NQASEL--R----
---------DFANLLLRIGEG---R-H-DTF---AG-----LDPSLAKIPHD----MIVPHT-------------AN-PTNDLNT-LI-DK--IY----P------DMQ----R----
----------------------------------------------------------------H-FQHP-SFF------------------------------S-D-----------
R-------------------AILSPLNVDVASVNNLVLD-RI----P---G--------------------------P-----------------------E-Q-E-Y----------
----------R---------------------SVD------TLVNP-----------------EE--------------------HEHL----------------------QLPSE--
-----Y--------------------------------------------------------LNT-----------------------------------LN-----------V-S--
------GIP----------VHRL-----RLKRFAPVLLLRNLN----------S----D--M-----G-----LCNGTRLQIVGL
>Hel_P_inf
-----QLNESQRV-VYDQI------IEA----V-ECP-------------EE------GK-----KLFFVDGPGGTGKSTLLRNILAK---V-R--L--S------------------
-----G-KIAIAVASSGIASLLLM------GG------RTAHST-FKI---P--L------KL-N---E---S-------ST---C---------------------GI--R--K---
-N----S---H-------------IQELI--KH--ASLIIWDEAPM--AHRHAFEA-VDRTLRDIL----------------D---------N-----------------------D-
-------------------------TEPFGG-----KVFVLSGDFRQILPVV-----------------KN---G--T-PVE----------------------------T-------
------------------------IDA-----------------CLKSS------------------RLWPQ--------F-QTFRLTENMRV---R-T-ADTA-DTAEEM--A----
---------AFSELLLQVGEG---R-H-DVN---PS-----LGNEYMKIPRD----MLIEN--PPVPEDEDREIRPGVIPRGMDR-II-DE--MY----G------EIN----NP---
----------------------------------------------------------------E-VATD-EYF------------------------------A-N-----------
R-------------------TILTTTNAIVHRINEAVTD-RL----T---G--------------------------Q-----------------------A-R-E-Y----------
----------M---------------------SSD------SVQDD---------------------------------------GDGN----------------------FFEQE--
-----V--------------------------------------------------------LHS-----------------------------------MN-----------I-S--
------GMP----------PHKL-----TLKVGMPIMMMRNLN----------P----D--L-----G-----LCNGTRLRIVAL
>KAE9276432.1_Phytophthora_rubi
-----QLNDGQRA-IYDEI------LQA----V-DGS-------------AV------GE-----KLFFIDGPGGTGKSTLLRHILAK---V-R--L--S------------------
-----G-KIAIAVASSGIASLLLM------GG------RTAHST-FRI---P--L------KL-N---D---K-------ST---C---------------------AI--Y--K---
-Q----S---N-------------LKTLI--QR--ASLVIWDEAPM--THRHAFEA-VDRTLRDIM----------------D---------N-----------------------D-
-------------------------QEPFGG-----KVFVLSGDFRQILPVV-----------------VR---G--T-PAE----------------------------T-------
------------------------IDA-----------------CLKSS------------------SLWSH--------F-KQVHLTENMRV---Q-S-ARSE-STAAEL--A----
---------AFSEFLLQVGEG---R-H-EVN---RS-----LGKDFVKIPRD----MLIDNTEPDQDMDEDEDILPGAVPRGLKN-II-DV--MY----A------DIN----NP---
----------------------------------------------------------------D-IATD-EYF------------------------------A-D-----------
R-------------------TILTTTNAVVQGINEAVSQ-RL----S---G--------------------------D-----------------------S-H-E-Y----------
----------L---------------------SVD------SVDDD---------------------------------------NEGN----------------------FFEPE--
-----V--------------------------------------------------------LHT-----------------------------------VN-----------I-N--
------GIP----------PHKL-----TLKEGAPIMMMRNLN----------P----D--L-----G-----LCNGTRLRVVKL
>Hel_C_sup
-----QLNEEQRI-AYDRL------IQA----V-NSG---------------------SG-----GIYFLDSPGGTGKTFLITLLLAK---I-R--S--Q------------------
-----N-EVALALASSGIAATLLE------GG------RTAHSA-LKL---P--L------NM-H-INE---T-------PV---C---------------------NI--A--K---
-N----S---A-------------MAKTL--QV--CKLIIWDECTM--AHKRSLEA-LDRTLKDLR----------------D---------N-----------------------Q-
--------------------------NIFGG-----AMILLSGDFRQTLPVI-----------------PR---S--T-VAD----------------------------E-------
------------------------INA-----------------CLKSS------------------NLWRH--------V-KTLQLTTNMRV---F-L-QQDQ-----TA--T----
---------VFSKQLLDIGNG---K-V-A-----VD-----SSTGLMTFPTD----FCHF-------------------TESKEE-LI-QR--VF----P------DIK----Q----
----------------------------------------------------------------Q-YNNH-DWL------------------------------S-E-----------
R-------------------AILAAKNKDVDDLNATIQN-FL----P---G--------------------------E-----------------------L-F-T-Y----------
----------K---------------------SVD------TATNQ---------------------------------------DDVV----------------------NYPTE--
-----F--------------------------------------------------------LNS-----------------------------------LD-----------L-P--
------GLP----------PHNL-----KLKVGSVVIMLRNIN----------Q----P--R-----------LCNGTRLVVKKL
>Hel_M_dem
-----LMNEEQRT-IYDRI------MLA----V-SAG---------------------QG-----GFFFLDAPGGTGKTFVISLILAE---I-R--S--N------------------
-----N-GIALAVASSGIAATLLD------GG------RTAHSV-FKL---P--L------NI-Q-NNP---D-------AV---C---------------------NI--K--K---
-Q----S---S-------------MATVL--KR--CKIIIWDECTM--AHKYSLEA-LNRTLKDIK----------------N---------S-----------------------D-
--------------------------KLFGG-----TLLVLSGDFRQTLPVI-----------------PR---S--T-YAD----------------------------E-------
------------------------INA-----------------CLKSS------------------PLWRN--------V-EKLQLKINMRV---Q-M-LQDP-----SA--E----
---------TFSKQLLDIGDG---K-V-A------I-----DETGYVKLPTD----FCTI-------------------ADSQDT-LI-EQ--IF----P------DVH----T----
----------------------------------------------------------------R-YINH-EWL------------------------------A-E-----------
R-------------------VILAAKNVDVDNLNLKIQM-LL----P---G--------------------------N-----------------------L-V-S-Y----------
----------K---------------------SID------TVCDD---------------------------------------SEAV----------------------NFPTE--
-----F--------------------------------------------------------LNS-----------------------------------LD-----------L-P--
------GMP----------PHNL-----QLKVGSPIILLRNLN----------P----P--R-----------LCNGTRLVIQKL
>CEO98944.1_Plasmodiophora_brassicae
-----NMNQDQRA-ALDRI------IAS----V-RNP-------------AD------SE-----KTFFVDGPGGTGKTTLFTTLLKL---A-R--S--Q------------------
-----Q-VRCLAVASSGIAACLLP------AG------RTAHSA-LAI---P--L------EI-H---D---K-------ST---C---------------------MV--N--A---
-E----S---D-------------LANRL--RR--TSLMVYDEVAM--AHRYAPEA-VDRTLRDIR----------------S---------V-----------------------D-
--------------------------LPNGS-----MIVVYGGDFRQILPVI-----------------PG---G--S-RRQ----------------------------V-------
------------------------VQA-----------------CLKKS------------------YLWRH--------V-KVLPLTINMRL---Q-T-----------R--P----
```

```
---------DFQQYLLDVGEG---K-S-GPE----------------VVPQP----WMQT----------------E--GNSKES-LI-TE--IF-----------------------
-------------------------------------------------------------------NDP-NDF------------------------------S-D-----------
R-------------------VILTVRNDDAQDINRKITE-ML----P---G--------------------------E-----------------------K-I-S-V----------
----------Y---------------------SAD------KVAND---------------------------------------DDAA----------------------LYPIE--
-----F--------------------------------------------------------LNS-----------------------------------LL-----------P-S--
------GVA----------PHHL-----DLKVGQYVMLLRNLN----------P----A--R-----G-----LCNGTRMQVKQV

>Hel_X_lae
-----SLNTLQST-AFNKI------IIA----A-EDN-------------RT------MP-----KCYFLDGPGGSGKTYLYETLIHF---F-R--A--K------------------
-----N-LSFLASATTGIAANLLI------DG------RTCHSL-FKL---P--V------PI-T---E---T-------SV---S---------------------NM--K--M---
-D----S---D-------------SANEI--RL--AKLLILDECTM--ASSHLLNT-IDKLLRELM----------------D---------N-----------------------D-
--------------------------IPFGG-----KLLLLGGDFRQCLAIV-----------------PH---A--M-RSA----------------------------I-------
------------------------VQS-----------------SLKYA------------------ENWHY--------F-EKVTLVENMR----C-A-----------D--P----
---------QYNNWLLLLGNG---K-L-TND---FE-----LHPDIIQIPKE----FIC------------------------ED-LV-TE--IF----G------KEI---------
----------------------------------------------------------------S-LDQI-PFL------------------------------A-K-----------
R-------------------AILSPKNIDVDMINNQVIA-LL----P---G--------------------------Q-----------------------S-C-V-F----------
----------L---------------------STD------CIDSE-----------------DE--------------------SEKL----------------------NFPLE--
-----Y--------------------------------------------------------LNT-----------------------------------IN-----------P-A--
------GLP----------QHNL-----ILKVGTIVMLLRNLN----------T----K--Q-----G-----LCNGTRLVVKSM

>Hel_F_can
-----TLNKEQLQ-AFEKI------ETA----M-NSS-------------DG------TE-----KCFFLDGPGGSGKTYLYKTFLSH---V-R--G--Q------------------
-----G-ETALPVASTGIAANLLK------GG------RTYHSQ-YKV---P--I------NL-N---E---T-------SV---S---------------------GI--E--M---
-T----S---K-------------DAKVI--RD--AKLLIWDEATM--ASANALHC-IDRLLKEIM----------------K---------S-----------------------D-
--------------------------LAFGG-----KVLLLGGDFRQTLPII-----------------PH---A--D-AVA----------------------------I-------
------------------------VQA-----------------SIKFS------------------HLWRK--------F-QVLKLDSNVR----S-T-----------D--I----
---------EYSEWLMKLGDG---E-L-TNE---HS-----LGENIIEIPES----MLAS-----------------------EN-IV-KD--IF----G------DCL---------
----------------------------------------------------------------T-PENV-EQF------------------------------C-N-----------
R-------------------AILCPTNAEVDKINNQVLQ-IL----Q---G--------------------------E-----------------------C-K-T-Y----------
----------L---------------------STD------SIVTD-----------------ED--------------------SSRD----------------------DYPVE--
-----F--------------------------------------------------------LNT-----------------------------------LN-----------P-S--
------GSS----------PHEL-----KLKVGALIMLLRNLN----------T----K--R-----G-----LCNGTRLVVTEL

>Hel_E_jap
-----KLNVEQKV-ISDKV------LHA----V-KNK---------------------IP-----NCYFIDGPGGSGKTFIYQTLCYM---L-R--S--E------------------
-----N-KVVLPVAWTGIAASLLP------GG------RTSHSI-FKL---P--V------PI-L---D---T-------SV---S---------------------SI--R--T---
-H----T---K-------------DAQLL--RE--SDLIIWDEVSM--VPKDALRI-VDRLLKDIM----------------N---------N-----------------------N-
--------------------------LPFGG-----KIILFGGDFRQVLPVV-----------------RH---A--S-RTA----------------------------I-------
------------------------VEN-----------------TVKRS------------------PLWSH--------V-TTYKLTQNMRT---C-N-----------D--A----
---------VFTEWLLKLGNG---N-L-EAQ---TD-----YYDEAIAIPRN----CYCH----------------------YDE-LI-TT--IF----N----VPEI----------
----------------------------------------------------------------N-EQNV-SQF------------------------------Y-S-----------
M-------------------AILCPKNDECITINEYIISNLL----P---G--------------------------E-----------------------E-K-I-Y----------
----------L---------------------SSD------SVQAD-----------------ET--------------------DNNQ----------------------LYPME--
-----F--------------------------------------------------------LNS-----------------------------------LN-----------P-S--
------GLP----------PHKL-----LLKKNTVIMLIRNLN----------A----N--Q-----G-----LINGTRLVVTDL

>Hel_C_ele
-----TLNDQQKR-AADQI------LAA----L-DDAS--------------------LP-----RLFYLDGPGGSGKTYLYITLYNI---C-V--G--R------------------
-----G-LKVACTAWTGIAANLLP------LG------RTSASL-FKL---D--I------RN-Q---C---K-------SS---L---------------------H---Q--R---
-Q----L---K-------------EAQEL--AE--NDVFIWDEASM--VPKTALDT-VDVLLRDLT----------------K---------I-----------------------D-
--------------------------QPFGG-----KILILGGDFRQILPVV-----------------ER---S--S-RAD----------------------------Q-------
------------------------VDA-----------------CIKRS------------------PLWTE--------F-QILHLISNMRV---T-S----------GD--S----
---------DWIQFLLNVGDG---S-A-N------------DSDSKVTLPLS----VMCDHN------------------------IV-EE--VF----G------AVI----DP---
----------------------------------------------------------------T-TSDP----------------------------------C-D-----------
N-------------------VILTPKNVDVAQLNDDVHN-RM----V---G--------------------------E-----------------------E-R-I-Y----------
----------L---------------------SRD------EVIVE-----------------HQ--------------------ADTM----------------------HYPTE--
-----F--------------------------------------------------------LNK-----------------------------------MS-----------P-S--
------SLP----------PHIL-----KLKKGSVIILLRNLD----------V----S--A-----G-----LCNGSRFIVETL

>Hel_A_can
-----TLNAQQER-ACNTI------LSS----V-SDPT--------------------RP-----RLFFIDGPGGSGKTYLYNALFNI---L-I--G--Q------------------
-----N-NKVICTAWTGIAANLLP------NG------RTAASL-FKL---D--I------GN-D-------L-------KT---S---------------------SM--R--R---
-Q----Q---K-------------EARAL--AE--VNVIIWDEASM--IPRRALET-VDELLRDIM----------------Q---------N-----------------------E-
--------------------------QPFGG-----KTMILGGDFRQVLPVV-----------------QR---G--N-RSD----------------------------T-------
------------------------VNS-----------------CIKAS------------------ALWSN--------F-TTLELTSNMRV---T-S----------GD--S----
---------EWINFLLRVGNG---T-E-N------------DEDGRVTLPTE----IMCAGN------------------------IV-TA--VY----G------ENI---------
----------------------------------------------------------------D-ARDT-DNL------------------------------S-T-----------
K-------------------AILAPRNRSVDQLNTEVLS-RM----N---S--------------------------E-----------------------E-R-I-Y----------
----------K---------------------SID------EAVTE-----------------DP--------------------SDAI----------------------HFQPE--
-----F--------------------------------------------------------LHK-----------------------------------LD-----------P-S--
------GMP----------PHEL-----RLRKGAIVMLLRNLD----------V----S--A-----G-----LCNGTRLVVEQF

>Hel_N_ame
-----SLNTHQKR-AADDI------LAA----M-NRS---------------------ES-----RCFFIDGPGGTGKTYLYNTIYNL---A-V--G--Q------------------
-----R-RQVLCVAWTGIAANLLP------GG------RTVTSA-FKL---N--M------AD-G-------N-------RT---S---------------------LM--K--R---
-Q----Q---K-------------EARQL--MA--TEIIIWDEISM--APKCALEA-VECLLRDIM----------------Q---------N-----------------------D-
--------------------------KPFGG-----KLFIIGGDFRQVLPIV-----------------EH---G--Q-RDD----------------------------F-------
------------------------VNS-----------------CVTNS------------------VLWSL--------F-KTHRLQVNMRA---R-E----------AG--L----
---------EWANFLLNLGNG---N-A-N------------DDNGRVQISEE----FRCQRS------------------------IV-TE--IF----G------ETI---------
------------------------------------------------------------------SADD-TDL------------------------------Y-E-----------
R-------------------AILAPTNMSVRQLNNDALQ-RLCTSSP---H--------------------------D-----------------------E-R-V-Y----------
----------K---------------------SID------EALYH-----------------EG--------------------SSDE----------------------LYPME--
-----Y--------------------------------------------------------LNT-----------------------------------LE-----------P-T--
------GMP----------PHEL-----RLKKGAIIMLLRNLD----------V----L--N-----G-----LCNGTRLRIETL
```

```
>Hel_C_pur
-----QLNQDQET-AFKAV------TEA----V-RDDP--------------------ST-----AHFYLQGPGGTGKTFLYETLACH---Y-R--S--E------------------
-----G-KTVICAASTGIAALLLP------GG------RTSHSQ-FML---P--I------DL-H---A---E-------ST---C---------------------NI--A--K---
-Q----S---K-------------TGRLL--AS--ADLIIWDEVPM--QHKYCFEA-VHRLLVDLR----------------G---------T---------------------DED-
--------------------------VLFGG-----VPVILGGDFAQILPVI-----------------RN---G--S-EGQ----------------------------I-------
------------------------VHA-----------------CLRKS------------------FVWPR--------L-KQLALRINMRV---Q-D-SEHG---------N----
---------AFVRWVQSIPY--------------DP-----ALRTMVTLPAY----V-------------------K-QPSTVSE-LI-DH--VY----PA-----DLL----R----
----------------------------------------------------------------NASQDH-ATF------------------------------A-G-----------
R-------------------CLLSTLNTTVTELNNTILD-RMSV--P---AR-------------------------H-----------------------Q-R-T-Y----------
----------A---------------------AVN------TQRTD-----------------PG------------------TAEERY----------------------QLPPE--
-----V--------------------------------------------------------LQS-----------------------------------LE-----------L-P--
------SLP----------PGEL-----RLKIGAPVMLLRNIC----------P----Q--E-----G-----LCNGSRMVVTDL
>Hel_R_del
-----MMNIGQKD-VFDEI------IDS----I-SSNP--------------------NT-----AHFFLQGPAGTGKTFVYNTLCHY---F-R--R--Q------------------
-----G-KIVVCVASSGIASLLLP------GG------RTSHSR-FKI---P--L------NI-Y---P---D-------SV---C---------------------PI--K--K---
-N----S---D-------------LAAML--MQ--CSLIIWDEVPM--QHRHCFEA-VNRTLQDIC----------------S---------N----------------------FG-
--------------------------SLFGG-----IPVVLGGDFAQIGPVV-----------------KN---G--Q-RHH----------------------------I-------
------------------------VEA-----------------SLAKSI-----------------EIWPN--------L-KKLKLTENMRL---S-G-SSPI------D--Q----
---------SFSQWIGSLSYN---S-L---------------LNGKIFLPRY----IA------------------Q--YHSLTT-FV-DS--IY----PK-----EIM----E----
----------------------------------------------------------------Q-TLDP-EFF------------------------------Q-E-----------
R-------------------TIIAPKNDLVDEINRYVLD-QL----P---G--------------------------N-----------------------K-I-S-L----------
----------F---------------------AVD------RVTQE-----------------DS------------------TGSEDR----------------------QMPTE--
-----Y--------------------------------------------------------LQS-----------------------------------LN-----------P-H--
------GLP----------PSVL-----ELKVGMPVMILRNIN----------V----E--K-----G-----LCNGTRVTVLSI
>XP_005716008.1_Chondrus_crispus
-----LLNTNQRS-LSAVA------GPS----M-PTR---------------------SG-----RLFFLDAPGGTGKTFVLSAIQDF---L-R--T--R------------------
-----R-KQVIAVATSAVAAVLLD------GG------RTAHSV-FKI---P--I------PV-S---A---E-------ST---C---------------------SF--S--A---
-N----S---D-------------TGRTL--QQ--VDLIIWDEIVM--CHRHCIET-VDRSLRDLM----------------Q---------T-----------------------D-
--------------------------RPFGG-----NFLVLAGDFRQILPVV-----------------PG---G--S-RGQ----------------------------I-------
------------------------MSA-----------------CVKAS------------------PLYRE--------C-RFLRLTENMRL---A-A-LRADPAADVEA--L----
---------NFPEFLLSVGEG---R-L-Q-----GE-----QRPEWISLPQS----VAFEHT--------------------IRN-LC------------------------------
--------------------------------------------------------------------------------------------------------L-K-----------
R-------------------VILTTKNRPLEEVNEVIGN-MI----P---G--------------------------S-----------------------Y-R-T-Y----------
----------L---------------------SAD------KVENE-----------------DT---------------------NAL----------------------IYPTE--
-----M--------------------------------------------------------LNT-----------------------------------LT-----------AGS--
------ALP----------DHKL-----KLKKGFIVMLLRNLD----------P----A--T-----G-----HVNGARYVIENM
>OSX80228.1_Porphyra_umbilicalis
-----TLQPDQKV-VWDAV------SVS----I-DGS---------------------MG-----RLFFLDAPGGAGKTYLAETLLNY---T-R--G--S------------------
-----G-HIGLAVASSAIAATLMP------LG------RTAHSR-FKI---P--I------EI-N---Q---T-------SF---C---------------------GF--T--Q---
-S----T---D-------------VAKML--KK--TKLIVWDEASM--AHRHCFEA-VDRSIVDVM----------------G---------P-----------------------D-
---------------------------VASQ-----ITWLVCGDFRQVPAVV-----------------PK---G--S-IAQ----------------------------I-------
------------------------IRA-----------------SLRKSP-----------------LMWSC--------F-TRMQLTTNMRV---K-T-RADA-GQAEEA--SL---
-------FEAFGKWLLAIGDG---V-I-RPGSTAAE-----QCTTKIRIPRA----MCLP-K--------------G---------KM-ST--VN----P-------------T----
----------------------------------------------------------------N-AAAL-DAM------------------------------G-E-----------
R-------------------MILTTLNKDVLDLNELAID-QF----P---G--------------------------Q-----------------------A-R-T-Y----------
----------Y---------------------SID------TVSDE-----------------DM--------------------ELAE----------------------VYTTE--
-----F--------------------------------------------------------LNT-----------------------------------ID-----------H-S--
------SVP----------THAM-----VLKVGMTVMLLRNLA----------A----Q--N-----G-----DCNGTRYIVTRL
>KAA6365738.1_Streblomastix_strix
-----QLNDDQKE-IAQLI------IGI----L-NNSYLQ------------------HS-----RLIFVDGPAGTGKTFLYNIINKI---V-N--L--I------------------
-----G-KKILICAWTGIAACLLP------YG------QSSHSL-FKL---P--V------PL-T---S---S-------KN---S--------------------SKI-----I---
-Q----A---K------YEL----LWNQL--QL--VDVILWDEAPM--ASKWAIES-VDKKLKEIR----------------K---------N-----------------------N-
--------------------------KDFGG-----VLMIFGGDFRQVLPIV-----------------KF---G--G-RNE----------------------------Q-------
------------------------VNA-----------------SIQKS------------------NLQKK--------F-DCLKLKKNMRT---G-D----------GS--E----
---------EFSSFLMQIGNG---T-M-Q------Q-----DKNEMIDIPNI---------------------------CMSQEN-LI-KD--VF----G------------------
----------------------------------------------------------------DEILKT-NQQ------------------------------A-N-----------
D-------------------VILCTTNSDSDDINFCVLR-LL----K---C--------------------------E-----------------------P-I-Q-L----------
----------L---------------------SSD------KATLK-----------------NG--------------------ESLD----------------------EITRD--
-----V--------------------------------------------------------LNN-----------------------------------LT-----------L-A--
------ALP----------PHIL-----NIKIGASVMLLRNMN----------I----K--L-----G-----LCNGTRLKVIAI
>Methanothrix_sp
-----SLSEEQTK-AVQHI------TH-------------------------------GK-----DVSILVGRAGSGKSYTLGAIREV---Y-G--A--Q------------------
-----G-YRVRGVALAGIAAEGLQ----NDSG---IHSKTLHRQMWDW----------------------------------------------------------------------
------D---Q-------------GRDLL--SN--NDIIVLDEAAM--VGTRQMHQ-LLTHVKD------------------------------------------------------
-----------------------------AG-----AKIIMVGDQDQAQSIE-----------------AG-----------------------------------------------
--------------------------------------------GIFRA-------------------AKQS--------L-GAVKLTQIRR----Q-K-----------E--V----
---------WQKEATLEFAGDGI-E-V-S---------------------RG------------------------------IE--LY-HQHGHVKDFEKRDAAKEELV----KDWA-
-----------------------------------------------------------------EYNHA-AQ---------------------R---------G-R-----------
S-------------------IILAYTNKDVEDLNAMARE-QR---KLH--G--------------------------E-----------------------------L----------
---------------------------------------------------------------NK--------------------N-----------------------------E--
---VIF----------------------------------------------------------A-----------------------------------TD-----------R----
--------------------GKK-----AFVQGDRILFLKNER---------------S--M-----G-----VRNGTVGTIESI
>D_bacterium
-----SMSNEQEA-AFRYI------AD-------------------------------SG-----DVTCMIGFAGAGKSYTLGAVREA---Y-E--A--A------------------
-----G-YKCKGMALSGIAAEGLE----ISSG---ITSKTIHKAMLDI----------------------------------------------------------------------
------E---H-------------NREQF--TR--KDIIIIDEAGM--VATRQMQK-IISEART------------------------------------------------------
-----------------------------AG-----AKVVLVGDPHQLQPIE-----------------AG-----------------------------------------------
--------------------------------------------GAFRA-------------------ILDR--------V-GYVEISEIRR----Q-K-----------L--D----
```

```
---------WMKEASKDFARN---R-V-S---------------------QA------------------------------LD--AY-NKNGHVKSFDKFDDAKENLI----
QEWVKDR--------------------------------------------------------------LEAKDQ-NK---------------------D---------S-T------
-----A-------------------IILAYRNKDIQDLNSRART-AL---LQA--K--------------------------A-----------------------------L-----
--------------------------------------------------------------------KT--------------------E---------------------------
--G-----VEV----------------------------------------------------------Q-----------------------------------TE-----------
R------------------------GKR-----ILTEGDRILFLRNEK---------------S--L-----G-----VKNGSIGTLEKI

>D_cetonica

-----CLSDEQKN-VLEQI------SK-------------------------------GG-----DLCAVVGHAGTGKSYTLRAVREA---F-E--S--Q------------------
-----G-CTVQGIALAGVAAEGLE----TSSG---IHSTTIHRKLFDW----------------------------------------------------------------------
------D---N-------------GRSRL--DN--KSVLVIDEAGM--VGTRQMDR-ILAEANN------------------------------------------------------
-----------------------------AG-----AKVIVVGDTKQTQAVE-----------------AG-----------------------------------------------
--------------------------------------------GAFRG-------------------ILER--------V-ETSRLSEVWR----Q-K-----------K--D----
---------WQKEATRLLSGD---R-A-SIH-------------------QA------------------------------LD--MY-HKEGYVARYDRYDLASESML----DFYV-
-----------------------------------------------------------------QHYSA-ES---------------------------------T-----------
S-------------------VMIAHRNEDVDRLNTLCRN-DL---RTK--T--------------------------D---------------------------L-L----------
----------------------------------------------------------------G--------------------QK----------------------------E--
---TKV----------------------------------------------------------A-----------------------------------TT-----------T----
--------------------GLK-----AFSNGDRVLFLRNEK---------------S--I-----G-----VKNGTFGTVERF

>Sphingobium_sp

-----SLGDQQQE-ALAHI------TG-------------------------------RD-----DLAIVVGYAGTGKSTMLGVARDE---W-E--R--A------------------
-----G-FNVRGAALSGIAAEGLE----GGSG---IQSRTIASMEYQW----------------------------------------------------------------------
------D---Q-------------GRELL--SP--RDVLVIDEAGM--IGTRQMER-VLSEASQ------------------------------------------------------
-----------------------------AG-----AKVVLVGDAEQLQAIE-----------------AG-----------------------------------------------
--------------------------------------------AAFRS-------------------LAER--------H-GAAEISEVRR----Q-H-----------E--D----
---------WQKDATRALATG---R-T-G-------------------EA------------------------------IH--AY-AEHGMVHAADTREAARAELI----DTWD-
-----------------------------------------------------------------AQRLA-DS---------------------D---------K-T-----------
R-------------------IILTHTNAEVRDLNLAARD-RL---RDA--G--------------------------E-----------------------------L----------
----------------------------------------------------------------G--------------------Q-----------------------------D--
---VAV----------------------------------------------------------S-----------------------------------AE-----------R----
--------------------GAR-----EFATGDRIMFLKNER---------------G--M-----G-----VKNGTLGKVERV

>Sphingomonas_sp

-----ILSGEQRD-AFDHV------TG-------------------------------NA-----GLASVVGYAGSGKSAMLGVAREA---W-E--G--Q------------------
-----G-YTVRGAALSGIAAENLE----GGSG---IASRTIASLEHAW----------------------------------------------------------------------
------G---Q-------------GREQL--GP--RDVLVVDEAGM--IGSRQMER-VLSQARD------------------------------------------------------
-----------------------------AG-----AKVVMVGDPEQLQAIE-----------------AG-----------------------------------------------
--------------------------------------------AAFRS-------------------ITER--------H-GAAEITEIRR----Q-R-----------E--D----
---------WQKEATRQLATG---R-T-G-------------------EA------------------------------IR--AY-DAHGMVHGYATREEARAGLV----DDWD-
-----------------------------------------------------------------GKRQA-EP---------------------G---------R-S-----------
Q-------------------IIFTHTNAEVRELNGEARE-RM---RAT--E--------------------------D-----------------------------L----------
----------------------------------------------------------------G--------------------D-----------------------------D--
---VAV----------------------------------------------------------K-----------------------------------AD-----------R----
--------------------GER-----AFAAGDRLMFLRNER---------------S--L-----G-----VKNGTLGTIEGV

>S_macrogoltabida

-----VLSGEQRD-AFDRI------TE-------------------------------GQ-----GLTSVIGYAGTGKSAMLGVAREA---W-E--R--E------------------
-----G-YQVRGAALSGIAAENLE----GGSG---IQSRTIASLEHAW----------------------------------------------------------------------
------A---Q-------------GRDQL--SR--NDVLVVDEAGM--IGTRQMER-VLSHARD------------------------------------------------------
-----------------------------AG-----AKVVLVGDPEQMQAIE-----------------AG-----------------------------------------------
--------------------------------------------AAFRS-------------------ISER--------H-GAAEITEVRR----Q-R-----------G--D----
---------WQKEATRSLATG---R-T-G-------------------EA------------------------------LH--AY-ESRGMVQAADTREAARGELV----DGWD-
-----------------------------------------------------------------RQRQA-EP---------------------D---------K-T-----------
R-------------------IILTHTNAEVRALNEEARG-RM---RAG--G--------------------------E-----------------------------L----------
----------------------------------------------------------------G--------------------Q-----------------------------D--
---VGV----------------------------------------------------------T-----------------------------------VE-----------R----
--------------------GRR-----DFASGDRIMFLRNER---------------S--M-----G-----VKNGTLGTLEHV

>A_tumefaciens

-----KLSGEQAE-ALVHV------TD-------------------------------GR-----DLGIVVGYAGTGKSAMLGVAREV---W-E--A--E------------------
-----G-YAVRGVALSGIAAENLE----SGSG---ISSRTIASMEHGW----------------------------------------------------------------------
------K---Q-------------GRDTL--TS--RDVLVIDEAGM--VGTRQMER-VLSHAQE------------------------------------------------------
-----------------------------VG-----AKVVLAGDPQQLQSIE-----------------AG-----------------------------------------------
--------------------------------------------AAFRS-------------------IHER--------H-GGVEISEVRR----Q-R-----------E--D----
---------WQRDATRDLATG---N-M-G-------------------EA------------------------------IH--AY-ERNDMVHAAETREQARNDLI----EGWD-
-----------------------------------------------------------------RQRQE-NP---------------------D---------A-S-----------
R-------------------IILTHTNVEVRELNEAARA-KV---RDA--G--------------------------N-----------------------------L----------
----------------------------------------------------------------G--------------------E-----------------------------D--
---VCI----------------------------------------------------------T-----------------------------------VE-----------
TRDGQ-------------------AGER-----SFAAGDRVMFMANER---------------G--LGGDGGG-----VKNGTLGTIEEV

>Mesorhizobium_sp

-----VLSGEQAD-ALDHI------TD-------------------------------GH-----GLGVVVGFAGTGKSAMLGVARQA---W-A--A--A------------------
-----G-YEVKGAALSGITAENLE----SGSG---IASRTVASLEHGW----------------------------------------------------------------------
------G---Q-------------GGDLL--TA--RDVLVIDEAGM--VGTRQLER-VLSHAAE------------------------------------------------------
-----------------------------VG-----AKIVLVGDPQQLQAIE-----------------AG-----------------------------------------------
--------------------------------------------AAFRS-------------------IHER--------H-GGVEIGQVRR----Q-R-----------E--D----
---------WQRDATRNLATG---R-I-G-------------------AA------------------------------ID--AY-EAKGMVHQAATRDQARGDLV----ERWD-
-----------------------------------------------------------------RDRRA-DP---------------------E---------A-S-----------
R-------------------IILTHTNDEVRALNKAARE-RM---HAA--G--------------------------D-----------------------------L----------
----------------------------------------------------------------G--------------------D-----------------------------D--
---VQV----------------------------------------------------------R-----------------------------------VD-----------R----
--------------------GAR-----SFATGDRIMFLRNER---------------G--L-----G-----VKNGTLGIVEEV
```

```
>Phenylobacterium_sp
-----ALGGEQRD-ALEHI------TG-------------------------------GQ-----DLSMVVGYAGSGKSAMLGVAREA---W-E--A--Q------------------
-----G-YQVRGAALSGIAAESLE----AGSS---IPSRTIASLEHSW----------------------------------------------------------------------
------G---Q-------------GRDLL--TS--SDVLVIDEAGM--IGSRQMDR-VLLAAER------------------------------------------------------
-----------------------------AG-----AKVVLVGDAEQLQAIE-----------------AG-----------------------------------------------
--------------------------------------------ASFRA-------------------LTER--------H-GAAEITEIRR----Q-R-----------E--S----
---------WQREATRELATG---R-T-G---------------------AA------------------------------LE--RY-DAAGMVRAHETREAAREALV----DGWE-
-----------------------------------------------------------------AVRRE-AP---------------------G---------A-S-----------
Q-------------------IMLAHTRADVAELNHLARV-RM---RDA--G--------------------------E-----------------------------L----------
----------------------------------------------------------------G--------------------E-----------------------------D--
---LAL----------------------------------------------------------A-----------------------------------TE-----------R----
--------------------GER-----TFAAGDRIMFLRNER---------------S--L-----G-----VKNGTLGTVERI
>A_excentricus
-----GMSDEQKD-AVRHI------TG-------------------------------DA-----QIAVVVGFAGAGKSTLLSAAKEA---W-E--A--Q------------------
-----G-YTVHGAALAGKAVGGLE----ESAG---IEGRTLASWDTRW----------------------------------------------------------------------
------K---M-------------GTSEL--GP--GDVLVIDEAGM--IGSRQMDR-FVSEAER------------------------------------------------------
-----------------------------TG-----AKLVLVGDHEQLQAIG-----------------AG-----------------------------------------------
--------------------------------------------APFRA-------------------IAER--------V-GHASVEDIRR----Q-R-----------S--D----
---------WQRDASKAFATQ---R-T-A---------------------QG------------------------------LA--AY-IEHGHVHLKADQSEATTALV----RDYV-
-----------------------------------------------------------------KDVEA-RP---------------------D---------G-S-----------
R-------------------AAMAHRRVDVRELNNGIRE-EL---KAR--G--------------------------H-----------------------------L----------
---------------------------------------------------------------KG--------------------E-----------------------------D--
---VPF----------------------------------------------------------N-----------------------------------TD-----------D----
--------------------GQR-----NFTEGDRLVFLQNDR---------------E--M-----G-----VKNGTLGTVEGI
```

9.3 Material suplementar do Capítulo 3

Supplementary Material

Table S1

Rank				Species	# of sequences	Geographical location	Submission Institution	Submission Date
Class Insecta								
	Order Lepidoptera							
		Superfamily Papilionoidea						
			Family Nymphalidae					
				Pararge aegeria	31	Scotland/UK	Wellcome Sanger Institute/UK, Stockholm University/SWE	2021-01-28, 2018-08-08
				Fabriciana adippe	1	Romania	Wellcome Sanger Institute/UK	2021-04-15
				Heliconius wallacei	1	Peru	University of Cambridge/UK	2015-11-29
				Vanessa cardui	5	Scotland/UK	Wellcome Sanger Institute/UK	2021-02-13
				Dryas iulia	7	Costa Rica	Cornell University/USA	2021-06-28
				Danaus melanippus	1	India	Iridian Genomes/USA	2020-01-30
				Nymphalis polychloros	1	Spain	Wellcome Sanger Institute/UK	2021-02-13
			Family Riodinidae					
				Apodemia ares	1	USA	Florida Museum of Natural History/USA	2021-05-03
				Emesis lacrines	1	Costa Rica	Florida Museum of Natural History/USA	2021-05-03
				Emesis aurimna	1	Costa Rica	Florida Museum of Natural History/USA	2021-05-03
				Emesis ocypore	1	Peru	Florida Museum of Natural History/USA	2021-05-03
				Emesis heterochroa	1	Peru	Florida Museum of Natural History/USA	2021-05-03
			Family Papilionidae					
				Parnassius apollo	5	Germany, Italy	Florida Museum of Natural History/USA, Stockholm University/SWE	2021-05-03, 2021-06-20
				Parnassius imperator	1	China	Florida Museum of Natural History/USA	2021-05-03
				Parnassius smintheus	1	Canada	Florida Museum of Natural History/USA	2021-05-03
				Zerynthia polyxena	1	Italy	Florida Museum of Natural History/USA	2021-05-03
				Archon apollinus	1	Greece	Florida Museum of Natural History/USA	2021-05-03
				Protesilaus protesilaus	2	Peru	Florida Museum of Natural History/USA	2021-05-03
			Family Lycaenidae					
				Curetis bulis	1	Myanmar	Florida Museum of Natural History/USA	2021-05-03
				Cyaniris semiargus	3	Romania	Wellcome Sanger Institute/UK	2021-01-25
				Lysandra coridon	2	Romania	Wellcome Sanger Institute/UK	2021-02-13
				Lycaena phlaeas	1	Scotland/UK	Wellcome Sanger Institute/UK	2021-03-17
				Aricia agestis	1	Romania	Wellcome Sanger Institute/UK	2021-01-25
				Lysandra bellargus	1	Spain	Wellcome Sanger Institute/UK	2021-03-17
				Lepidochrysops patricia	1	South Africa	Florida Museum of Natural History/USA	2021-05-03
				Eumaeus atala	7	USA	University of Texas Southwestern/USA	2021-03-02
			Family Pieridae					
				Pieris rapae	3	Scotland/UK	Wellcome Sanger Institute/UK	2021-01-25
			Family Hesperiidae					
				Pyrgus malvae	3	Romania	Wellcome Sanger Institute/UK	2021-07-21
				Satarupa nymphalis	1	China	Florida Museum of Natural History/USA	2021-05-03
				Gindanes brontinus	1	Costa Rica	Florida Museum of Natural History/USA	2021-05-03
				Pyrrhopyge telassa	1	Peru	Florida Museum of Natural History/USA	2021-05-03
				Pyrrhopyge sergius	1	Peru	Florida Museum of Natural History/USA	2021-05-03
				Pyrrhopyge hadassa	1	Peru	Florida Museum of Natural History/USA	2021-05-03
				Pyrrhopyge kelita	1	Peru	Florida Museum of Natural History/USA	2021-05-03
				Pyrrhopyge crida	1	Costa Rica	Florida Museum of Natural History/USA	2021-05-03
				Pyrrhopyge pelota	1	Bolivia	Florida Museum of Natural History/USA	2021-05-03
				Celaenorrhinus cf. opalinus	1	Kenya	Florida Museum of Natural History/USA	2021-05-03
				Katreus holocausta	1	Cameroon	Florida Museum of Natural History/USA	2021-05-03
				Morvina fissimacula	1	Costa Rica	Florida Museum of Natural History/USA	2021-05-03
				Ouleus salvina	1	Costa Rica	Florida Museum of Natural History/USA	2021-05-03
				Cecropterus casica	1	USA	Florida Museum of Natural History/USA	2021-05-03
				Mylon lassia	1	Costa Rica	Florida Museum of Natural History/USA	2021-05-03
				Eburuncus unifasciata	1	Panama	Florida Museum of Natural History/USA	2021-05-03
				Oxynetra roscius	1	Brazil	Florida Museum of Natural History/USA	2021-05-03
				Duroca duroca	1	Brazil	Florida Museum of Natural History/USA	2021-05-03
				Charidia lucaria	1	Peru	Florida Museum of Natural History/USA	2021-05-03
				Aurina azines	1	Guyana	Florida Museum of Natural History/USA	2021-05-03
				Mimia cf. chiapaensis	1	Ecuador	Florida Museum of Natural History/USA	2021-05-03
				Pythonides amaryllis	1	Costa Rica	Florida Museum of Natural History/USA	2021-05-03
				Zopyrion sandace	1	Mexico	Florida Museum of Natural History/USA	2021-05-03
				Mimoniades ocyalus	1	Brazil	Florida Museum of Natural History/USA	2021-05-03
				Dalla cyprius	1	Peru	Florida Museum of Natural History/USA	2021-05-03
				Signeta flammeata	1	Australia	Florida Museum of Natural History/USA	2021-05-03
				Erynnis tages	1	Romania	Wellcome Sanger Institute/UK	2021-01-25
				Ectomis octomaculata	1	Costa Rica	Florida Museum of Natural History/USA	2021-05-03
				Cecropterus confusis	1	USA	Florida Museum of Natural History/USA	2021-05-03
				Thymelicus sylvestris	6	England/UK	Wellcome Sanger Institute/UK	2021-07-21
				Piruna pirus	1	USA	Florida Museum of Natural History/USA	2021-05-03
				Timochares trifasciata	1	Costa Rica	Florida Museum of Natural History/USA	2021-05-03
				Autochton oryx	1	Ecuador	Florida Museum of Natural History/USA	2021-05-03
		Superfamily Geometroidea						
			Family Geometridae					
				Campaea margaritaria	2	England/UK	Wellcome Sanger Institute/UK	2021-08-18
				Hydriomena furcata	4	England/UK	Wellcome Sanger Institute/UK	2021-08-18
				Ectropis grisescens	9	China	Institute of Plant Physiology and Ecology/CHN	2021-03-22
		Superfamily Noctuoidea						
			Family Noctuidae					
				Amphipyra tragopoginis	2	England/UK	Wellcome Sanger Institute/UK	2021-02-13
				Griposia aprilina	1	England/UK	Wellcome Sanger Institute/UK	2021-09-30
				Atethmia centrago	4	England/UK	Wellcome Sanger Institute/UK	2021-03-17
				Mythimna ferrago	1	England/UK	Wellcome Sanger Institute/UK	2021-07-06
				Autographa pulchrina	1	England/UK	Wellcome Sanger Institute/UK	2021-04-14
				Autographa gamma	1	England/UK	Wellcome Sanger Institute/UK	2021-01-25
				Trichoplusia ni	1	USA	Cornell University/USA	2018-10-01
				Mamestra brassicae	2	Wales/UK	Wellcome Sanger Institute/UK	2021-01-25
				Sesamia nonagrioides	1	France	Paris-Saclay University/FRA	2021-04-13
			Family Notodontidae					
				Clostera curtula	3	England/UK	Wellcome Sanger Institute/UK	2021-04-14
				Ptilodon capucinus	4	England/UK	Wellcome Sanger Institute/UK	2021-09-11
			Family Erebidae					
				Eilema sororculum	1	England/UK	Wellcome Sanger Institute/UK	2021-09-24

			Spilosoma lubricipeda	3	England/UK	Wellcome Sanger Institute/UK	2021-02-13
			Euproctis similis	3	England/UK	Wellcome Sanger Institute/UK	2021-01-25
			Spilarctia lutea	11	England/UK	Wellcome Sanger Institute/UK	2021-09-18
			Schrankia costaestrigalis	1	England/UK	Wellcome Sanger Institute/UK	2021-04-14
			Arctia plantaginis	3	Finland?	University of Cambridge/UK	2020-04-10
			Lymantria monacha	6	England/UK	Wellcome Sanger Institute/UK	2021-01-25
			Lymantria dispar	5	Japan, China	Laval University/CAN	2021-05-04
	Superfamily Bombycoidea						
		Family Bombycidae					
			Bombyx mori	3	Japan	The University of Tokyo/JPN	2020-11-06
		Family Sphingidae					
			Laothoe populi	7	England/UK	Wellcome Sanger Institute/UK	2021-02-13
			Hyles vespertilio	1	Italy	Max Planck Institute of Molecular Cell Biology and Genetics/DEU	2020-01-29
		Family Saturniidae					
			Samia ricini	7	India*	Gakushuin University/JPN	2020-06-20
	Superfamily Pyraloidea						
		Family Crambidae					
			Chilo suppressalis	1	China	Huazhong Agricultural University/CHN	2019-01-08
			Chrysoteuchia culmella	2	England/UK	Wellcome Sanger Institute/UK	2021-07-06
	Superfamily Gelechioidea						
		Family Blastobasidae					
			Blastobasis lacticolella	15	England/UK	Wellcome Sanger Institute/UK	2021-01-25
			Blastobasis adustella	4	England/UK	Wellcome Sanger Institute/UK	2021-05-19
	Superfamily Drepanoidea						
		Family Drepanidae					
			Habrosyne pyritoides	1	England/UK	Wellcome Sanger Institute/UK	2021-05-11
	Superfamily Tortricoidea						
		Family Tortricidae					
			Apotomis turbidana	1	England/UK	Wellcome Sanger Institute/UK	2021-01-25
Order Diptera							
	Superfamily Diopsoidea						
		Family Diopsidae					
			Teleopsis dalmanni	4	Malaysia	SUNY Geneseo/USA, University of Maryland/USA	2020-09-23, 2020-10-30
	Superfamily Syrphoidea						
		Family Syrphidae					
			Cheilosia vulpina	1	England/UK	Wellcome Sanger Institute/UK	2021-09-30
			Melanostoma mellinum	3	England/UK	Wellcome Sanger Institute/UK	2021-09-11
	Superfamily Tephritoidea						
		Family Tephritidae					
			Bactrocera dorsalis	1	USA	Agricultural Research Service-USDA/USA	2014-12-03
	Superfamily Ephydroidea						
		Family Drosophilidae					
			Drosophila biarmipes	7	India to SE Asia*	University of Pennsylvania/USA	2019-05-08
			Drosophila ficusphila	1	Taiwan	Stanford University/USA	2021-04-28
			Drosophila auraria	1	Japan	University of California, Berkeley/USA	2019-08-21
			Drosophila bifasciata	1	Japan	University of California, Berkeley/USA	2019-11-15
			Drosophila obscura	3	Europe*, Serbia	National Institute of Genetics/JPN, Stanford University/USA	2017-10-14, 2021-04-28
			Drosophila ambigua	1	Serbia	Stanford University/USA	2021-04-28
			Drosophila guanche	1	Canary Islands/ESP	Centro Nacional de Análisis Genómico/ESP	2018-09-20
			Scaptomyza montana	2	USA*	Stanford University/USA	2021-06-16
			Scaptomyza flava	1	USA	University of California, Berkeley/USA	2018-12-17
	Superfamily Oestroidea						
		Family Tachinidae					
			Tachina fera	1	England/UK	Wellcome Sanger Institute/UK	2021-02-13
Order Orthoptera							
	Superfamily Grylloidea						
		Family Gryllidae					
			Teleogryllus occipitalis	4	Japan	Waseda university/JPN	2020-02-22
			Gryllus bimaculatus	1	Japan	Tokushima University/JPN	2021-02-13
	Superfamily Eumastacoidea						
		Family Morabidae					
			Vandiemenella viatica	1	Australia	Uppsala University/SWE	2021-08-07
Order Hymenoptera							
	Superfamily Ichneumonoidea						
		Family Braconidae					
			Cotesia vestalis	1	South Korea	Andong National University/KOR	2015-03-18
			Cotesia vestalis bracovirus segment c35	1	China	Zhejiang University/CHN	2011-05-09
		Family Ichneumonidae					
			Mesochorus sp.	1	Costa Rica	University of Georgia/USA	2021-06-16
Order Coleoptera							
	Superfamily Tenebrionoidea						
		Family Pyrochroidae					
			Pyrochroa serraticornis	5	England/UK	Wellcome Sanger Institute/UK	2021-03-17
Order Neuroptera							
		Family Chrysopidae					
			Chrysoperla carnea	1	England/UK	Wellcome Sanger Institute/UK	2021-04-14
Order Siphonaptera							
	Superfamily Pulicoidea						
		Family Pulicidae					
			Ctenocephalides felis	5	USA	West Virginia University/USA	2018-08-24
Order Phasmatodea							
		Family Phasmatidae					
			Clitarchus hookeri	1	New Zealand	Landcare Research/NZL	2017-11-16

Class Arachnida								
	Order Araneae							
		Superfamily Araneoidea						
			Family Nephilidae					
				Trichonephila inaurata madagascariensis	1	Madagascar	Institute for Advanced Biosciences - Keio University/JPN	2021-07-22
			Family Linyphiidae					
				Oedothorax gibbosus	1	Belgium	Royal Belgian Institute of Natural Sciences/BEL	2021-07-22

*Original or known distribution of the species (geographical location of biosample not available).

Table S2. Average base differences per site between groups in the main clade containing CvBV Hel_c35.

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36	37	38	39	40
1.*Cotesia vestalis* bracovirus																																								
2.*Pararge aegeria*	0.0002																																							
3.*Pyrgus malvae*	0.0001	0.0003																																						
4.*Cotesia vestalis*	0.0002	0.0004	0.0003																																					
5.*Campaea margaritaria*	0.0011	0.0013	0.0012	0.0012																																				
6.*Hydriomena furcata*	0.0018	0.0020	0.0019	0.0021	0.0014																																			
7.*Amphipyra tragopoginis*	0.0022	0.0024	0.0023	0.0024	0.0019	0.0029																																		
8.*Pyrochroa serraticornis*	0.0043	0.0046	0.0044	0.0043	0.0044	0.0051	0.0054																																	
9.*Bombyx mori*	0.0061	0.0063	0.0062	0.0063	0.0063	0.0068	0.0072	0.0066																																
10.*Satarupa nymphalis*	0.0034	0.0036	0.0035	0.0035	0.0034	0.0041	0.0045	0.0040	0.0031																															
11.*Laothoe populi*	0.0063	0.0066	0.0065	0.0064	0.0063	0.0071	0.0074	0.0070	0.0061	0.0034																														
12.*Habrosyne pyritoides*	0.0061	0.0063	0.0061	0.0062	0.0063	0.0068	0.0071	0.0067	0.0057	0.0031	0.0061																													
13.*Fabriciana adippe*	0.0051	0.0053	0.0052	0.0050	0.0050	0.0058	0.0061	0.0058	0.0073	0.0048	0.0077	0.0074																												
14.*Clostera curtula*	0.0059	0.0061	0.0060	0.0058	0.0056	0.0066	0.0070	0.0065	0.0058	0.0028	0.0053	0.0058	0.0073																											
15.*Apotomis turbidana*	0.0072	0.0074	0.0073	0.0073	0.0067	0.0080	0.0079	0.0078	0.0096	0.0066	0.0093	0.0098	0.0090	0.0086																										
16.*Gindanes brontinus*	0.0146	0.0146	0.0145	0.0148	0.0142	0.0150	0.0153	0.0132	0.0159	0.0128	0.0154	0.0159	0.0157	0.0153	0.0143																									
17.*Apodemia ares*	0.0116	0.0116	0.0115	0.0113	0.0109	0.0123	0.0123	0.0112	0.0132	0.0104	0.0127	0.0130	0.0130	0.0122	0.0126	0.0090																								
18.*Pyrrhopyge telassa*	0.0134	0.0134	0.0133	0.0131	0.0126	0.0141	0.0142	0.0138	0.0149	0.0119	0.0149	0.0148	0.0148	0.0150	0.0170	0.0091	0.0096																							
19.*Pyrrhopyge sergius*	0.0160	0.0160	0.0159	0.0158	0.0157	0.0167	0.0167	0.0165	0.0176	0.0147	0.0177	0.0174	0.0174	0.0175	0.0197	0.0118	0.0120	0.0091																						
20.*Pyrrhopyge hadassa*	0.0166	0.0166	0.0165	0.0165	0.0161	0.0173	0.0173	0.0170	0.0182	0.0153	0.0182	0.0180	0.0180	0.0182	0.0199	0.0121	0.0128	0.0091	0.0071																					
21.*Pyrrhopyge kelita*	0.0173	0.0173	0.0172	0.0172	0.0168	0.0179	0.0182	0.0177	0.0189	0.0162	0.0195	0.0187	0.0185	0.0188	0.0201	0.0120	0.0133	0.0118	0.0081	0.0016																				
22.*Pyrrhopyge crida*	0.0160	0.0160	0.0159	0.0158	0.0153	0.0166	0.0167	0.0164	0.0178	0.0143	0.0176	0.0176	0.0174	0.0178	0.0193	0.0110	0.0124	0.0090	0.0074	0.0065	0.0082																			
23.*Pyrrhopyge pelota*	0.0201	0.0202	0.0200	0.0199	0.0204	0.0208	0.0211	0.0208	0.0215	0.0187	0.0217	0.0214	0.0216	0.0215	0.0241	0.0166	0.0171	0.0143	0.0123	0.0129	0.0140	0.0130																		
24.*Emesis lacrines*	0.0122	0.0122	0.0121	0.0121	0.0120	0.0127	0.0130	0.0110	0.0138	0.0111	0.0135	0.0139	0.0134	0.0133	0.0131	0.0121	0.0106	0.0120	0.0146	0.0152	0.0160	0.0147	0.0196																	
25.*Eilema sororculum*	0.0148	0.0148	0.0147	0.0150	0.0143	0.0153	0.0156	0.0142	0.0168	0.0138	0.0164	0.0169	0.0159	0.0162	0.0160	0.0175	0.0158	0.0170	0.0197	0.0201	0.0199	0.0195	0.0239	0.0156																
26.*Celaenorrhinus cf opalinus*	0.0116	0.0116	0.0115	0.0115	0.0108	0.0121	0.0121	0.0123	0.0136	0.0107	0.0135	0.0135	0.0131	0.0131	0.0140	0.0139	0.0121	0.0156	0.0177	0.0183	0.0195	0.0181	0.0229	0.0124	0.0114															
27.*Katreus holocausta*	0.0127	0.0129	0.0128	0.0127	0.0121	0.0136	0.0138	0.0128	0.0148	0.0115	0.0142	0.0149	0.0143	0.0137	0.0146	0.0152	0.0142	0.0171	0.0197	0.0201	0.0205	0.0191	0.0238	0.0147	0.0135	0.0108														
28.*Teleopsis dalmanni*	0.0262	0.0262	0.0261	0.0262	0.0260	0.0267	0.0268	0.0223	0.0280	0.0249	0.0265	0.0284	0.0275	0.0267	0.0209	0.0296	0.0270	0.0223	0.0246	0.0246	0.0235	0.0243	0.0274	0.0281	0.0236	0.0157	0.0253													
29.*Morvina fissimacula*	0.0209	0.0209	0.0208	0.0209	0.0207	0.0214	0.0219	0.0215	0.0228	0.0194	0.0228	0.0229	0.0225	0.0225	0.0243	0.0127	0.0168	0.0176	0.0198	0.0205	0.0221	0.0196	0.0256	0.0198	0.0234	0.0227	0.0239	0.0278												
30.*Ouleus salvina*	0.0164	0.0166	0.0165	0.0165	0.0166	0.0173	0.0175	0.0176	0.0183	0.0154	0.0185	0.0187	0.0183	0.0185	0.0203	0.0154	0.0139	0.0170	0.0193	0.0197	0.0192	0.0189	0.0235	0.0154	0.0207	0.0194	0.0208	0.0256	0.0247											
31.*Emesis aurimna*	0.0194	0.0194	0.0193	0.0197	0.0194	0.0199	0.0201	0.0189	0.0210	0.0178	0.0208	0.0207	0.0201	0.0202	0.0211	0.0192	0.0177	0.0203	0.0232	0.0237	0.0250	0.0232	0.0275	0.0140	0.0227	0.0209	0.0223	0.0361	0.0277	0.0238										
32.*Hyles vespertilio*	0.0156	0.0155	0.0154	0.0156	0.0151	0.0161	0.0165	0.0160	0.0173	0.0146	0.0177	0.0177	0.0167	0.0172	0.0181	0.0167	0.0154	0.0172	0.0205	0.0207	0.0224	0.0197	0.0257	0.0158	0.0146	0.0144	0.0154	0.0181	0.0266	0.0213	0.0246									
33.*Emesis ocypore*	0.0178	0.0178	0.0177	0.0179	0.0181	0.0183	0.0186	0.0182	0.0190	0.0161	0.0195	0.0194	0.0189	0.0187	0.0214	0.0185	0.0173	0.0194	0.0220	0.0224	0.0244	0.0217	0.0273	0.0176	0.0215	0.0211	0.0221	0.0257	0.0279	0.0236	0.0255	0.0242								
34.*Chrysoperla carnea*	0.0211	0.0213	0.0212	0.0211	0.0215	0.0218	0.0222	0.0216	0.0233	0.0205	0.0234	0.0238	0.0223	0.0230	0.0247	0.0238	0.0226	0.0245	0.0269	0.0276	0.0277	0.0272	0.0319	0.0216	0.0196	0.0203	0.0206	0.0247	0.0321	0.0283	0.0303	0.0229	0.0293							
35.*Tachina fera*	0.0221	0.0221	0.0220	0.0217	0.0215	0.0228	0.0230	0.0225	0.0235	0.0208	0.0233	0.0239	0.0235	0.0228	0.0251	0.0243	0.0230	0.0256	0.0285	0.0293	0.0309	0.0290	0.0335	0.0229	0.0210	0.0209	0.0215	0.0250	0.0338	0.0295	0.0314	0.0267	0.0318	0.0304						
36.*Lepidochrysops patricia*	0.0182	0.0184	0.0183	0.0180	0.0184	0.0189	0.0185	0.0187	0.0205	0.0177	0.0205	0.0210	0.0203	0.0198	0.0212	0.0210	0.0192	0.0207	0.0245	0.0248	0.0266	0.0236	0.0312	0.0196	0.0172	0.0174	0.0172	0.0226	0.0312	0.0265	0.0286	0.0225	0.0296	0.0288	0.0287					
37.*Schrankia costaestrigalis*	0.0268	0.0268	0.0267	0.0267	0.0259	0.0273	0.0276	0.0272	0.0278	0.0252	0.0283	0.0280	0.0275	0.0279	0.0294	0.0277	0.0268	0.0291	0.0315	0.0322	0.0338	0.0305	0.0368	0.0271	0.0294	0.0292	0.0298	0.0325	0.0373	0.0335	0.0354	0.0325	0.0309	0.0366	0.0403	0.0363				
38.*Drosophila ficusphila*	0.0416	0.0418	0.0417	0.0416	0.0416	0.0420	0.0425	0.0427	0.0438	0.0412	0.0441	0.0431	0.0434	0.0438	0.0448	0.0426	0.0419	0.0448	0.0468	0.0453	0.0470	0.0473	0.0525	0.0426	0.0448	0.0437	0.0446	0.0480	0.0539	0.0474	0.0501	0.0469	0.0446	0.0520	0.0564	0.0503	0.0503			
39.*Danaus melanippus*	0.0271	0.0273	0.0272	0.0272	0.0270	0.0279	0.0275	0.0265	0.0289	0.0265	0.0286	0.0293	0.0285	0.0281	0.0279	0.0297	0.0275	0.0299	0.0321	0.0323	0.0320	0.0316	0.0360	0.0282	0.0246	0.0243	0.0248	0.0386	0.0362	0.0328	0.0349	0.0289	0.0331	0.0343	0.0340	0.0318	0.0410	0.0557		
40.*Ctenocephalides felis*	0.0507	0.0507	0.0506	0.0508	0.0483	0.0510	0.0517	0.0508	0.0524	0.0499	0.0526	0.0530	0.0524	0.0519	0.0528	0.0530	0.0501	0.0534	0.0563	0.0561	0.0559	0.0562	0.0610	0.0507	0.0531	0.0531	0.0530	0.0566	0.0602	0.0567	0.0596	0.0544	0.0554	0.0605	0.0626	0.0552	0.0634	0.0751	0.0620	
41.*Gryllus bimaculatus*	0.0337	0.0337	0.0336	0.0319	0.0325	0.0332	0.0341	0.0324	0.0355	0.0329	0.0343	0.0355	0.0353	0.0350	0.0317	0.0370	0.0357	0.0347	0.0374	0.0377	0.0370	0.0371	0.0417	0.0317	0.0340	0.0300	0.0314	0.0472	0.0408	0.0386	0.0419	0.0322	0.0390	0.0379	0.0399	0.0350	0.0465	0.0623	0.0465	0.0676

Figure S1. Same Maximum Likelihood phylogeny as Fig. 1 (main text), displaying taxa names and branch support values. Distinct Lepidoptera superfamilies are represented by different colors and non-lepidopteran arthropods are represented in black. See Materials and Methods for details of the phylogenetic inference procedures.

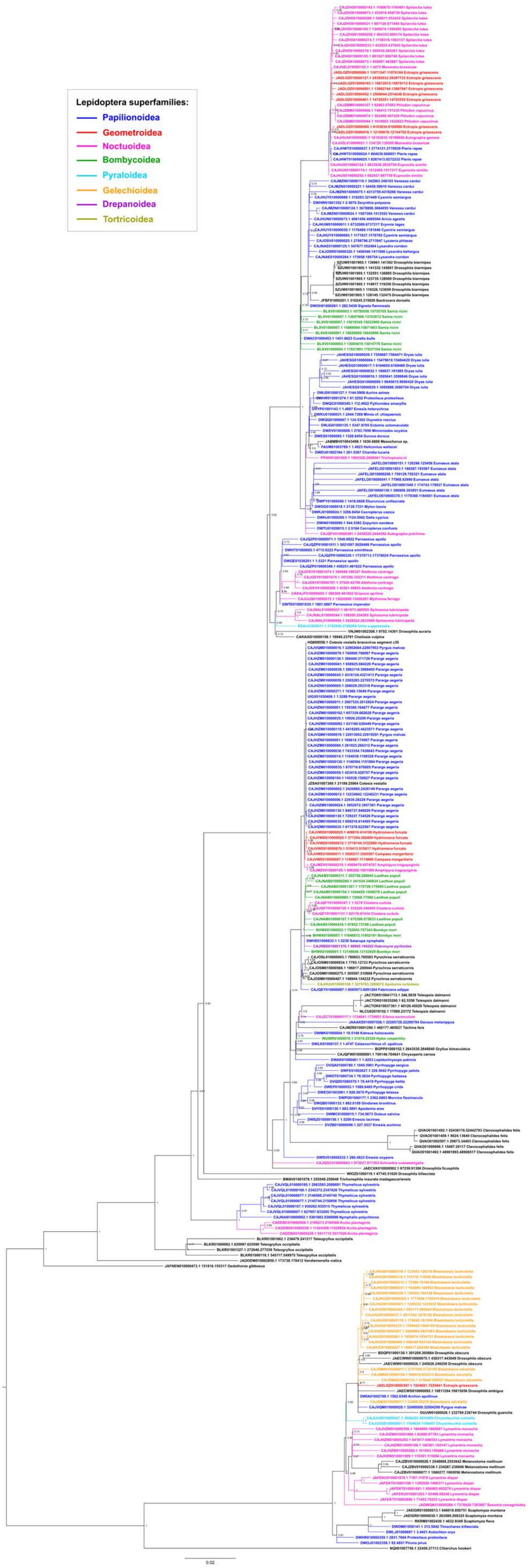
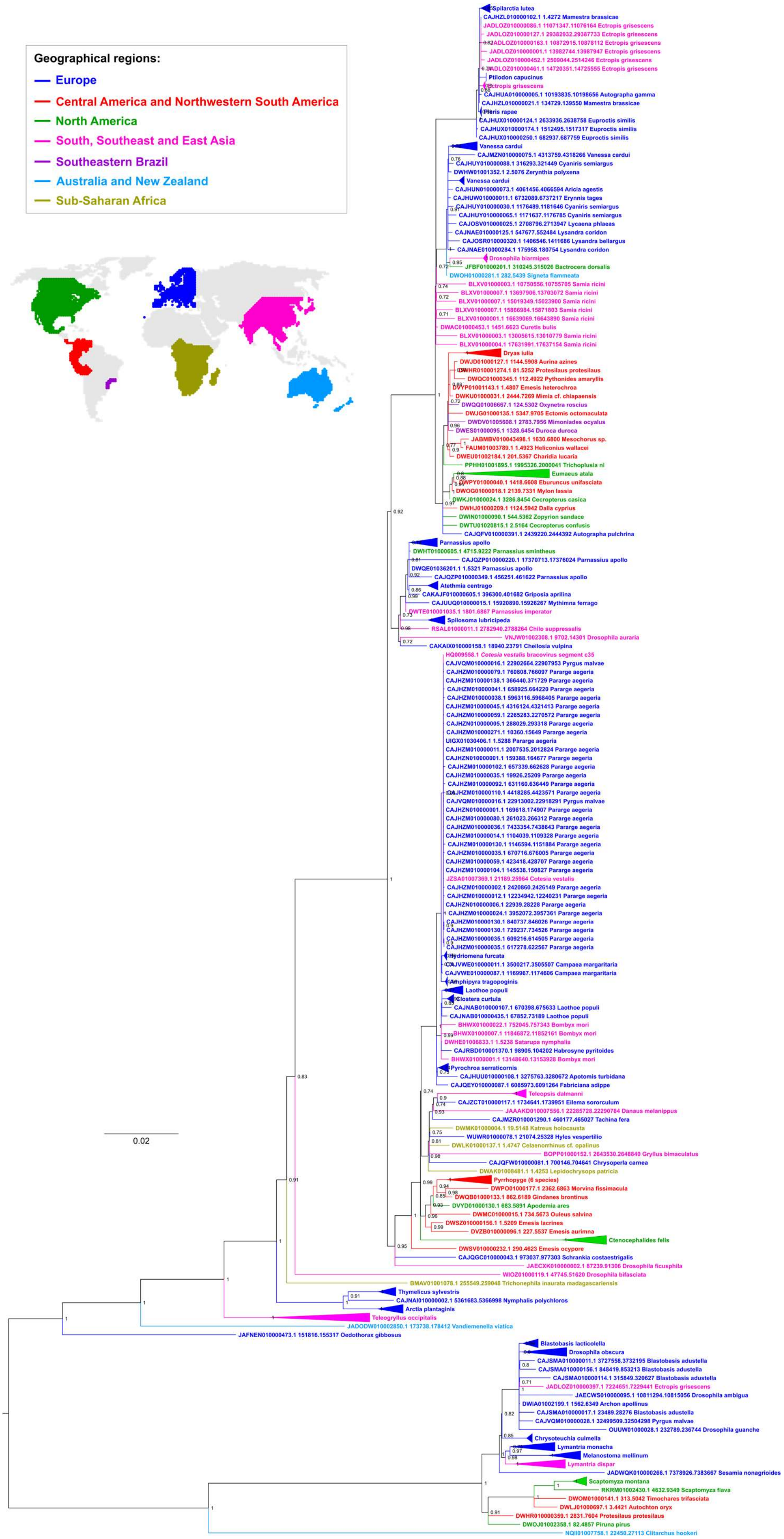

Lepidoptera superfamilies:
Papilionoidea
Geometroidea
Noctuoidea
Bombycoidea
Pyraloidea
Gelechioidea
Drepanoidea
Tortricoidea
0.02

Figure S2. Same Maximum Likelihood phylogeny as Fig. 2 (main text), displaying taxa names and branch support values. Colors correspond to geographical locations where the species were sampled (Table S1).


Geographical regions:
Europe
Central America and Northwestern South America
North America
South, Southeast and East Asia
Southeastern Brazil
Australia and New Zealand
Sub-Saharan Africa
0.02
Spilarctia lutea
CAJHZL010000102.1 1.4272 Mamestra brassicae
JADLOZ010000086.1 11071347.11076164 Ectropis grisescens
JADLOZ010000127.1 29382932.29387733 Ectropis grisescens
JADLOZ010000163.1 10872915.10878112 Ectropis grisescens
JADLOZ010000001.1 13982744.13987947 Ectropis grisescens
JADLOZ010000452.1 2509044.2514246 Ectropis grisescens
JADLOZ010000461.1 14720351.14725555 Ectropis grisescens
Ptilodon capucinus
Ectropis grisescens
CAJHUA010000005.1 10193835.10198656 Autographa gamma
CAJHZL010000021.1 134729.139550 Mamestra brassicae
Pieris rapae
CAJHUX010000124.1 2633936.2638758 Euproctis similis
CAJHUX010000174.1 1512495.1517317 Euproctis similis
CAJHUX010000250.1 682937.687759 Euproctis similis
Vanessa cardui
CAJMZN010000075.1 4313759.4318266 Vanessa cardui
CAJHUY010000088.1 316293.321449 Cyaniris semiargus
DWHW01001352.1 2.5076 Zerynthia polyxena
Vanessa cardui
CAJHUN010000073.1 4061456.4066594 Aricia agestis
CAJHUW010000011.1 6732089.6737217 Erynnis tages
CAJHUY010000030.1 1176489.1181646 Cyaniris semiargus
CAJHUY010000065.1 1171637.1176785 Cyaniris semiargus
CAJOSV010000025.1 2708796.2713947 Lycaena phlaeas
CAJNAE010000125.1 547677.552484 Lysandra coridon
CAJOSR010000320.1 1406546.1411686 Lysandra bellargus
CAJNAE010000284.1 175958.180754 Lysandra coridon
Drosophila biarmipes
JFBF01000201.1 310245.315026 Bactrocera dorsalis
DWOH01000281.1 282.5439 Signeta flammeata
BLXV01000003.1 10750556.10755705 Samia ricini
BLXV01000007.1 13697906.13703072 Samia ricini
BLXV01000007.1 15019349.15023900 Samia ricini
BLXV01000007.1 15866984.15871803 Samia ricini
BLXV01000001.1 16639069.16643890 Samia ricini
DWAC01000453.1 1451.6623 Curetis bulis
BLXV01000003.1 13005615.13010779 Samia ricini
BLXV01000004.1 17631991.17637154 Samia ricini
Dryas iulia
DWJD01000127.1 1144.5908 Aurina azines
DWHR010001274.1 81.5252 Protesilaus protesilaus
DWQC01000345.1 112.4922 Pythonides amaryllis
DVYP01001143.1 1.4807 Emesis heterochroa
DWKU01000031.1 2444.7269 Mimia cf. chiapaensis
DWQQ01006667.1 124.5302 Oxynetra roscius
DWJG01000135.1 5347.9705 Ectomis octomaculata
DWDV01005608.1 2783.7956 Mimoniades ocyalus
DWES01000095.1 1328.6454 Duroca duroca
JABMBV010043498.1 1630.6800 Mesochorus sp.
FAUM01003789.1 1.4923 Heliconius wallacei
DWEU01002184.1 201.5367 Charidia lucaria
PPHH01001895.1 1995326.2000041 Trichoplusia ni
Eumaeus atala
DWPY01000040.1 1418.6608 Eburuncus unifasciata
DWOG01000018.1 2139.7331 Mylon lassia
DWKJ01000024.1 3286.8454 Cecropterus casica
DWHJ01000209.1 1124.5942 Dalla cyprius
DWIN01000090.1 544.5362 Zopyrion sandace
DWTU01020815.1 2.5164 Cecropterus confusis
CAJQFV010000391.1 2439220.2444392 Autographa pulchrina
Parnassius apollo
DWHT01000605.1 4715.9222 Parnassius smintheus
CAJQZP010000220.1 17370713.17376024 Parnassius apollo
DWQE01036201.1 1.5321 Parnassius apollo
CAJQZP010000349.1 456251.461622 Parnassius apollo
Atethmia centrago
CAKAJF010000605.1 396300.401682 Griposia aprilina
CAJUUQ010000015.1 15920890.15926267 Mythimna ferrago
DWTE010001035.1 1801.6867 Parnassius imperator
Spilosoma lubricipeda
RSAL01000011.1 2782940.2788264 Chilo suppressalis
VNJW01002308.1 9702.14301 Drosophila auraria
CAKAIX010000158.1 18940.23791 Cheilosia vulpina
HQ009558.1 Cotesia vestalis bracovirus segment c35
CAJVQM010000016.1 22902664.22907953 Pyrgus malvae
CAJHZM010000079.1 760808.766097 Pararge aegeria
CAJHZM010000138.1 366440.371729 Pararge aegeria
CAJHZM010000041.1 658925.664220 Pararge aegeria
CAJHZM010000038.1 5963116.5968405 Pararge aegeria
CAJHZM010000045.1 4316124.4321413 Pararge aegeria
CAJHZM010000059.1 2265283.2270572 Pararge aegeria
CAJHZN010000005.1 288029.293318 Pararge aegeria
CAJHZM010000271.1 10360.15649 Pararge aegeria
UIGX01030406.1 1.5288 Pararge aegeria
CAJHZM010000011.1 2007535.2012824 Pararge aegeria
CAJHZN010000001.1 159388.164677 Pararge aegeria
CAJHZM010000102.1 657339.662628 Pararge aegeria
CAJHZM010000035.1 19926.25209 Pararge aegeria
CAJHZM010000092.1 631160.636449 Pararge aegeria
CAJHZM010000110.1 4418285.4423571 Pararge aegeria
CAJVQM010000016.1 22913002.22918291 Pyrgus malvae
CAJHZN010000001.1 169618.174907 Pararge aegeria
CAJHZM010000080.1 261023.266312 Pararge aegeria
CAJHZM010000036.1 7433354.7438643 Pararge aegeria
CAJHZM010000014.1 1104039.1109328 Pararge aegeria
CAJHZM010000130.1 1146594.1151884 Pararge aegeria
CAJHZM010000035.1 670716.676005 Pararge aegeria
CAJHZM010000059.1 423418.428707 Pararge aegeria
CAJHZM010000104.1 145538.150827 Pararge aegeria
JZSA01007369.1 21189.25964 Cotesia vestalis
CAJHZM010000002.1 2420860.2426149 Pararge aegeria
CAJHZM010000012.1 12234942.12240231 Pararge aegeria
CAJHZN010000006.1 22939.28228 Pararge aegeria
CAJHZM010000024.1 3952072.3957361 Pararge aegeria
CAJHZM010000130.1 840737.846026 Pararge aegeria
CAJHZM010000130.1 729237.734526 Pararge aegeria
CAJHZM010000035.1 609216.614505 Pararge aegeria
CAJHZM010000035.1 617278.622567 Pararge aegeria
Hydriomena furcata
CAJVWE010000011.1 3500217.3505507 Campaea margaritaria
CAJVWE010000087.1 1169967.1174606 Campaea margaritaria
Amphipyra tragopoginis
Laothoe populi
Clostera curtula
CAJNAB010000107.1 670398.675633 Laothoe populi
CAJNAB010000435.1 67852.73189 Laothoe populi
BHWX01000022.1 752045.757343 Bombyx mori
BHWX01000007.1 11846872.11852161 Bombyx mori
DWHE01006833.1 1.5238 Satarupa nymphalis
CAJRBD010001370.1 98905.104202 Habrosyne pyritoides
BHWX01000001.1 13148640.13153928 Bombyx mori
Pyrochroa serraticornis
CAJHUU010000108.1 3275763.3280672 Apotomis turbidana
CAJQEY010000087.1 6085973.6091264 Fabriciana adippe
Teleopsis dalmanni
CAJZCT010000117.1 1734641.1739951 Eilema sororculum
JAAAKD010007556.1 22285728.22290784 Danaus melanippus
CAJMZR010001290.1 460177.465027 Tachina fera
DWMK01000004.1 19.5148 Katreus holocausta
WUWR01000078.1 21074.25328 Hyles vespertilio
DWLK01000137.1 1.4747 Celaenorrhinus cf. opalinus
BOPP01000152.1 2643530.2648840 Gryllus bimaculatus
CAJQFW010000081.1 700146.704641 Chrysoperla carnea
DWAK01008481.1 1.4253 Lepidochrysops patricia
Pyrrhopyge (6 species)
DWPO01000177.1 2362.6863 Morvina fissimacula
DWQB01000133.1 862.6189 Gindanes brontinus
DVYD01000130.1 683.5891 Apodemia ares
DWMC01000015.1 734.5673 Ouleus salvina
DWSZ010000156.1 1.5209 Emesis lacrines
DVZB010000096.1 227.5537 Emesis aurimna
Ctenocephalides felis
DWSV010000232.1 290.4623 Emesis ocypore
CAJQGC010000043.1 973037.977303 Schrankia costaestrigalis
JAECXK010000002.1 87239.91306 Drosophila ficusphila
WIOZ01000119.1 47745.51620 Drosophila bifasciata
BMAV01001078.1 255549.259048 Trichonephila inaurata madagascariensis
Thymelicus sylvestris
CAJNAI010000002.1 5361683.5366998 Nymphalis polychloros
Arctia plantaginis
Teleogryllus occipitalis
JADODW010002850.1 173738.178412 Vandiemenella viatica
JAFNEN010000473.1 151816.155317 Oedothorax gibbosus
Blastobasis lacticolella
Drosophila obscura
CAJSMA010000011.1 3727558.3732195 Blastobasis adustella
CAJSMA010000156.1 848419.853213 Blastobasis adustella
CAJSMA010000114.1 315849.320627 Blastobasis adustella
JADLOZ010000397.1 7224651.7229441 Ectropis grisescens
JAECWS010000095.1 10811294.10815056 Drosophila ambigua
DWIA01002199.1 1562.6349 Archon apollinus
CAJSMA010000017.1 23489.28276 Blastobasis adustella
CAJVQM010000028.1 32499509.32504298 Pyrgus malvae
OUUW01000028.1 232789.236744 Drosophila guanche
Chrysoteuchia culmella
Lymantria monacha
Melanostoma mellinum
Lymantria dispar
JADWQK010000266.1 7378926.7383667 Sesamia nonagrioides
Scaptomyza montana
RKRM01002430.1 4632.9349 Scaptomyza flava
DWOM01000141.1 313.5042 Timochares trifasciata
DWLJ01000697.1 3.4421 Autochton oryx
DWHR010000359.1 2831.7604 Protesilaus protesilaus
DWOJ01002358.1 82.4857 Piruna pirus
NQII01007758.1 22450.27113 Clitarchus hookeri

Data S1. Biopython script to only include sequences with > 70% (3705 bp) and to edit FASTA descriptions to contain only the hit accession number, the sequence match range and the species name.

```
>>> from Bio import SeqIO
>>> large_sequences = []
>>> for record in SeqIO.parse("blast_results.txt", "fasta"):
        if len(record.seq) > 3705:
                large_sequences.append(record)

>>> SeqIO.write(large_sequences, "large_seq.fasta", "fasta")
>>> clean_sequences = []
>>> for seq_record in SeqIO.parse("large_seq.fasta", "fasta"):
        seq_record.id = ((seq_record.description.split()[0])+(" ")+
                         (seq_record.description.split()[1])+(" ")+
                         (seq_record.description.split()[2]))
        seq_record.description = ("")
        clean_sequences.append(seq_record)

>>> SeqIO.write(clean_sequences, "clean_seq.fasta", "fasta")
```

Data S2. List of sequences descriptions used in the analysis, with their accession number, match range and the species name.

```
HQ009558.1 Cotesia vestalis bracovirus segment c35
CAJHZN010000006.1_22939.28228_Pararge_aegeria
CAJHZN010000001.1_159388.164677_Pararge_aegeria
CAJHZN010000001.1_169618.174907_Pararge_aegeria
CAJHZM010000138.1_366440.371729_Pararge_aegeria
CAJHZM010000104.1_145538.150827_Pararge_aegeria
CAJHZM010000080.1_261023.266312_Pararge_aegeria
CAJHZM010000059.1_423418.428707_Pararge_aegeria
CAJHZM010000045.1_4316124.4321413_Pararge_aegeria
CAJHZM010000014.1_1104039.1109328_Pararge_aegeria
CAJHZM010000011.1_2007535.2012824_Pararge_aegeria
CAJVQM010000016.1_22902664.22907953_Pyrgus_malvae
CAJHZM010000035.1_19926.25209_Pararge_aegeria
CAJHZN010000005.1_288029.293318_Pararge_aegeria
CAJHZM010000130.1_729237.734526_Pararge_aegeria
CAJHZM010000130.1_840737.846026_Pararge_aegeria
UIGX01030406.1_1.5288_Pararge_aegeria
CAJHZM010000036.1_7433354.7438643_Pararge_aegeria
CAJHZM010000035.1_609216.614505_Pararge_aegeria
CAJHZM010000035.1_617278.622567_Pararge_aegeria
CAJVQM010000016.1_22913002.22918291_Pyrgus_malvae
CAJHZM010000079.1_760808.766097_Pararge_aegeria
CAJHZM010000059.1_2265283.2270572_Pararge_aegeria
CAJHZM010000271.1_10360.15649_Pararge_aegeria
JZSA01007369.1_21189.25964_Cotesia_vestalis
CAJHZM010000038.1_5963116.5968405_Pararge_aegeria
CAJHZM010000012.1_12234942.12240231_Pararge_aegeria
CAJHZM010000002.1_2420860.2426149_Pararge_aegeria
CAJHZM010000035.1_670716.676005_Pararge_aegeria
CAJHZM010000041.1_658925.664220_Pararge_aegeria
CAJHZM010000110.1_4418285.4423571_Pararge_aegeria
CAJHZM010000130.1_1146594.1151884_Pararge_aegeria
CAJHZM010000102.1_657339.662628_Pararge_aegeria
CAJHZM010000092.1_631160.636449_Pararge_aegeria
CAJHZM010000024.1_3952072.3957361_Pararge_aegeria
CAJVWE010000011.1_3500217.3505507_Campaea_margaritaria
CAJVWE010000087.1_1169967.1174606_Campaea_margaritaria
CAJVWD010000025.1_408819.414108_Hydriomena_furcata
CAJVWD010000070.1_510413.515617_Hydriomena_furcata
CAJVWD010000025.1_377294.382499_Hydriomena_furcata
CAJVWD010000012.1_3716744.3722060_Hydriomena_furcata
CAJMZV010000185.1_996280.1001569_Amphipyra_tragopoginis
CAJMZV010000218.1_4569479.4574767_Amphipyra_tragopoginis
CAJOSM010000467.1_148944.154232_Pyrochroa_serraticornis
CAJOSM010000275.1_305597.310868_Pyrochroa_serraticornis
CAJOSM010000934.1_7793.12723_Pyrochroa_serraticornis
CAJOSL010000003.1_780653.785583_Pyrochroa_serraticornis
CAJOSM010000568.1_196017.200944_Pyrochroa_serraticornis
BHWX01000001.1_13148640.13153928_Bombyx_mori
DWHE01006833.1_1.5238_Satarupa_nymphalis
CAJNAB010000311.1_202758.208045_Laothoe_populi
CAJNAB010000435.1_67852.73189_Laothoe_populi
BHWX01000007.1_11846872.11852161_Bombyx_mori
CAJNAB010000107.1_670398.675633_Laothoe_populi
CAJNAB010000260.1_341534.346834_Laothoe_populi
CAJRBD010001370.1_98905.104202_Habrosyne_pyritoides
CAJQEY010000087.1_6085973.6091264_Fabriciana_adippe
CAJNAB010000869.1_72668.77560_Laothoe_populi
CAJQFY010000135.1_335220.340495_Clostera_curtula
CAJQFY010000347.1_1.5279_Clostera_curtula
CAJQFY010001131.1_62176.67434_Clostera_curtula
BHWX01000022.1_752045.757343_Bombyx_mori
CAJNAB010000154.1_1044455.1049578_Laothoe_populi
CAJNAB010001387.1_170726.175895_Laothoe_populi
```

```
CAJHUU010000108.1_3275763.3280672_Apotomis_turbidana
DWQB01000133.1_862.6189_Gindanes_brontinus
DVYD01000130.1_683.5891_Apodemia_ares
DWEO01003961.1_926.5870_Pyrrhopyge_telassa
DVQA01000780.1_1045.5981_Pyrrhopyge_sergius
DWOT01000734.1_76.5034_Pyrrhopyge_hadassa
DVQD01000579.1_78.4419_Pyrrhopyge_kelita
DWEP01000053.1_1589.6485_Pyrrhopyge_crida
DWSZ010000156.1_1.5209_Emesis_lacrines
CAJZCT010000117.1_1734641.1739951_Eilema_sororculum
DWLK01000137.1_1.4747_Celaenorrhinus_cf._opalinus
DWMK01000004.1_19.5148_Katreus_holocausta
DWFE01002627.1_226.5042_Pyrrhopyge_pelota
NLCU02019152.1_17889.23172_Teleopsis_dalmanni
JACTOK010041713.1_346.5639_Teleopsis_dalmanni
JACTOK010037361.1_40126.45029_Teleopsis_dalmanni
JACTOK010035260.1_62.5356_Teleopsis_dalmanni
DWPO01000177.1_2362.6863_Morvina_fissimacula
DWMC01000015.1_734.5673_Ouleus_salvina
DVZB010000096.1_227.5537_Emesis_aurimna
WUWR01000078.1_21074.25328_Hyles_vespertilio
DWSV010000232.1_290.4623_Emesis_ocypore
CAKAJF010000605.1_396300.401682_Griposia_aprilina
DWQE01036201.1_1.5321_Parnassius_apollo
DWTE010001035.1_1801.6867_Parnassius_imperator
DWHT01000605.1_4715.9222_Parnassius_smintheus
CAJOSY010001674.1_384948.390327_Atethmia_centrago
CAJOSY010000266.1_43501.48893_Atethmia_centrago
CAJOSY010000707.1_37529.42796_Atethmia_centrago
CAJOSY010001674.1_347256.352311_Atethmia_centrago
CAJQZP010000349.1_456251.461622_Parnassius_apollo
CAJQZP010001011.1_5021097.5026488_Parnassius_apollo
CAJQZP010000971.1_1549.6922_Parnassius_apollo
CAKAIX010000158.1_18940.23791_Cheilosia_vulpina
CAJUUQ010000015.1_15920890.15926267_Mythimna_ferrago
RSAL01000011.1_2782940.2788264_Chilo_suppressalis
CAJNAL010000544.1_199300.204585_Spilosoma_lubricipeda
CAJNAL010000084.1_2628222.2633568_Spilosoma_lubricipeda
CAJNAL010000037.1_461673.466985_Spilosoma_lubricipeda
CAJQZP010000220.1_17370713.17376024_Parnassius_apollo
DWKJ01000024.1_3286.8454_Cecropterus_casica
DWOG01000018.1_2139.7331_Mylon_lassia
DWPY01000040.1_1418.6608_Eburuncus_unifasciata
DWAC01000453.1_1451.6623_Curetis_bulis
DWQQ01006667.1_124.5302_Oxynetra_roscius
DWES01000095.1_1328.6454_Duroca_duroca
DVYP01001143.1_1.4807_Emesis_heterochroa
DWEU01002184.1_201.5367_Charidia_lucaria
DWHR010001274.1_81.5252_Protesilaus_protesilaus
DWJD01000127.1_1144.5908_Aurina_azines
DWKU01000031.1_2444.7269_Mimia_cf._chiapaensis
DWQC01000345.1_112.4922_Pythonides_amaryllis
DWIN01000090.1_544.5362_Zopyrion_sandace
DWDV01005608.1_2783.7956_Mimoniades_ocyalus
DWHJ01000209.1_1124.5942_Dalla_cyprius
FAUM01003789.1_1.4923_Heliconius_wallacei
JABMBV010043498.1_1630.6800_Mesochorus_sp.
CAJQFV010000391.1_2439220.2444392_Autographa_pulchrina
DWOH01000281.1_282.5439_Signeta_flammeata
DWHW01001352.1_2.5076_Zerynthia_polyxena
CAJHUW010000011.1_6732089.6737217_Erynnis_tages
CAJHUY010000088.1_316293.321449_Cyaniris_semiargus
CAJHUY010000030.1_1176489.1181646_Cyaniris_semiargus
CAJNAE010000125.1_547677.552484_Lysandra_coridon
CAJOSV010000025.1_2708796.2713947_Lycaena_phlaeas
CAJNAE010000284.1_175958.180754_Lysandra_coridon
CAJHUN010000073.1_4061456.4066594_Aricia_agestis
CAJHUY010000065.1_1171637.1176785_Cyaniris_semiargus
```

```
CAJOSR010000320.1_1406546.1411686_Lysandra_bellargus
CAJMZN010000024.1_1507394.1512552_Vanessa_cardui
CAJMZN010000124.1_3678898.3684055_Vanessa_cardui
JFBF01000201.1_310245.315026_Bactrocera_dorsalis
BLXV01000007.1_15866984.15871803_Samia_ricini
DWJG01000135.1_5347.9705_Ectomis_octomaculata
BLXV01000001.1_16639069.16643890_Samia_ricini
BLXV01000007.1_15019349.15023900_Samia_ricini
BLXV01000007.1_13697906.13703072_Samia_ricini
BLXV01000004.1_17631991.17637154_Samia_ricini
BLXV01000003.1_13005615.13010779_Samia_ricini
DWTU01020815.1_2.5164_Cecropterus_confusis
PPHH01001895.1_1995326.2000041_Trichoplusia_ni
CAJMZN010000221.1_54458.59610_Vanessa_cardui
CAJMZN010000118.1_342983.348103_Vanessa_cardui
SZUW01001905.1_136961.141302_Drosophila_biarmipes
SZUW01001905.1_119326.123659_Drosophila_biarmipes
SZUW01001905.1_128145.132475_Drosophila_biarmipes
SZUW01001905.1_123735.128069_Drosophila_biarmipes
SZUW01001905.1_132551.136885_Drosophila_biarmipes
SZUW01001905.1_114917.119250_Drosophila_biarmipes
SZUW01001905.1_141332.145691_Drosophila_biarmipes
BLXV01000003.1_10750556.10755705_Samia_ricini
JADLOZ010000452.1_2509044.2514246_Ectropis_grisescens
JADLOZ010000461.1_14720351.14725555_Ectropis_grisescens
JADLOZ010000163.1_10872915.10878112_Ectropis_grisescens
JADLOZ010000001.1_13982744.13987947_Ectropis_grisescens
CAJHUX010000250.1_682937.687759_Euproctis_similis
CAJHUX010000124.1_2633936.2638758_Euproctis_similis
CAJHUX010000174.1_1512495.1517317_Euproctis_similis
CAJHZL010000021.1_134729.139550_Mamestra_brassicae
CAJHZL010000102.1_1.4272_Mamestra_brassicae
CAJHUA010000005.1_10193835.10198656_Autographa_gamma
CAJHWT010000024.1_864030.868851_Pieris_rapae
CAJHWT010000025.1_8267413.8272232_Pieris_rapae
CAJHWT010000037.1_2774121.2778939_Pieris_rapae
CAJZHO010000673.1_459067.463887_Spilarctia_lutea
CAJZHO010000165.1_1345674.1350495_Spilarctia_lutea
CAJZHO010000105.1_881927.886748_Spilarctia_lutea
CAJZHO010000219.1_260539.265361_Spilarctia_lutea
CAJZHO010000021.1_867128.871949_Spilarctia_lutea
CAJZHO010000673.1_453918.458739_Spilarctia_lutea
CAJZHO010000143.1_1190670.1195491_Spilarctia_lutea
JADLOZ010000086.1_11071347.11076164_Ectropis_grisescens
JADLOZ010000127.1_29382932.29387733_Ectropis_grisescens
CAJZHO010000233.1_432824.437645_Spilarctia_lutea
CAJZHO010000266.1_548611.553432_Spilarctia_lutea
CAJZHO010000274.1_1798316.1803137_Spilarctia_lutea
CAJZHO010000256.1_664353.669174_Spilarctia_lutea
CAJZBM010000107.1_62863.67683_Ptilodon_capucinus
CAJZBM010000088.1_746415.751235_Ptilodon_capucinus
CAJZBM010000074.1_502408.507228_Ptilodon_capucinus
CAJZBM010000044.1_1818003.1822823_Ptilodon_capucinus
JADLOZ010000416.1_12189676.12194702_Ectropis_grisescens
JADLOZ010000492.1_6103834.6108866_Ectropis_grisescens
JAHESG010000032.1_186837.191985_Dryas_iulia
JAHESG010000017.1_6164655.6169460_Dryas_iulia
JAHESG010000029.1_7559667.7564471_Dryas_iulia
JAHESG010000004.1_15479610.15484420_Dryas_iulia
JAHESG010000016.1_3595041.3599846_Dryas_iulia
JAHESG010000029.1_3085886.3090704_Dryas_iulia
JAHESG010000009.1_9945615.9950420_Dryas_iulia
CAJMZN010000075.1_4313759.4318266_Vanessa_cardui
CAJQFW010000081.1_700146.704641_Chrysoperla_carnea
CAJMZR010001290.1_460177.465027_Tachina_fera
DWAK01008481.1_1.4253_Lepidochrysops_patricia
JAFELO010000258.1_750129.755321_Eumaeus_atala
JAFELO010000541.1_77968.82990_Eumaeus_atala
```

```
JAFELO010000139.1_388908.393951_Eumaeus_atala
JAFELO010001003.1_188367.193567_Eumaeus_atala
JAFELO010000370.1_1179360.1184501_Eumaeus_atala
JAFELO010000151.1_120296.125456_Eumaeus_atala
JAFELO010001549.1_174743.179937_Eumaeus_atala
CAJQGC010000043.1_973037.977303_Schrankia_costaestrigalis
JAECXK010000002.1_87239.91306_Drosophila_ficusphila
JAAAKD010007556.1_22285728.22290784_Danaus_melanippus
VNJW01002308.1_9702.14301_Drosophila_auraria
QVAO01001492.1_48901893.48906517_Ctenocephalides_felis
QVAO01000696.1_15497.20117_Ctenocephalides_felis
QVAO01001492.1_52438176.52442793_Ctenocephalides_felis
QVAO01002597.1_29873.34493_Ctenocephalides_felis
QVAO01001409.1_9024.13640_Ctenocephalides_felis
CAJVQL010000107.1_930262.935515_Thymelicus_sylvestris
CAJVQL010000166.1_2342372.2347626_Thymelicus_sylvestris
CAJVQL010000077.1_2140588.2145740_Thymelicus_sylvestris
CAJVQL010000077.1_2145744.2150856_Thymelicus_sylvestris
CAJVQL010000007.1_627957.633205_Thymelicus_sylvestris
CAJVQL010000185.1_2083593.2088691_Thymelicus_sylvestris
CAJNAI010000002.1_5361683.5366998_Nymphalis_polychloros
CADEBC010000506.1_2189273.2194586_Arctia_plantaginis
CADEBD010000226.1_11524308.11529654_Arctia_plantaginis
CADEBD010000226.1_5511712.5517026_Arctia_plantaginis
BMAV01001078.1_255549.259048_Trichonephila_inaurata_madagascariensis
BLKR01001327.1_272646.277539_Teleogryllus_occipitalis
BLKR01000082.1_620997.625590_Teleogryllus_occipitalis
BLKR01001662.1_236479.241317_Teleogryllus_occipitalis
BLKR01000118.1_545117.549975_Teleogryllus_occipitalis
WIOZ01000119.1_47745.51620_Drosophila_bifasciata
BOPP01000152.1_2643530.2648840_Gryllus_bimaculatus
JADODW010002850.1_173738.178412_Vandiemenella_viatica
JAFNEN010000473.1_151816.155317_Oedothorax_gibbosus
DWOJ01002358.1_82.4857_Piruna_pirus
DWHR010000359.1_2831.7604_Protesilaus_protesilaus
CAJHUO010000233.1_1599422.1604185_Blastobasis_lacticolella
CAJHUO010000267.1_2646904.2651683_Blastobasis_lacticolella
DWIA01002199.1_1562.6349_Archon_apollinus
CAJHUO010000037.1_3811342.3816132_Blastobasis_lacticolella
CAJHUO010000261.1_1929974.1934751_Blastobasis_lacticolella
CAJHUO010000046.1_898348.903124_Blastobasis_lacticolella
CAJHUO010000427.1_199617.204394_Blastobasis_lacticolella
CAJHUO010000318.1_112116.116892_Blastobasis_lacticolella
CAJHUO010000318.1_123543.128316_Blastobasis_lacticolella
CAJHUO010000345.1_691171.695944_Blastobasis_lacticolella
CAJHUO010000401.1_1220232.1225032_Blastobasis_lacticolella
CAJHUO010000228.1_159363.164129_Blastobasis_lacticolella
CAJHUO010000331.1_164805.169582_Blastobasis_lacticolella
CAJHUO010000118.1_176842.181594_Blastobasis_lacticolella
CAJHUO010000212.1_73386.78146_Blastobasis_lacticolella
CAJVQM010000028.1_32499509.32504298_Pyrgus_malvae
CAJHUO010000203.1_1777648.1782415_Blastobasis_lacticolella
CAJSMA010000017.1_23489.28276_Blastobasis_adustella
CAJSMA010000011.1_3727558.3732195_Blastobasis_adustella
CAJSMA010000156.1_848419.853213_Blastobasis_adustella
JADLOZ010000397.1_7224651.7229441_Ectropis_grisescens
CAJSMA010000114.1_315849.320627_Blastobasis_adustella
BDQP01000130.1_301209.305884_Drosophila_obscura
JAECWW010000026.1_245028.249259_Drosophila_obscura
CAJUUO010000047.1_3826225.3831009_Chrysoteuchia_culmella
CAJUUO010000001.1_1104634.1109407_Chrysoteuchia_culmella
JAECWW010000075.1_438317.443049_Drosophila_obscura
CAJHZW010001009.1_110301.115094_Lymantria_monacha
CAJHZW010000306.1_1864909.1869687_Lymantria_monacha
CAJHZW010000292.1_841817.846553_Lymantria_monacha
CAJHZW010000280.1_101693.106468_Lymantria_monacha
JAFEKU010001978.1_7197.11978_Lymantria_dispar
JAFEKT010001198.1_1395550.1400311_Lymantria_dispar
```

```
JAFEKT010002896.1_71452.76223_Lymantria_dispar
CAJHZW010001466.1_82990.87783_Lymantria_monacha
CAJHZW010000168.1_160361.165147_Lymantria_monacha
JAFEKT010001681.1_858485.863270_Lymantria_dispar
JAFEKU010001203.1_63466.68240_Lymantria_dispar
JAECWS010000095.1_10811294.10815056_Drosophila_ambigua
DWOM01000141.1_313.5042_Timochares_trifasciata
DWLJ01000697.1_3.4421_Autochton_oryx
JAEIGR010000030.1_263589.268325_Scaptomyza_montana
JAEIGR010000013.1_846018.850751_Scaptomyza_montana
RKRM01002430.1_4632.9349_Scaptomyza_flava
CAJZBV010000028.1_2548868.2553642_Melanostoma_mellinum
CAJZBV010000338.1_234267.239006_Melanostoma_mellinum
CAJZBV010000077.1_1060277.1065056_Melanostoma_mellinum
JADWQK010000266.1_7378926.7383667_Sesamia_nonagrioides
OUUW01000028.1_232789.236744_Drosophila_guanche
NQII01007758.1_22450.27113_Clitarchus_hookeri
```