



Universidade de Brasília
Faculdade de Ciências da Saúde
Curso de Graduação em Farmácia

**Prospecção de Enzimas do Complexo CRISPR/Cas a partir do
Microbioma Intestinal do Cumpim *Syntermes wheeleri***

Autor: João Lucas da Silva Correia
Orientador: Prof. Dr. Ricardo Henrique Krüger
Co-orientadora: Carla Simone Vizzotto

Brasília, 2022

Prospecção de Enzimas do Complexo CRISPR/Cas a partir do Microbioma Intestinal do Cumpim *Syntermes wheeleri*

Trabalho de Conclusão de Curso apresentada como requisito parcial para obtenção do título de Bacharel em Farmácia, pela Faculdade de Ciências da Saúde, Universidade de Brasília.

Orientador: Ricardo Henrique Krüger

Co-orientadora: Carla Simone Vizzotto

Brasília, 2022



Universidade de Brasília
Faculdade de Ciências da Saúde
Curso de Graduação em Farmácia

João Lucas da Silva Correia

Banca Examinadora

Professor orientador: Dr. Ricardo Henrique Krüger – Instituto de Biologia –
Universidade de Brasília – UnB

Co-orientadora: Dra. Carla Simone Vizzotto – Departamento de Engenharia
Civil e Ambiental, Faculdade de Tecnologia – Universidade de Brasília - UnB

Examinadora externa: Professora Dra. Eliane Ferreira Noronha – Instituto de
Biologia – Universidade de Brasília – UnB

AGRADECIMENTOS

Agradeço ao meu professor Dr. Ricardo Henrique Krüger pela orientação. À Dra. Carla Simone Vizzotto pelo apoio e co-orientação. Agradeço também à Universidade de Brasília por toda estrutura disponibilizada. Agradeço a minha família que esteve comigo este tempo todo.

... serei brutal e malfeita como uma pedra, serei leve e vaga como o que se sente e não se entende, me ultrapassarei em ondas, ah, Deus, e que tudo venha e caia sobre mim, até a incompreensão de mim mesma em certos momentos brancos porque basta me cumprir e então nada impedirá meu caminho até a morte-sem-medo, de qualquer luta ou descanso me levantarei forte e bela como um cavalo novo.

Clarice Lispector

RESUMO

O sequenciamento metagenômico em plataforma *NGS Illumina* da microbiota intestinal do cupim *Syntermes wheeleri* permitiu, além do aprofundamento na compreensão do seu nicho ecológico, a prospecção de produtos biotecnológicos. Esse trabalho soma-se aos esforços na prospecção e caracterização de enzimas de interesse biotecnológico, tendo como objetivo prospectar enzimas do complexo *CRISPR/Cas* a partir de dados metagenômicos da microbiota intestinal do cupim *Syntermes wheeleri*. As sequências metagenômicas da microbiota associada ao cupim *Syntermes wheeleri* serão analisadas para a identificação de enzimas associadas ao complexo *CRISPR/Cas* de edição gênica. A seleção das enzimas de interesse ocorreu a partir de sistema de busca pelo nome de cada enzima em banco de dados *ggKbase* (<https://ggkbase.berkeley.edu/>). As sequências obtidas foram avaliadas quanto à qualidade seguindo parâmetros de completude, tamanho e similaridade com bancos de dados de proteínas caracterizadas experimentalmente. Ao todo foram obtidas 288 enzimas do complexo *CRISPR/Cas* no banco de dados do microbioma intestinal de *Syntermes wheeleri*. Dessas, 127 apresentavam códon de início e parada da tradução. O tamanho de 110 enzimas foi compatível com o tamanho de proteínas já caracterizadas experimentalmente. No total, 105 sequências apresentaram ao menos uma sequência de AA com identidade no banco de dados *PDB*. Ao final, foram selecionadas 66 proteínas que tem potencial para futuras sínteses por expressão heteróloga e caracterização estrutural e enzimática

Palavras-chave: Microbiota intestinal, Cupim, *CRISPR*, *Cas*, Enzimas, Biotecnologia.

Sumário

AGRADECIMENTOS	4
RESUMO	6
INTRODUÇÃO	1
Contexto Ambiental.....	1
Metagenômica e bioprospecção	3
Enzimas e Biotecnologia.....	5
Enzimas do complexo <i>CRISPR/Cas</i>	7
OBJETIVOS	10
Objetivo geral	10
Objetivos específicos	10
METODOLOGIA.....	11
Sequências metagenômicas	11
Seleção de enzimas do complexo <i>CRISPR/Cas</i>	11
RESULTADOS E DISCUSSÃO.....	12
PERSPECTIVAS.....	19
CONCLUSÃO.....	20
REFERÊNCIAS.....	21

INTRODUÇÃO

Contexto Ambiental

O solo possui grande diversidade de microrganismos, que apresentam diferentes estratégias de vida e papéis funcionais (DELGADO-BAQUERIZO et al., 2018). Esse ambiente é rico em como bactérias, fungos e outros microrganismos, sendo vital para o equilíbrio no planeta (RICHTER et al., 2022). A complexidade da microbiota dos solos ainda não é completamente conhecida. Porém, é bastante consolidado que as microbiotas do planeta Terra são essenciais para sustentar a vida no planeta, ou seja, os organismos vegetais, humanos e animais são dependentes em algum grau dos microbiomas. (MUJTAR et al., 2019; CAMERON et al., 2018; DELGADO-BAQUERIZO et al., 2018; MORAN, 2015).

O ambiente e os seres vivos albergam vasta composição filogenética com amplo potencial genômico (ZANEVELD et al., 2011; STALEY al., 2014; LANGILLE et al., 2013; WARNECKE et al., 2007). Os microbiomas de diferentes habitats apresentam-se como alternativas para a busca por novas moléculas bioativas, funcionando como reservatório de produtos com potencial biotecnológico. Mesmo diante de tamanha oportunidade, a pesquisa por compostos de interesse biotecnológico em alguns ambientes ainda é pouco explorada (WARNECKE et al., 2019).

A diversidade microbiana exerce funções fundamentais em reações bioquímicas e estruturação do solo, decomposição da matéria orgânica, controle de pragas, favorecimento do crescimento de vegetais, processos de ciclo biogeoquímicos, biodegradação de pesticidas e outras moléculas, supressão de doenças e no próprio funcionamento do ecossistema (DINI-ANDREOTE e VAN ELSAS, 2019; COTTA, 2016; BELL et al., 2005). Moléculas sintetizadas pela comunidade microbiana, influenciam em funções fisiológicas e ecológicas nesse ambiente (NEWMAN et al., 2020; DAVIES, 2013). Essas características conferem aos microrganismos um grande potencial biotecnológico. Suas contribuições podem ser usadas, por exemplo, na produção de bioinoculantes para produção agropecuária, na atividade de biorremediação, controle biológico

de pragas, produção de moléculas com funções farmacológicas, como enzimas, antibióticos, e outras substâncias químicas. (MOREIRA e SIQUEIRA, 2006; PYE et al., 2017). Além da diversidade microbiana encontrada diretamente nos solos, os animais que habitam esse ambiente também possuem uma rica microbiota simbiótica com potencial biotecnológico.

Os cupins são insetos que regulam atividades no solo, como a decomposição de matéria orgânica, a manutenção e desenvolvimento da estrutura do solo (KHAN et al., 2018; ENAGBONMA e BABALOLA, 2019; MYER e FORSCHLER, 2019). A ação desse invertebrado propicia um solo mais rico em nutrientes pelo acúmulo de materiais orgânicos e minerais em diversas profundidades (ENAGBONMA e BABALOLA, 2020; GRIFFITHS et al., 2021; GEDOZ et al., 2021). Alguns estudos também relacionam os cupins com a preservação de água no solo (WATSON, 1969; SHANBHAG et al., 2019).

O nicho ecológico ocupado pelos cupins e a necessidade de avanços na exploração econômica de biomassa vegetal despertaram o interesse pelo seu potencial para a prospecção de enzimas capazes de degradar lignina, hemicelulose e celulose (RASHAMUSE et al., 2017). Esses processos enzimáticos dependem do metabolismo da sua microbiota intestinal (ENAGBONMA et al., 2020). Estudos realizados com o intestino do cupim *Syntermes wheeleri* (figura 1), utilizando técnicas de sequenciamento de DNA e RNA, evidenciaram grande potencial para a prospecção de novas enzimas a partir de microrganismos ainda não cultivados (SANTANA et al., 2015; LIMA et al., 2019).

Os avanços em técnicas de sequenciamento de material genético, proteômica e biologia molecular favorecem o estudo dos microrganismos cultiváveis e não cultiváveis, permitindo os estudos sobre o papel da microbiota na relação simbiótica com o hospedeiro, porém ainda existem muitas lacunas de conhecimento a serem preenchidas (WARNECKE et al., 2007; JOYNSON et al., 2017; LIU et al., 2019; XIA et al., 2017).

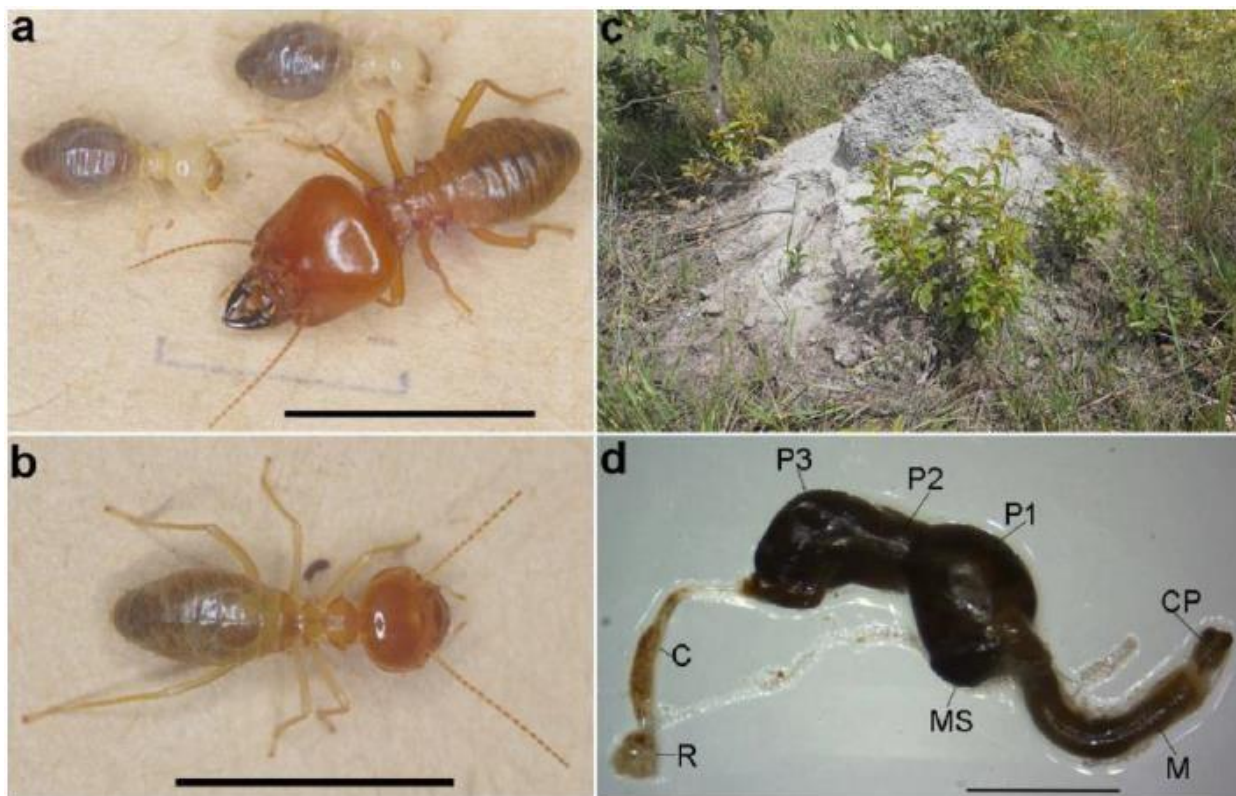


Figura 1 – Operário do cupim neotropical *Syntermes wheeleri*: (a) soldado com cabeça branca, (b) trabalhador de cabeça vermelha, (c) cupinzeiro, (d) compartimentos intestinais: papo(CP), intestino médio (M), segmento misto (MS), valva entérica (P2), tecerico segmento proctodeal (pança-P3), colon (C), primeiro segmento proctodeal (P1), reto (R). (SANTANA et al.,2015; LIMA, 2019).

Estudos revelaram que os microrganismos intestinais dos cupins possuem variadas enzimas, atuando em vários processos fisiológicos e não somente na digestão da lignocelulose (ARORA et al., 2022). A microbiota intestinal tem forte influência na saúde do hospedeiro, pois possibilita reações enzimáticas que o hospedeiro não consegue viabilizar, fornece produtos metabólicos, desenvolve o sistema imunológico e defende contra possíveis patógenos (FLINT et al., 2012; MARANDUBA et al., 2015).

Metagenômica e bioprospecção

Os microrganismos possuem complexas relações em seus habitats naturais. Os níveis de interação entre si, a abundância e o funcionamento dessas comunidades são de interesse da pesquisa científica e tecnológica (LEY ET AL.,2006; GAGNIERE et al., 2016; DICK, 2018).

Os estudos de Roberto Koch e o desenvolvimento de meios de cultura levaram a microbiologia a denominar os microrganismos em cultiváveis e não cultiváveis

(HANDELSMAN, 2004). A observação em microscópio de amostras ambientais apresentava números maiores de células quando comparadas à observação do crescimento em placa de cultivo. Assim, deduziu-se que muitas bactérias ambientais se desenvolviam em meios de cultura, não eram viáveis em cultivos de laboratório (STALEY e KONOPKA, 1985). Segundo TORSVIK et al (1990), em uma amostra de solo era somente possível cultivar de 0,1% a 1% de bactérias presente naquele ambiente por meios de cultivo padrão. Esta dificuldade no cultivo limitava a pesquisa funcional de microbiomas (MEDEMA et al., 2021; ROBISON et al., 2021).

A limitação em recuperar material biológico de microrganismos não cultiváveis era um desafio, porém os avanços tecnológicos possibilitaram o sequenciamento de DNA/RNA sem a necessidade de cultivo. A exploração de bactérias não cultiváveis, agregou novas informações genômicas, concedendo dados sobre diversidade genética e estrutura populacional em amostras ambientais através da aquisição do DNA de uma comunidade inteira (HANDELSMAN et al., 1998; VENTER et al., 2004; WESTMANN et al., 2018).

Inicialmente realizavam-se a extração do DNA total de uma amostra, posterior clonagem em cromossomos artificiais de bactérias e depois o sequenciamento (SCHLEPER et al., 1998).

Esta técnica aumentou a compreensão da diversidade microbiana retirando o viés da cultura e suas possíveis interações. Acrescentando conhecimentos da diversidade genética, seus papéis ecológicos dos microrganismos, comportamento de populações microbiológicas (HANDELSMAN, 2004).

O avanço das técnicas a partir do desenvolvimento do sequenciamento de nova geração (NGS) e redução dos custos por base sequenciada levaram a um aumento exponencial da disponibilidade de genomas e metagenomas em bancos de dados públicos (Zhao et al, 2020). Os resultados desses avanços podem ser utilizados para bioprospecção de composto bioativos inéditos e novas enzimas de interesse industrial (COHEN et al., 2015; COLIN et al., 2015; FORSBER et al., 2016; UFARTÉ et al., 2017; SOMMER et al., 2010).

O sequenciamento metagenômico em plataforma NGS Illumina da microbiota intestinal do cupim *Syntermes wheeleri* permitiu, além do aprofundamento na compreensão do seu nicho ecológico, a prospecção de produtos biotecnológicos. Enzimas degradadoras de biomassa lignocelulósica estão caracterizadas e oxidoredutases com potencial antimicrobiano já estão sendo caracterizados pelo

grupo de trabalho do laboratório de Enzimologia da Universidade de Brasília. Esse trabalho soma-se aos esforços do grupo na prospecção e caracterização de enzimas de interesse biotecnológico. As sequências metagenômicas da microbiota associada ao cupim *Syntermes wheeleri* serão analisadas para a identificação de enzimas associadas ao complexo *CRISPR/Cas* de edição gênica.

Enzimas e Biotecnologia

A Biotecnologia é uma ciência multidisciplinar que utiliza de seres vivos ou parte de sua composição, com o objetivo de obter um determinado novo produto ou serviço (CBD,1992; VARGA et al., 2018). Com os avanços da ciência, pode associar os conhecimentos científicos e tecnológicos nos objetivos de produzir bens e serviços para a humanidade e ambiente com parte destes microrganismos (RENGIFO, 2009). Dentro da biotecnologia, a procura metodológica por organismos, enzimas, compostos, genes ou alguma parte extraída do ser vivo, com o objetivo de desenvolver um produto é conhecida como bioprospecção (JUNIOR, 2011). Normalmente, etapas para descobertas biotecnológicas incluem coleta de material de caráter biológico em determinado local, seleção de microrganismos e análises de caracterização e atividades potenciais com o material selecionado (Bull et al, 2000). A bioprospecção focada em material biológico acarreta processos sustentáveis, pois garante a preservação da natureza, atuando no campo da agricultura, biorremediação, prevenção do ambiente e gerando sustentabilidade (SCHENBERG, 2010).

Há relatos que o uso biotecnológico de microrganismos e suas enzimas está há muito tempo presente no cotidiano humano, desde as primeiras civilizações, como na fermentação de bebidas ou alimentos (FIGUEREDO et al.,2006). Podem ser citados exemplos como a fabricação de queijo a partir do leite de cabra, na obra *Odisséia* de Homero, a fabricação de pão pelos egípcios e a produção de cerveja na Babilônia em meados dos anos 6000 a.C (CARBALLO, 2002; JAY, 2000).

Todas as etapas do metabolismo de organismos vivos são realizadas por múltiplos e eficientes mecanismos de reações e sinalizações químicas, em sua maioria catalisados por enzimas. A presença das proteínas capazes de catalisar reações químicas, ou seja, acelerar a velocidade dessas reações, permite a vida aos seres vivos (SILVA, 2016; MORSHED et al., 2021).

As enzimas são proteínas que contribuem para reações químicas e constituem-

se como ferramentas indispensáveis para a viabilidade metabólica dos seres vivos. Elas podem estar relacionadas a degradação, alteração ou síntese de moléculas. As enzimas podem ser divididas em 6 classes,

denominadas pelas reações catalisadas: oxido redutases (reações de óxido-redução); isomerases (reação de mudança intramolecular); liases (reações de quebra de ligações); hidrolases (reação de hidrólise); transferase (transferência de grupos entre moléculas); ligases (ligações covalentes) (OARLANDELLI et al.,2012).

As enzimas são de grande interesse para o desenvolvimento biotecnológico. Atualmente são utilizadas nos mais diversos processos, como a biologia molecular, biomedicina, no tratamento de resíduos, na fabricação de produtos e desenvolvimento de metodologias analíticas. As enzimas são seletivas em seus alvos, suas reações são brandas, possível fácil manuseio, menor consumo de energia e menor impacto ambiental (OARLANDELLI et al., 2012; SANTOS, 2008).

O uso de microrganismos para produzir enzimas tem sido recorrente para áreas de interesse terapêutico, alimentício, têxtil, química, indústria de papel, indústria de detergente (SRIVASTAVA, 2019; MUSSATTO et al., 2007). É possível retirar as enzimas de células e usá-las como produto ou para engenharia genética (ROBISON, 2015, ORZETTI et al., 2022).

A engenharia genética faz parte da biotecnologia, modifica o genoma de um organismo. Essas alterações são caracterizadas por retirada ou introdução de determinadas características, também pode estar relacionada com produção de novas proteínas ou enzimas (DOS SANTOS e WIETHÖLTER, 2021).

A manipulação genética está envolvida com tratamentos de inúmeras doenças, criação de organismos com detalhes genéticos específicos, criando novos tratamentos e produtos terapêuticos mais efetivos (DOS SANTOS e WIETHÖLTER, 2021). Modificação de genes é interessante para áreas como oncologia, agricultura, outras áreas da medicina (ZHANG et al., 2022; HU e XIONG, 2014; HOCKEMEYER et al., 2011; MAEDER e GERSBACH, 2016).

Edição genética é dividida em duas fases principais: reconhecimento e a clivagem do DNA, e após isso o reparo. Existem várias técnicas de edição de genes, porém o uso da edição baseada em *CRISPR-Cas* (*Clustered Regulatory Interspaced Short Palindromic Repeats* - Repetições Palindrômicas Curtas

Agrupadas Regularmente-Interespaçadas) é considerado uma nova tecnologia

emergente que desperta muito interesse na área científica (TOBITA et al., 2015; AREND et al., 2016).

Enzimas do complexo *CRISPR/Cas*

A engenharia genômica apresentou significativos avanços a partir do reconhecimento e uso do sistema *CRISPR-Cas* (WANG, 2017). Atualmente, esta técnica de biologia molecular é utilizada em muitos estudos científicos para a edição de genes (PASSOS et al., 2016).

A sigla *CRISPR* (*Clustered Regulatory Interspaced Short Palindromic Repeats* - 'Repetições Palindrômicas Curtas Agrupadas Regularmente- Interespaçadas'), representa um mecanismo de defesa de procariontes contra invasores, vírus e plasmídeos (BHAYA et al., 2011). Nesse sistema, o reconhecimento do material genético do invasor se concretiza por uma interação de emparelhamento de bases de um RNA guia com o DNA alvo exógeno, seguida pela atividade de enzimas nucleases (HEIDARI, 2015).

O sistema *CRISPR* possui associado ao seu locus gênico, dentre outros, os genes que codificam as enzimas *Cas*. Essas enzimas possuem função de helicase, polimerase e nuclease. Tais atividades também levam à hipótese de que o sistema *CRISPR/Cas* também está associado à reparação de DNA. A enzima *Cas* mais conhecida é a proteína *Cas9*, sendo essa também a mais utilizada (RICHTER et al., 2012).

O sistema *CRISPR* reconhece invasores por amostragem, com sequências curtas de DNA, chamadas de protoespaçadores. Essas sequências se juntam ao próprio material genético da bactéria, em um local específico, a matriz de sequências repetidas *CRISPR*. As sequências que incluídas na matriz passam a ser denominadas de espaçadores. Essas sequências quando transcritas em pequenos RNAs serão orientadas por proteínas *Cas* e outras proteínas associadas ao *CRISPR* (complexo proteico multifuncional), que reconhecerão e clivarão o material genético exógeno (MANGERICAO et al., 2016; MAI et al., 2016). A sequência ligante ao ser introduzida na matriz, é adicionada na extremidade proximal, junto à sequência líder, rica em A-T (Adenina e Timina) e tem aproximadamente 500pb (pares de base) de comprimento. Ao integrar-se as sequências espaçadoras numa matriz, ocorre a defesa mais rapidamente em caso de nova infecção da célula (WATSON et al., 2015).

O mecanismo do sistema *CRISPR* ocorre em 3 etapas: obtenção do material genético invasor, que será integrada ao DNA do hospedeiro no locus *CRISPR* (protoespaçador). Na segunda etapa, o RNA sintetizado, um longo pré-crRNA obtido será utilizado para produzir pequenos crRNAs funcionais. Os crRNAs funcionais permitem a identificação do DNA invasor para degradação por proteínas Cas (Figura 2) (MAIER et al., 2015).

O promotor da matriz *CRISPR* se encontra na região líder, sendo capaz de gerar apenas um único transcrito de RNA, denominado pré-crRNA. Em *E.colio* estudo desse sistema identificou a existência de uma variedade de genes Cas que participam do processo, sendo que seu complexo é chamado de *Cascade*. O complexo contém uma subunidade que processa um longo transcrito em crRNAs solitários e curtos, cada um desses produtos possui comprimento de um espaçador e uma sequência repetida. Esta sequência de RNAs se mantém ligada ao complexo e auxiliam no direcionamento para os genomas de DNA invasor (WATSON et al., 2015).

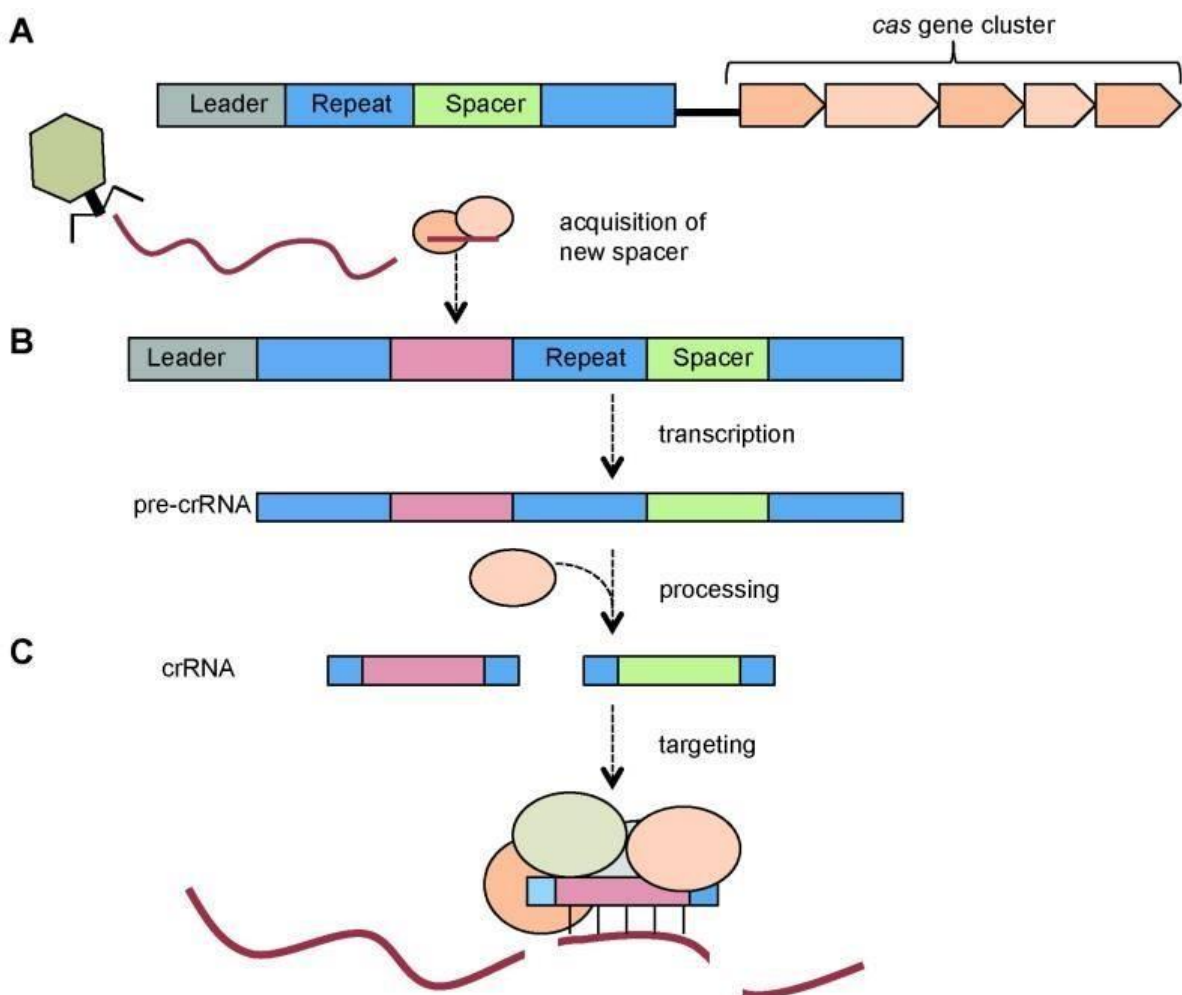


Figura 2 — *CRISPR*. *CRISPR* é formado por repetições curtas e sequências intervenientes provenientes dos invasores celulares. A célula bacteriana é invadida com um elemento estranho

(por exemplo, um fago), parte desse genoma será incorporado na extremidade líder dessa matriz, sendo possível duplicar esta sequência. (b) *CRISPR* é transcrito em pré-crRNAs, posteriormente são processados em crRNA maduro. O crRNA maduro junto com o complexo com a proteína CAS, se direciona para o elemento invasor semelhante ao anterior (RICHTER et al., 2012).

No sistema *CRISPR/Cas* tipo II, os elementos gênicos são transcritos em crRNA transativador (tracrRNA), proteína Cas9 e pré-crRNA. A ribonuclease III cliva pré-crRNA em crRNA nas proximidades das repetições diretas. Então forma o crRNA-tracrRNA-Cas9, tornando uma endonuclease guiada pelo crRNA ativo e direcionado à sequência alvo. O pareamento complementar de bases RNA-DNA é realizado e logo ocorre uma quebra da fita dupla por clivagem (GONG et al., 2020).

O sistema *CRISPR* foi classificado em 6 tipos, agrupados em duas classes. A classe 1 compreende os tipos I, III e IV; e a classe 2 compreende os tipos II, V e VI. No sistema *CRISPR* de classe 1 é necessário um complexo multiproteico para o reconhecimento e clivagem do DNA alvo. Esse complexo é chamado de *Cascade*, capaz de reconhecer o DNA alvo, que é clivado pela enzima Cas3. Para o sistema *CRISPR* classe II é necessário apenas uma endonuclease guiada por um RNA para a clivagem do DNA invasor. (MAKAROVA et al., 2011; WIEDENHEFT et al., 2011; MAKAROVA e KOONIN, 2015; WRIGHT et al., 2016; SHMAKOV et al., 2015; MAKAROVA et al., 2015).

O entusiasmo em relação ao sistema *CRISPR/Cas* está no seu poder de edição do DNA com especificidade e simplicidade, tendo forte potencial clínico em tratamento de doenças humanas (FINDLAY et al., 2014; GILBERT et al., 2014). Sua descoberta auxilia na validação de alvos de doenças (SHALEM et al., 2014; WANG et al., 2014), incremento de estudos epigenéticos (YAO et al., 2015; VORA et al., 20016), favorecimento da engenharia genética em plantas (RAITSKIN e PATRON, 2016; ZHANG et al., 2016). O *CRISPR/Cas* é uma ferramenta de edição gênica que já contribuiu para a terapia genética permitindo a descoberta de mais de 3000 genes que estão associados a mutação que causam doenças (COX et al., 2015).

OBJETIVOS

Objetivo geral

Prospectar enzimas do complexo CRISPR/Cas a partir de dados metagenômicos da microbiota intestinal do cupim *Syntermes wheeleri*.

Objetivos específicos

- Identificar a presença de sequências de proteínas do complexo *CRISPR/Cas* a partir de dados metagenômicos da microbiota intestinal do cupim *Syntermes wheeleri* depositados no banco de dados *ggKbase*;
- Avaliar a integridade das sequências do complexo *CRISPR/Cas* obtidas;
- Selecionar sequências de proteínas do complexo *CRISPR/Cas* que possam ser viáveis a caracterização.

METODOLOGIA

Sequências metagenômicas

Os dados utilizados nesse trabalho foram obtidos a partir de projeto anterior realizado por SANTANA, 2015. As amostras sequenciadas são provenientes da microbiota intestinal do cupim *Syntermes wheeleri*. Os dados estão depositados no banco de dados *ggKbase* (<https://ggkbase.berkeley.edu/>), gerido pela Universidade de Berkeley, Califórnia, USA. A plataforma *ggKbase* é *user-friendly* disponibilizada de forma *on-line* através de interface gráfica para usuários cadastrados. As sequências são armazenadas com anotação referente à sua nomenclatura, taxonomia e papel funcional. As análises foram realizadas com base em sequência de aminoácidos (AA).

Seleção de enzimas do complexo *CRISPR/Cas*

A partir da literatura científica foram identificados os genes presentes no sistema *CRISPR/Cas*. A seleção das enzimas de interesse ocorreu a partir de sistema de busca pelo nome de cada uma. As proteínas buscadas foram *Cas1*, *Cas2*, *Cas3*, *Cas5*, *Cas6*, *Cas9*, *Cascade - CasA*, *Cascade - CasC*, *Cascade - CasD*, *Cascade CasE*, *RAMP (Repeat Associated Mysterious Proteins)* *CasB*, *Cas7*, *Cas8*, *Cas10*, *Cas12*, *Cas13*, *Cas14*, *Cpf1*, *C2c1*, *C2c2* e *C2c3*.

Os arquivos com as sequências proteicas obtidas para cada gene passaram então por análises de critérios para exclusão de sequências. Os primeiros critérios utilizados para a exclusão de sequências foram a presença de códon de início (aminoácido metionina) e códon de parada da tradução. Após, foi realizada a análise de tamanho para cada proteína, através da consulta de proteínas caracterizadas experimentalmente no banco de dados público *Protein Data Bank (RCSB PDB)*. Proteínas com número menor de aminoácidos em relação às aquelas do *PDB* foram excluídas. As sequências remanescentes foram avaliadas pela ferramenta *BlastP* no disponibilizada pelo NCBI (*National Center for Biotechnology Information*) contra o banco de dados *PDB*. As sequências que não obtiveram ao menos uma proteína com similaridade no banco de dados foram excluídas. Posteriormente as sequências foram agrupadas entre si por alinhamento com a ferramenta *Muscle* (parâmetros *default*) dentro do *software Geneious*, para a eliminação de sequências com alta similaridade.

RESULTADOS E DISCUSSÃO

Ao todo foram obtidas 288 enzimas do complexo *CRISPR/Cas* no banco de dados do microbioma intestinal de *Syntermes wheeleri*. Dessas, 127 apresentavam códon de início e parada da tradução. O tamanho de 110 enzimas foi compatível com o tamanho de proteínas já caracterizadas experimentalmente. No total, 105 sequências apresentaram ao menos uma sequência de AA com identidade no banco de dados *PDB*. Ao final, foram selecionadas 66 proteínas que tem potencial para futuras sínteses por expressão heteróloga e caracterização estrutural e enzimática (tabela 1).

As enzimas *Cas1*, *Cas2*, *Cas3*, *Cas5*, *Cas6*, *Cas9*, *Cascade - CasA*, *Cascade - CasC*, *Cascade - CasD*, *Cascade CasE*, o domínio nuclease *RuvC* e proteínas *RAMP* foram identificadas nas sequências metagenômicas analisadas. Não foram encontradas as proteínas *CasB*, *Cas7*, *Cas8*, *Cas10*, *Cas12*, *Cas13*, *Cas14*, *Cpf1*, *C2c1*, *C2c2* e *C2c3*.

As proteínas *RAMP* (Repeat Associated Mysterious Proteins) formam um complexo que atua juntamente com as proteínas *Cas* no reconhecimento e clivagem do DNA/RNA exógeno. As proteínas *CasC*, *CasD* e *CasE* são consideradas proteínas do tipo *RAMP*. A busca por ambos os termos foi realizada para ampliar o número de sequências, considerando que podem ocorrer variações durante a anotação funcional das sequências. As proteínas *RAMP* possuem como característica peculiar uma dobra em sua estrutura que ajuda no reconhecimento do RNA invasor (WANG e LI, 2012).

RuvC é um domínio nuclease da proteína *Cas9* e está envolvido na clivagem da sequência alvo (NISHIMASU et al., 2014; ANDERS et al., 2014). Sua ação consiste em cortar a fita de DNA complementar à fita identificada pelo RNA guia (JINEK et al., 2012). A proteína *Cas9* é uma das enzimas *Cas* mais conhecidas e estudadas. A busca também pelo domínio funcional *RuvC*, além de *Cas9*, foi realizada pois a avaliação de possíveis variações nas sequências de aminoácidos e consequente estrutura desse domínio pode auxiliar o desenho de moléculas de maior eficiência ou precisão.

Tabela 1. Número de sequências únicas de proteínas após cada etapa de exclusão durante o processo de avaliação de qualidade. Dados obtidos a partir de metagenomas da microbiota intestinal do cupim *Syntermes wheelerii*, sequenciados em plataforma NGS - Illumina.

	Encontradas nos Metagenomas (busca por nome)	Apresentam <i>Start codon</i>	Apresentam <i>Stop codon</i>	Tamanho Mínimo estipulado <i>PDB</i> / número sequências	<i>BlastP</i>	Alinhamento Sequências
<i>Cas1</i>	54	39	29	>90 AA - 27	26	13
<i>Cas2</i>	38	32	27	>80 AA - 27	27	18
<i>Cas3</i>	43	14	13	>200 AA - 7	7	3
<i>Cas5</i>	30	23	20	>200 AA - 18	18	11
<i>Cas6</i>	4	1	1	>200 AA - 0	0	0
<i>Cas9</i>	23	11	4	>900 AA - 2	2	2
<i>RuvC</i>	49	10	8	> 150 AA - 7	7	5
<i>Cascade CasA</i>	5	1	1	>500 AA - 1	1	1
<i>Cascade CasC</i>	2	0	1	>200 AA - 0	0	0
<i>Cascade CasD</i>	7	7	7	>150 AA - 7	7	4
<i>Cascade CasE</i>	1	1	1	>200 AA - 0	0	0
<i>RAMP</i>	32	19	15	>200 AA - 14	10	7
Total de Sequências	288	158	127	0	105	64

*Não estão presentes *CasB*, *Cas7*, *Cas8*, *Cas10*, *Cas12*, *Cas13*, *Cas14*, *Cpf1*, *C2c1*, *C2c2*, *C2c3*

Existem duas classes de *CRISPR/Cas*; a primeira (Classe 1) que reúne os tipos I, III e IV, que necessitam formar um complexo proteico para a atividade nuclease; a segunda (Classe 2) inclui os tipos II, V e VI (HILLARY & CEASAR, 2022; ZHENG et al., 2020) e não dependem da formação de complexo para a atividade nuclease.

O complexo *Cascade* é um *cluster* de genes *Cas* (*CasA*, *CasB*, *CasC*, *CasD*, *CasE*) (BROUNS et al., 2008). A enzima *CasB* não foi identificada dentre as sequências avaliadas. É possível que seja apenas uma questão técnica, ocorrida durante sequenciamento ou anotação. Porém levanta-se também a possibilidade de que outras proteínas, do tipo *RAMP* por exemplo, cumpram a função de *CasB* no complexo proteico com atividade nuclease naquela microbiota específica. Em termos de aplicação biotecnológica, testar diretamente a atividade nuclease dessas enzimas com as sequências obtidas nesse trabalho pode não ser viável, uma vez que não foram encontradas a enzima *CasB*, e as enzimas *CasC* e *CasE* não atenderam ao padrão de qualidade estabelecido (códon de início e parada, e tamanho; respectivamente).

As enzimas que fazem parte do complexo *CRISPR/Cas* de classe 2 apresentam atividade nuclease independente de formação de complexo proteico. A enzima *Cas9* é um de seus exemplos mais conhecidos. Duas sequências de *Cas9* foram identificadas na microbiota analisada. Além disso, vários domínios *RuvC* também foram encontrados e atendem igualmente aos critérios de qualidade previamente estabelecidos. Somados, esses fatos indicam que essa biblioteca de sequências pode ser utilizada para a caracterização experimental dessas proteínas. A avaliação do efeito da modificação ou troca entre os diferentes domínios funcionais *RuvC* também é uma das possibilidades de testes a serem feitos com esse banco de sequências.

As enzimas *Cas1* e *Cas2* estão relacionadas ao processo de clivagem do material genético do invasor celular durante o processo de aquisição dos protoespaçadores do sistema *CRISPR/Cas*, para posterior incorporação no DNA do hospedeiro (HILLARY & CEASAR, 2022). Ao final da análise de qualidade, *Cas1* e *Cas2* foram as enzimas do sistema *CRISPR/Cas* mais abundantes nos metagenomas da microbiota intestinal do cupim *Syntermes wheeleri*. Além da função dessas proteínas para o mecanismo de defesa celular dos

microrganismos, é possível que elas também estejam relacionadas à contribuição da microbiota intestinal para a imunidade do hospedeiro. Essas proteínas da base de dados analisada podem ser expressas e testadas quanto à sua especificidade para diferentes alvos, principalmente virais.

Durante o processo biotecnológico de expressão heteróloga de proteínas para a produção em escala uma das principais dificuldades encontradas pelos pesquisadores é a secreção das proteínas produzidas pelo organismo hospedeiro produtor. Kruasuwan et al (2021) utilizou a edição genética da levedura *Ogataea thermomethanolica* (TBRC 656) com *CRISPR/Cas9* para alterar os níveis de secreção de xilanases e fitases produzidas pela levedura. Nesse sentido, as sequências *CRISPR/Cas* analisadas nesse trabalho podem ser testadas e utilizadas nos processos de expressão heteróloga de outras enzimas da microbiota de *Syntermes wheeleri*, que já estão em andamento no grupo de pesquisa.

PERSPECTIVAS

Esse trabalho apresenta a etapa inicial de seleção de sequências únicas de enzimas do complexo *CRISPR/Cas* a partir de dados metagenômicos da microbiota intestinal do cupim *Syntermes wheeleri*. As enzimas aqui prospectadas devem passar por etapa complementar de análise *in silico* para avaliação da sua similaridade linear, estrutural e funcional com outras proteínas já descritas em bancos de dados curados. As etapas posteriores de caracterização e viabilidade de uso biotecnológico demandam longo trabalho experimental que envolve a escolha de hospedeiros e condições para expressão. O investimento de recursos é necessário para que as enzimas aqui apresentadas possam tornar-se recursos biotecnológicos viáveis.

CONCLUSÃO

A avaliação de metagenomas da microbiota associada ao intestino do cupim *Syntermes wheeleri* revelou a presença de 66 sequências de enzimas do complexo *CRISPR/Cas* que possivelmente podem ser expressas, caracterizadas e avaliadas quanto ao seu potencial uso biotecnológico, para desenvolvimento da sua possível aplicação tecnológica. Diferentes tipos e classes de proteínas *Cas* e *RAMP* foram identificadas. Análises posteriores são necessárias para a definição das sequências prioritárias para o início da fase experimental de caracterização das proteínas. A prospecção de enzimas a partir de dados ambientais e metagenômicos constitui-se como uma das formas de aproveitamento biotecnológico da biodiversidade.

REFERÊNCIAS

ANDERS, Carolin et al. Structural basis of PAM-dependent target DNA recognition by the Cas9 endonuclease. **Nature**, v. 513, n. 7519, p. 569-573, 2014.

AREND, Marcela Corso; PEREIRA, Jessica Olivaes; MARKOSKI, Melissa Medeiros. O Sistema CRISPR/Cas9 e a possibilidade de edição genômica para a cardiologia. **Arquivos Brasileiros de Cardiologia**, v. 108, p. 81-83, 2017.

BAYM, Michael; KRYAZHIMSKIY, Sergey; LIEBERMAN, Tami D.; CHUNG, Hattie; DESAI, Michael M.; KISHONY, ROY. Inexpensive multiplexed library preparation for megabase-sized genomes. **PloS one**, v. 10, n. 5, p. e0128036, 2015.

BELL, Thomas; NEWMAN, J. A.; SILVERMAN, B. W.; TURNER, S. I.; LILEY, A. K. The contribution of species richness and composition to bacterial services. **Nature**, v. 436, n. 7054, p. 1157-1160, 2005.

BERVOETS, Indra; CHARLIER, Daniel. Diversity, versatility and complexity of bacterial gene regulation mechanisms: opportunities and drawbacks for applications in synthetic biology. **FEMS microbiology reviews**, v. 43, n. 3, p. 304-339, 2019.

BHAYA, D.; DAVISON, M.; BARRANGOU, R. CRISPR-Cas systems in bacteria and archaea: versatile small RNAs for adaptive defense and regulation. *Annual review of genetics*. 2011 Dec 15;45(1):273-97.

BISTA, Iliana et al. Performance of amplicon and shotgun sequencing for accurate biomass estimation in invertebrate community samples. **Molecular Ecology Resources**, v. 18, n. 5, p. 1020-1034, 2018.

BLAYA, Josefa; MARHUENDA, Frutos C.; PASCUAL, Jose A.; ROS, Margarita. Microbiota characterization of compost using omics approaches opens new perspectives for Phytophthora root rot control. **PLoS One**, v. 11, n. 8, p. e0158048, 2016.

BLOCKLEY, Alix; ELLIOTT, David R.; ROBERTS, Adam P.; SWEET, Michael. Symbiotic microbes from marine invertebrates: driving a new era of natural product drug discovery. **Diversity**, v. 9, n. 4, p. 49, 2017.

BROUNS, Stan JJ et al. Small CRISPR RNAs guide antiviral defense in prokaryotes. **Science**, v. 321, n. 5891, p. 960-964, 2008.

BULL, Alan T .; Bunch, Alan W .; ROBINSON, Gary K. Biocatalisadores para produtos e processos industriais limpos. Opinião atual em microbiologia , v. 2, n. 3, p. 246-251, 1999.

CALVETE, Crislaine Lambiase; CASEIRO, Marcos Montani; DE SOUZA, Cleide Barbieri. BIOTECNOLOGIA: TRANSFORMAÇÃO BACTERIANA POR MÉTODO DE CHOQUE TÉRMICO. **UNILUS Ensino e Pesquisa**, v. 12, n. 26, p. 41-53, 2015.

CAMERON, E. K.; MARTINS, Inês S.; LAVELLE, Patrick; MATHIEU, Jérôme; TEDERSOO, Leho; GOTTSCHALL, Felix; GUERRA, Carlos A.; HINES, Jes; PATOINE, Guillaume; SIEBERT, Julia; WINTER, Marten; CESARZ, Simone; DELGADO-BAQUERIZO, Manuel; FERLIAN, Olga; FIERE, Noah; KREFT, Holger; LOVEJOY, Thomas E.; MONTANARELLA, Luca; ORGIAZZI, Alberto; PEREIRA, Henrique M.; PHILLIPS, Hellen R.P.; SETTELE, Josef; WALL, Diana H.; EISENHAUER, Nico. Global gaps in soil biodiversity data. *Nature Ecology and Evolution* 2, 1042-1043. 2018.

CARBALLO, J. Biotecnología e alimentos transxénicos. *Ciencia y Tecnología Alimentaria, Sociedad Mexicana de Nutrición y Tecnología de alimentos*, v. 3, n. 5, p. 314-321, 2002.

CHIRIAC, Luiza Silvia; MURARIU, Dumitru T. Application of metagenomics in ecology: a brief overview. **Current Trends in Natural Sciences**, v. 10, n. 19, p. 346-353, 2021.

COHEN, Louis J.; KANG, Hahk-Soo; CHU, John; HUANG, Yun-Han; GORDON, Emma A.; REDDY, Boojala Vijay B.; TERNEI, Melinda A.; CRAIG, Jeffrey W.; BRADY, Sean F. Functional metagenomic discovery of bacterial effectors in the human microbiome and isolation of commendamide, a GPCR G2A/132 agonist. **Proceedings of the National Academy of Sciences**, v. 112, n. 35, p. E4825-E4834, 2015.

COLIN, Pierre-Yves; KINTSES, Balint; GIELEN, Fabrice; MITON, Charlotte M.; FISCHER, Gerhard; MOHAMED, Mark F.; HYVÖNEN, Marko; MORGAVI, Diego P.; JANSSEN, Dick B.; HOLLFELDER, Florian. Ultrahigh-throughput discovery of promiscuous enzymes by picodroplet functional metagenomics. **Nature communications**, v. 6, n. 1, p. 1-12, 2015.

COLLET, Jean-Francois; BARDWELL, James CA. Oxidative protein folding in bacteria. **Molecular microbiology**, v. 44, n. 1, p. 1-8, 2002.

CONVENTION ON BIOLOGICAL DIVERSITY (CBD). [s.l.: s.n.], 1992. Disponível em: <https://www.cbd.int/doc/legal/cbd-en.pdf>; Acesso em: 17 nov. 2020.

COTTA, Simone Raposo. O solo como ambiente para a vida microbiana. **Microbiologia do solo**, p. 23, 2016.

COX, D.B.T., PLATT, R.J.; ZHANG, F. Therapeutic genome editing: prospects and challenges. *Nature medicine*, 21(2), pp.121-131. 2015.

CRUMP, Byron C.; HOBBI, John E. Synchrony and seasonality in bacterioplankton communities of two temperate rivers. **Limnology and Oceanography**, v. 50, n. 6, p. 1718-1729, 2005.

DAVIES, Julian. Specialized microbial metabolites: functions and origins. **The Journal of antibiotics**, v. 66, n. 7, p. 361-364, 2013.

DELGADO-BAQUERIZO, M.; OLIVERIO, A. M.; BREWER, T. E.; BENAVENTGONZÁLEZ, A.; ELDRIDGE, D. J.; BARDGETT, R. D.; MAESTRE, F. T.; SINGH, B. K.; FIERER, N. A global atlas of the dominant bacteria found in soil. *Science*, 2018. v. 359, n. 6373, p. 320-325.

DICK, Gregory. **Genomic approaches in earth and environmental sciences**. John Wiley & Sons, 2018.

DINI-ANDREOTE, Francisco; VAN ELSAS, Jan Dirk. The soil microbiome—An overview. **Modern Soil Microbiology**, p. 37-48, 2019.

DOS SANTOS, Vinicius Sartor; WIETHÖLTER, Paula. Contribuições da engenharia genética no tratamento de doenças. **Brazilian Journal of Development**, v. 7, n. 3, p. 31157-31176, 2021.

ENAGBONMA, Ben Jesuorsemwon; BABALOLA, Olubukola Oluranti. Environmental sustainability: A review of termite mound soil material and its bacteria. **Sustainability**, v. 11, n. 14, p. 3847, 2019.

ENAGBONMA, Ben Jesuorsemwon; AJILOGBA, Caroline Fadeke; BABALOLA, Olubukola Oluranti. Metagenomic profiling of bacterial diversity and community structure in termite mounds and surrounding soils. **Archives of Microbiology**, v. 202, n. 10, p. 2697-2709, 2020.

FIERER, Noah; LAUBER, Christian L.; RAMIREZ, Kelly S.; ZANEVELD, Jesse; BRADFORD, Mark A.; KNIGHT, Rob. Comparative metagenomic, phylogenetic and physiological analyses of soil microbial communities across nitrogen gradients. **The ISME journal**, v. 6, n. 5, p. 1007-1017, 2012.

FIGUEIREDO, LHM; PENTEADO, MIO; MEDEIROS, P. T. Patentes em Biotecnologia. *Biotecnologia Ciência & Desenvolvimento*. Brasília/DF, n. 36, p. 32-39, 2006.

FINDLAY, G.M.; BOYLE, E.A.; HAUSE, R.J.; KLEIN, J.C.; SHENDURE, J. Edição de saturação de regiões genômicas por reparo dirigido por homologia multiplex. *Natureza*. 2014 set;513(7516):120-3.

FLINT, Harry J. et al. The role of the gut microbiota in nutrition and health. **Nature reviews Gastroenterology & hepatology**, v. 9, n. 10, p. 577-589, 2012.

FORSBERG, Kevin J.; PATEL, Sanket; WITT, Evan; WANG, Bin; ELLISON, Tyler D.; DANTAS, Gautam. Identification of genes conferring tolerance to lignocellulose-derived inhibitors by functional selections in soil metagenomes. **Applied and environmental microbiology**, v. 82, n. 2, p. 528-537, 2016.

FORTUNATO, Caroline S.; EILER, Alexander; HERFORT, Lydie; NEEDOBA, Joseph A.; PETERSON, Tawnya D.; CRUMP, Byron C. Determining indicator taxa across spatial and seasonal gradients in the Columbia River coastal margin. **The ISME Journal**, v. 7, n. 10, p. 1899-1911, 2013.

FREY, K.G.; HERRERA-GALEANO, J.E.; REDDEN, C.L.; LUU, T.V.; SERVETAS, S.L.; MATECZUN, A.J.; MOKASHI, VISHWESH, P.; BISHOP-LILLY, Kimberly A. Comparison of three next-generation sequencing platforms for metagenomic sequencing and identification of pathogens in blood. **BMC genomics**, v. 15, n. 1, p. 1-14, 2014.

GAGNIÈRE, Johan; RAISCH, Jennifer; VEZIAN, Julie; BARNICH, Nicolas; BONNET, Richard; BUC, Emmanuel; BRINGER, Marie-agnès; PEZET, Denis; BONNET, Mathilde. Gut microbiota imbalance and colorectal cancer. **World journal of gastroenterology**, v. 22, n. 2, p. 501, 2016.

GEDOZ, Milena et al. Edaphic invertebrates as indicators of soil integrity quality. **Floresta e Ambiente**, v. 28, 2021.

Gilbert LA, Horlbeck MA, Adamson B, Villalta JE, Chen Y, Whitehead EH, Guimaraes C, Panning B, Ploegh HL, Bassik MC, Qi LS. Genome-scale CRISPR-mediated control of gene repression and activation. *Cell*. 2014 Oct 23;159(3):647-61.

GOMES, R. N.; LIMA, P. S.; KURIYAMA, S. N.; NETO, A. A. F. Desenvolvimento da química verde no cenário industrial brasileiro. *Revista Fitos*, p. 80-89, 2018.

GABER, Yasser et al. Heterologous expression of lytic polysaccharide monoxygenases (LPMOs). **Biotechnology advances**, v. 43, p. 107583, 2020.

GONG, Tao; ZENG, Jumei; BOYU, Tang; ZHOU, Xuedong; LI, Yuqing. CRISPR-Cas systems in oral microbiome: From immune defense to physiological regulation. **Molecular Oral Microbiology**, v. 35, n. 2, p. 41-48, 2020.

GRIFFITHS, Hannah M.; ASHOTON, Louise A.; PARR, Catherine L.; EGGLETON, Paul. The impact of invertebrate decomposers on plants and soil. **New Phytologist**, v. 231, n. 6, p. 2142-2149, 2021.

GUO, Jianhua; LI, Jie; CHEN, Hui; BOND, Philip L.; YUAN, Zhiguo. Metagenomic analysis reveals wastewater treatment plants as hotspots of antibiotic resistance genes and mobile genetic elements. **Water research**, v. 123, p. 468-478, 2017.

HANDELSMAN, Jo. Metagenomics: application of genomics to uncultured microorganisms. **Microbiology and molecular biology reviews**, v. 68, n. 4, p. 669-685, 2004.

HEIDARI, Raheleh; SHAW, David Martin; ELGER, Bernice Simone. CRISPR and the rebirth of synthetic biology. **Science and engineering ethics**, v. 23, n. 2, p. 351-363, 2017.

Hillary, V.E., Ceasar, S.A. A Review on the Mechanism and Applications of CRISPR/Cas9/Cas12/Cas13/Cas14 Proteins Utilized for Genome Engineering. **Mol Biotechnol**, 2022.

HOCHSTRASSER, Megan L.; et al. CasA mediates Cas3-catalyzed target degradation during CRISPR RNA-guided interference. **Proceedings of the National Academy of Sciences**, v. 111, n. 18, p. 6618-6623, 2014.

HOCKEMEYER, Dirk; WANG, Hayoi; KIANI, Samira; LAI, Christine S.; GAO, Qing; CASSADY, John P.; COST, Gregory J.; ZHANG, Lei; SANTIAGO, Yolanda; MILLER, Jeffrey C.; ZEITLER, Bryan; CHERONE, Jennifer M.; MENG, Xiangdong; HINKLEY, REBAR, Edward J.; GREGORY, Philip D.; URNOV, Fyodor D.; JAENISCH, Rudolf. Genetic engineering of human pluripotent cells using TALE nucleases. **Nature biotechnology**, v. 29, n. 8, p. 731-734, 2011. HANDELSMAN, Jo; RONDON, Michele R.; BRADY, SEAN F.; CLARDY, Jon; GOODMAN, Robert M. Molecular biological access to the chemistry of unknown soil microbes: a new frontier for natural products. **Chemistry & biology**, v. 5, n. 10, p. R245-R249, 1998.

HU, Honghong; XIONG, Lizhong. Genetic engineering and breeding of drought-resistant crops. **Annual review of plant biology**, v. 65, p. 715-741, 2014.

HUANG, Chung-Jr; LIN, Henry; YANG, Xiaoming. Industrial production of recombinant therapeutics in *Escherichia coli* and its recent advancements. **Journal of Industrial Microbiology and Biotechnology**, v. 39, n. 3, p. 383-399, 2012.

JINEK, Martin et al. A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity. **science**, v. 337, n. 6096, p. 816-821, 2012.

JORE, Matthijs M. et al. Structural basis for CRISPR RNA-guided DNA recognition by Cascade. **Nature structural & molecular biology**, v. 18, n. 5, p. 529-536, 2011.

JOYNSON, Ryan et al. Metagenomic analysis of the gut microbiome of the common black slug *Arion ater* in search of novel lignocellulose degrading enzymes. **Frontiers in microbiology**, v. 8, p. 2181, 2017.

JULLESSON, David et al. Impact of synthetic biology and metabolic engineering on industrial production of fine chemicals. **Biotechnology advances**, v. 33, n. 7, p. 1395-1402, 2015.

LIU, Ning et al. Functional metagenomics reveals abundant polysaccharide-degrading gene clusters and cellobiose utilization pathways within gut microbiota of a wood-feeding higher termite. **The ISME Journal**, v. 13, n. 1, p. 104-117, 2019.

LIMA, Raul Alcântara Teixeira et al. Functional and structural characterization of a novel GH3 β -glucosidase from the gut metagenome of the Brazilian Cerrado termite *Syntermes wheeleri*. **International Journal of Biological Macromolecules**, v. 165, p. 822-834, 2020.

LUTZ, Stefanie; THUERIG, Barbara; OBERHAENSLI, Thomas; MAYERHOFER, Johanna; FUCHS, Jacques G.; WIDMER, Franco; FREIMOSER, Florian M.; AHRENS, Christian H. Harnessing the microbiomes of suppressive composts for plant protection: From metagenomes to beneficial microorganisms and reliable diagnostics. **Frontiers in Microbiology**, v. 11, p. 1810, 2020

JAY, J. M. Modern food microbiology. 6. ed., Aspen Publition, 2000.

JUNIOR, Nilo Luiz Saccaro. Desafios da bioprospecção no Brasil. Texto para Discussão, 2011.

KAN, Jinjun; CRUMP, Byron C.; WANG, Kui; CHEN, Feng. Bacterioplankton community in Chesapeake Bay: predictable or random assemblages. **Limnology and Oceanography**, v. 51, n. 5, p. 2157-2169, 2006.

KENNEDY, Jonathan; MARCHESI, Julian R.; DOBSON, Alan DW. Metagenomic approaches to exploit the biotechnological potential of the microbial consortia of marine sponges. **Applied microbiology and biotechnology**, v. 75, n. 1, p. 11-20, 2007.

KHAN, Md. Aslam; AHMAD, Wasim; PAUL, Bishwajeet. Ecological impacts of termites. In: **Termites and Sustainable Management**. Springer, Cham, 2018. p. 201-216.

KIRIGA, Waititu J.; YU, Qingyi; BILL, Ratemo. Breeding and genetic engineering of drought-resistant crops. **International Journal of Agriculture and Crop Sciences**, v. 9, n. 1, p. 7, 2016.

KIRSE, Ameli et al. Unearthing the potential of Soil eDNA metabarcoding—towards best practice advice for invertebrate biodiversity assessment. **Frontiers in Ecology and Evolution**, p. 337, 2021.

KRUASUWAN, Worarat et al. Modulation of heterologous protein secretion in the thermotolerant methylotrophic yeast *Ogataea thermomethanolica* TBRC 656 by CRISPR-Cas9 system. **PloS one**, v. 16, n. 9, p. e0258005, 2021.

LANGILLE, Morgan G.I.; ZANEVELD, Jesse; CAPORASO, J. GREGORY; McDONALD, Daniel; KNIGHTS, Dan; REYES, Joshua A.; CLEMENTE, Jose C.; BRUKEPILE, Deron E.; THURBER, Rebecca L. Vega; KNIGHT, Rob; BEIKO, Robert G.; HUTTENHOWER, Curtis. Predictive functional profiling of microbial communities using 16S rRNA marker gene sequences. **Nature biotechnology**, v. 31, n. 9, p. 814-821, 2013.

LEY, Ruth E.; TURNBAUGH, Peter J.; KLEIN, Samuel; GORDON, Jeffrey I. Human gut microbes associated with obesity. **nature**, v. 444, n. 7122, p. 1022-1023, 2006.

LI, An-Dong; LI, Li-Guan; ZHANG, Tong. Exploring antibiotic resistance genes and metal resistance genes in plasmid metagenomes from wastewater treatment plants. **Frontiers in microbiology**, v. 6, p. 1025, 2015.

LIMA, Raul Alcântara Teixeira. Caracterização bioquímica e biofísica de uma β -glicosidase obtida do metagenoma de intestino de *Syntermes wheeleri*. 2019.

LIMA, Raul Alcântara Teixeira. Caracterização bioquímica e biofísica de uma β -glicosidase obtida do metagenoma de intestino de *Syntermes wheeleri*. 2019.

MAEDER, Morgan L.; GERSBACH, Charles A. Genome-editing technologies for gene and cell therapy. **Molecular Therapy**, v. 24, n. 3, p. 430-446, 2016.

MAIER, Lisa-Katharina; STACHLER, Aris-Edda; SAUNDERS, Sita J.; BACKOFEN, Rolf; MARCHFELDER, Anita. An active immune defense with a minimal CRISPR (clustered

regularly interspaced short palindromic repeats) RNA and without the Cas6 protein. **Journal of Biological Chemistry**, v. 290, n. 7, p. 4192-4201, 2015.

MAKAROVA, K.S.; HAFT, D.H.; BARRANGOU, R.; BROUNS, S.J.; CHARPENTIER, E.; HORVATH, P.; MOINEAU, S.; MOJICA, F.J.; WOLF, Y.I.; YAKUNIN, A.F.; VAN, Der Oost J. Evolution and classification of the CRISPR-Cas systems. *Nature Reviews Microbiology*. 2011 Jun;9(6):467-77.

MAKAROVA, K.S.; KOONIN, E.V. Annotation and classification of CRISPR-Cas systems. *CRISPR*. 2015:47-75.

MAKAROVA, K.S.; WOLF, Y.I.; ALKHNABASHI, O.S.; COSTA, F.; SHAH, S.A.; SAUNDERS, S.J.; BARRANGOU, R.; BROUNS, S.J.; CHARPENTIER, E.; HAFT, D.H.; HORVATH, P. An updated evolutionary classification of CRISPR-Cas systems. *Nature Reviews Microbiology*. 2015 Nov;13(11):722-36.

MAKAROVA, Kira S. et al. Evolutionary classification of CRISPR-Cas systems: a burst of class 2 and derived variants. **Nature Reviews Microbiology**, v. 18, n. 2, p. 67-83, 2020.

MANGERICO, Tatiana C.; PENG, Zhanhao; ZHANG, Xuegong. Computational prediction of CRISPR cassettes in gut metagenome samples from Chinese type-2 diabetic patients and healthy controls. In: **BMC systems biology**. BioMed Central, 2016. p. 81-87.

MARANDUBA, Carlos Magno da Costa et al. Intestinal microbiota as modulators of the immune system and neuroimmune system: impact on the host health and homeostasis. **Journal of immunology research**, v. 2015, 2015.

MEDEMA, Marnix H.; DE ROND, Tristan; MOORE, Bradley S. Mining genomes to illuminate the specialized chemistry of life. **Nature Reviews Genetics**, v. 22, n. 9, p. 553-571, 2021.

MORAN, Mary Ann. The global ocean microbiome. **Science**, v. 350, n. 6266, p. aac8455, 2015.

MOREIRA, Fátima; SIQUEIRA, José Oswaldo. **Microbiologia e Bioquímica**. Editora Ufla, 2006.

MOUNCEY, Nigel J.; OTANI, Hiroshi; UDWARAY, Daniel; YOSHIKUNI, Yasuo. New voyages to explore the natural product galaxy. **Journal of Industrial Microbiology and Biotechnology**, v. 46, n. 3-4, p. 273-279, 2019.

MORSHED, M. N.; BEHARY, N.; BOUAZIZI, N.; JINPING, G. U. A. N.; NIERSTRASZ, V. A. An overview on biocatalysts immobilization on textiles: preparation, progress and application in wastewater treatment. *Chemosphere*, v. 279, p. 13081, 2021.

MUJTAR, V. E. L.; MUÑOZ, N.; PRACK M. C.; CORMICK, B.; PULLEMAN, M.; TITTONELL, P. Role and management of soil biodiversity for food security and nutrition; where do we stand? *Global Food Security*, 1 mar. 2019. v. 20, n. April 2018, p. 132-144.

MUSSATTO, Solange I.; FERNANDES, Marcela; MILAGRES, Adriane MF. Enzimas-Poderosa ferramenta na indústria. **Ciência Hoje**, v. 41, p. 28-33, 2007.

MYER, Angela; FORSCHLER, Brian T. Evidence for the role of subterranean termites (*Reticulitermes* spp.) in temperate forest soil nutrient cycling. **Ecosystems**, v. 22, n. 3, p. 602-618, 2019.

NAVEENKUMAR, R.; BASKAR, G. Process optimization, green chemistry balance and techno-economic analysis of biodiesel production from castor oil using heterogeneous nanocatalyst. *Bioresource Technology*, v. 320, p. 124347, 2021.

NEWMAN, David J.; CRAGG, Gordon M. Natural products as sources of new drugs over the nearly four decades from 01/1981 to 09/2019. **Journal of natural products**, v. 83, n. 3, p. 770-803, 2020.

NISHIMASU, Hiroshi et al. Crystal structure of Cas9 in complex with guide RNA and target DNA. **Cell**, v. 156, n. 5, p. 935-949, 2014.

ORLANDELLI, Ravelly Casarotti; SPECIAN, Vânia; FELBER, Aretusa Cristina; PAMPHILE, João Alencar. Enzimas de interesse industrial: produção por fungos e aplicações. **SaBios-Revista de Saúde e Biologia**, v. 7, n. 3, 2012.

ORZETTI, Sabrina; TOMMASI, Federica; BERTOLA, Antonella; BORTOLIN, Giorgia; CACCIN, Elisabetta; CECCO, Sara; FERRARIN, Emanuela; GIACOMIN, Elisa; BALDO, Paolo. Genetic Therapy and Molecular Targeted Therapy in Oncology: Safety, Pharmacovigilance, and Perspectives for Research and Clinical Practice. **International Journal of Molecular Sciences**, v. 23, n. 6, p. 3012, 2022.

OUTILI, N.; HALIMA, K.; CHAHRA, N.; RAYAN, M.; HASSEN, M. A. Biodiesel production optimization from waste cooking oil using green chemistry metrics. *Renewable Energy*, v. 145, p. 2525-2586, 2020.

OVERTON, Tim W. Recombinant protein production in bacterial hosts. **Drug discovery today**, v. 19, n. 5, p. 590-601, 2014

PACHIADAKI, Maria G.; BROWN, Julia M.; BROWN, Joseph; BEZUIDT, Oliver; BERUBE, Paul M.; BILLER, Steven J.; POULTON, Nicole J.; BURKART, Michael D.; CLAIR, James J. La; CHISHOLM, Sallie W.; STEPANAUSKAS, Ramunas. Charting the complexity of the marine microbiome through single-cell genomics. **Cell**, v. 179, n. 7, p. 1623-1635. e11, 2019.

PASSOS, Geraldo Aleixo Silva; HERNANDEZ, Cesar Augusto Speck; SOUSA, Larissa Cotrim de; FELÍCIO, Rafaela de Freitas Martins; SOUZA, Tiago Alves Jorge de. Aplicações da técnica. **Introdução à técnica de CRISPR**, 2016.

PEREIRA, T. C. et al. Introdução à técnica CRISPR. Ribeirão Preto: Sociedade Brasileira de Genética, 2016.

PYE, Cameron R.; BERTIN, Matthew J.; LOKEY, R. Scott; GERWICK, William H.; LININGTON, Roger G. Retrospective analysis of natural products provides insights for future discovery trends. **Proceedings of the National Academy of Sciences**, v. 114, n. 22, p. 5601-5606, 2017.

RENGIFO, V. Modelo analógico para enseñar biotecnología: una experiencia en el aula. **Paradigma**, v. 30, n. 2, p. 34-46, 2009.

RAITSKIN, O.; PATRON, N.J. Multi-gene engineering in plants with RNA-guided Cas9 nuclease. *Current Opinion in Biotechnology*, 37, pp.69-75. 2016.

RASHAMUSE, Konanani et al. Metagenomic mining of glycoside hydrolases from the hindgut bacterial symbionts of a termite (*Trinervitermes trinervoides*) and the characterization of a multimodular β -1, 4-xylanase (GH11). **Biotechnology and Applied Biochemistry**, v. 64, n. 2, p. 174-186, 2017.

RICHTER, Corinna; CHANG, James T.; FINERAN, Peter C. Function and regulation of clustered regularly interspaced short palindromic repeats (CRISPR)/CRISPR associated (Cas) systems. **Viruses**, v. 4, n. 10, p. 2291-2311, 2012.

RICHTER, Daniel J. et al. Genomic evidence for global ocean plankton biogeography shaped by large-scale current systems. **Elife**, v. 11, p. e78129, 2022.

Robinson, P. K. Enzymes: principles and biotechnological applications. *Essays Biochem.* 2015.

ROBINSON, Serina L.; PIEL, Jörn; SUNAGAWA, Shinichi. A roadmap for metagenomic enzyme discovery. **Natural Product Reports**, v. 38, n. 11, p. 1994-2023, 2021.

ROSANO, Germán L.; CECCARELLI, Eduardo A. Recombinant protein expression in *Escherichia coli*: advances and challenges. **Frontiers in microbiology**, v. 5, p. 172, 2014.

SANTANA, Renata Henrique et al. The gut microbiota of workers of the litter-feeding termite *Syntermes wheeleri* (Termitidae: Syntermitinae): archaeal, bacterial, and fungal communities. **Microbial ecology**, v. 70, n. 2, p. 545-556, 2015.

SANTOS, S. F. M.; MACEDO, G. R. Aplicação da metodologia de superfície de resposta no estudo da produção e extração da poligalacturonase. **Química Nova**, v. 31, p. 1973-1978, 2008.

SCHENBERG, Ana Clara Guerrini. Biotecnologia e desenvolvimento sustentável. *Estudos avançados*, v. 24, n. 70, p. 07-17, 2010.

SCHLEPER, C.; DELONG, E. F.; PRESTON, C. M.; FELDMAN, R. A.; WU, K. Y.; SWANSON, R. V. Genomic analysis reveals chromosomal variation in natural populations of the uncultured psychrophilic archaeon *Cenarchaeum symbiosum*. **Journal of Bacteriology**, v. 180, n. 19, p. 5003-5009, 1998.

SCHMID, Andreas; HOLLMANN, Frank; PARK, Jin Byung; BÜHLER, Bruno. The use of enzymes in the chemical industry in Europe. **Current Opinion in Biotechnology**, v. 13, n. 4, p. 359-366, 2002.

SHALEM, O.; SANJANA, N.E.; HARTENIAN, E.; SHI, X.; SCOTT, D.A.; MIKKELSEN, T.S.; HECKL, D.; EBERT, B.L.; ROOT, D.E.; DOENCH, J.G.; ZHANG, F. Genome-scale CRISPR-Cas9 knockout screening in human cells. *Science*. 2014 Jan 3;343(6166):84-7.

SHANBHAG, Rashmi R.; HARIT, A.; CHEIK, S.; CHAUDHARY, E.; BOTTINELLI, N.; SUNDARARAJ, R.; JOUQUET, P. Litter quality affects termite sheeting production and water infiltration in the soil. **Sociobiology**, v. 66, n. 3, p. 491-499, 2019.

SILVA, Patricia Siqueira; JUNIOR, Cyro Teixeira da Silva; ARAÚJO, Elizabeth Giestal de; KANAAN, Salim; XAVIER, Analúcia Rampazzo. Adenosina desaminase: uma enzima extraordinária e onipresente. **Pulmão RJ**, v. 25, n. 1, p. 11-16, 2016.

SOMMER, Morten O. A.; CHURCH, George M.; DANTAS, Gautam. A functional metagenomic approach for expanding the synthetic biology toolbox for biomass conversion. **Molecular systems biology**, v. 6, n. 1, p. 360, 2010.

SOUZA, L. T.; VERÍSSIMO, L. A.; PESSELA, B. C.; SANTORO, R. R.; RESENDE, R. R.; MENDES, A. A. Imobilização enzimática: princípios fundamentais e tipos de

suporte. In: Resende, R. R. *Biotecnologia aplicada à agro&indústria: fundamentos e aplicações*, v. 4. Editora Blucher, 2017.

STALEY, James T.; KONOPKA, Allan. Measurement of in situ activities of nonphotosynthetic microorganisms in aquatic and terrestrial habitats. **Annual review of microbiology**, v. 39, n. 1, p. 321-346, 1985.

STALEY, Christopher; GOULD, Trevor J.; WANG, Ping; PHILLIPS, Jane; COTNER, James B.; SADOWSKY, Michael J. Core functional traits of bacterial communities in the Upper Mississippi River show limited variation in response to land cover. **Frontiers in microbiology**, v. 5, p. 414, 2014.

SHMAKOV, S.; ABUDAYYEH, O.O.; MAKAROVA, K.S.; WOLF, Y.I.; GOOTENBERG, J.S.; SEMENOVA, E.; MINAKHIN, L.; JOUNG, J.; KONERMANN, S.; SEVERINOV, K.; ZHANG, F. Discovery and functional characterization of diverse class 2 CRISPR-Cas systems. *Molecular cell*. 2015 Nov 5;60(3):385-97.

SRIVASTAVA, Neha. Production of food-processing enzymes from recombinant microorganisms. In: **Enzymes in food biotechnology**. Academic Press, 2019. p. 739-767.

TEIXEIRA, I. S.; MILAGRE, C. D. Evolução dirigida de enzimas: pequenas modificações, melhores biocatalisadores. *Química Nova*, v. 43, n. 6, p. 773-786, 2020.

TRINDADE, Marla; ZYL, Leonardo Joaquim Van; NAVARRO-FERNÁNDEZ, José; ELRAZAK, Ahmed Abd. Targeted metagenomics as a tool to tap into marine natural product diversity for the discovery and production of drug candidates. **Frontiers in microbiology**, v. 6, p. 890, 2015.

TOBITA, Takamasa; GUZMAN-LEPE, Jorge; COLLIN DE L'HORTET, Alexandra. From hacking the human genome to editing organs. **Organogenesis**, v. 11, n. 4, p. 173-182, 2015.

TORSVIK, Vigdis Lid; GOKSOYR, Jostein. Determination of bacterial DNA in soil. **Soil Biology and Biochemistry**, v. 10, n. 1, p. 7-12, 1978.

UFARTÉ, Lisa; LAVILLE, Elisabeth; DUQUESNE, Sophie; MORGAVI, Diego; ROBE, Patrick; KLOPP, Christophe; RIZZO, Angeline; PIZZUT-SERIN, Sandra; POTOCKI-VERONESE, Gabrielle. Discovery of carbamate degrading enzymes by functional metagenomics. **PLoS One**, v. 12, n. 12, p. e0189201, 2017.

VARGAS, Bruna Damaceno; BASSO, Arthur; RODRIGUES, Thais Valença; SILVA, Luana Bizzi; GATZKE, Monica; FRIZZO, Matias Nunes. *Biotecnologia e alimentos*

geneticamente modificados: uma revisão. **Revista Contexto & Saúde**, v. 18, n. 35, p. 19-26, 2018.

VENTER, J. Craig; REMINGTON, Karin; HEIDELBERG, John F.; HALPERN, Aaron L.; RUSCH, Doug; EISEN, Jonathan A.; WU, Dongying; PAULSEN, Ian; NELSON, Karen E.; NELSON, William; FOUTS, Derrick E.; LEVY, Samuel; KNAP, Anthony H.; LOMAS, Michael W.; NEALSON, Ken; WHITE, Owen; PETERSON, Jeremy; HOFFMAN, Jeff; PARSONS, Rachel; BADEN-TILSSON, Holly; PFANNKUCH, Cynthia; ROGERS, Yu-Hui; SMITH, Hamilton. Environmental genome shotgun sequencing of the Sargasso Sea. **science**, v. 304, n. 5667, p. 66-74, 2004.

VORA, S.; TUTTLE, M.; CHENG, J.; CHURCH, G. Next stop for the CRISPR revolution: RNA-guided epigenetic regulators. *The FEBS journal*. 2016 Sep;283(17):3181-93.

WANG, Hong-Xia; LI, Miangqiang; LEE, Ciaran M.; CHAKRABORTY, Syandan; KIM, Hae-Won; BAO, Gang; LEONG, Kam W. CRISPR/Cas9-based genome editing for disease modeling and therapy: challenges and opportunities for nonviral delivery. **Chemical reviews**, v. 117, n. 15, p. 9874-9906, 2017.

WANG, Tim; WEI, Jenny J.; SABATINI, David M.; LANDER, Eric S. Genetic screens in human cells using the CRISPR-Cas9 system. **Science**, v. 343, n. 6166, p. 80-84, 2014.

WANG, Ruiying; LI, Hong. The mysterious RAMP proteins and their roles in small RNA-based immunity. **Protein Science**, v. 21, n. 4, p. 463-470, 2012.

WATSON, J. P. Water movement in two termite mounds in Rhodesia. **The Journal of Ecology**, p. 441-451, 1969.

WATSON, James D.; BAKER, Tania A.; BELL, Stephen P.; GANN, Alexander; LEVINE, Michael; LOSICKE, Richard. *Biologia Molecular do Gene*. In: RNAs Reguladores. Porto Alegre: Artmed, 2015. 7ed. Cap.20, p.701-732.

WARNECKE, Falk; LUGINBUHL, Peter; IVANOVA, Natalia; GHASSEMIAN, Majid; RICHARDSON, Toby H.; STEGE, Justin T. Metagenomic and functional analysis of hindgut microbiota of a wood-feeding higher termite. **Nature**, v. 450, n. 7169, p. 560-565, 2007.

WESTMANN, Cauã A.; ALVES, Luana de Fátima; SILVA-ROCHA, Rafael; GUAZZARONI, María-Eugenia. Mining novel constitutive promoter elements in soil metagenomic libraries in *Escherichia coli*. **Frontiers in microbiology**, v. 9, p. 1344, 2018.

WIEDENHEFT, B.; LANDER, G.C.; ZHOU, K.; JORE, M.M.; BROUNS, S.J.; VAN, der Oost J.; DOUDNA, J.A.; NOGALES, E. Structures of the RNA-guided surveillance complex from a bacterial immune system. *Nature*. 2011 Sep;477(7365):486-9.

WIETHÖLTENANDAKUMAR, R.; BABU, S.; VISWANATHAN, R.; RAGUCHANDER, T.; SAMIYAPPAN, R. Induction of systemic resistance in rice against sheath blight disease by *Pseudomonas fluorescens*. **Soil Biology and Biochemistry**, v. 33, n. 4-5, p. 603-612, 2001.

WRIGHT, A.V.; NUÑEZ, J.K.; DOUDNA, J.A. Biology and applications of CRISPR systems: harnessing nature's toolbox for genome engineering. *Cell*. 2016 Jan 14;164(1-2):29-44.

WRIGHT, Gerard D. Bacterial resistance to antibiotics: enzymatic degradation and modification. *Advanced drug delivery reviews*, v. 57, n. 10, p. 1451-1470, 2005.

YANG, Jian; YANG, Fan; REN, Lili; XIONG, Zhaohui; WU, Zhiqiang; DONG, Jie; SUN, Lilian; ZHANG, Ting; HU, Yongfeng; DU, Jiang; WANG, Jianwei; JIN, Qi. Unbiased parallel detection of viral pathogens in clinical samples by use of a metagenomic approach. **Journal of clinical microbiology**, v. 49, n. 10, p. 3463-3469, 2011.

YAO, S.; HE, Z.; CHEN, C. CRISPR/Cas9-mediated genome editing of epigenetic factors for cancer therapy. *Human gene therapy*. 2015 Jul 1;26(7):463-71.

YOUNG, Joshua K. et al. The repurposing of type IE CRISPR-Cascade for gene activation in plants. **Communications biology**, v. 2, n. 1, p. 1-7, 2019.

XIA, Xiaofeng et al. Metagenomic sequencing of diamondback moth gut microbiome unveils key holobiont adaptations for herbivory. **Frontiers in microbiology**, v. 8, p. 663, 2017.

ZANEVELD, Jesse R.R.; PARFREY, Laura Wegener; TREUREN, Will Van; LOZUPONE, Catherine; CLEMENTE, Jose C.; KNIGHTS, Dan; STOMBAUGH, Jesse; KUCXYNSKI, Justin; KNIGHT, Rob. Combined phylogenetic and genomic approaches for the high-throughput study of microbial habitat adaptation. **Trends in microbiology**, v. 19, n. 10, p. 472-482, 2011.

ZHANG, R.; MINER, J.J.; GORMAN, M.J.; RAUSCH, K.; RAMAGE, H.; WHITE, J.P.; ZUIANI, A.; ZHANG, P.; FERNANDEZ, E.; ZHANG, Q.; DOWD, K.A. A CRISPR screen defines a signal peptide processing pathway required by flaviviruses. *Nature*, 535(7610), pp.164-168. 2016.

ZHANG, Wei-Wei; TOMMASI, Federica; BERTOLA, Antonella; BORTOLIN, Giorgia; CACCIN, Elisabetta; CECCO, Sara; FERRARIN, Emanuela; GIACOMIN, Elisa; BALDO, PAOLO. The first approved gene therapy product for cancer Ad-p53 (Gendicine): 12 years in the clinic. **Human gene therapy**, v. 29, n. 2, p. 160-179, 2018.

ZHANG, Sai; COOPER-KNOCK, Johnathan; WEIMER, Annika K.; SHI, Minyi; MOLL, Tobias; MARSHALL, Jack N. G.; HARVEY, Calum; NEZHAD, Helia Ghahremani; FRANKLIN, John; SOUZA, Cleide do Santos; NING, Ke; WANG, Cheng; LI, Jingjing; DILLIOTT, Allison A.; FARHAN, Sali; ELHAIK, Sali; ELHAIK, Eran; PASNICEANU, Iris; LIVESEY, Matthew; SNYDER, Michael P. Genome-wide identification of the genetic basis of amyotrophic lateral sclerosis. **Neuron**, v. 110, n. 6, p. 992-1008. e11, 2022.

ZERBS, Sarah; FRANK, Ashley M.; COLLART, Frank R. Bacterial systems for production of heterologous proteins. **Methods in enzymology**, v. 463, p. 149-168, 2009.

ZHENG, Yanli et al. Endogenous type I CRISPR-Cas: from foreign DNA defense to prokaryotic engineering. **Frontiers in bioengineering and biotechnology**, v. 8, p. 62, 2020.

Zhao Z. et al. Keeping up with the genomes: efficient learning of our increasing knowledge of the tree of life. **BMC Bioinformatics** v 21 (412), 2020.

MATERIAL SUPLEMENTAR

Sequências selecionadas a partir dos dados metagenômicos da microbiota intestinal do cupim *Syntermes wheeleri*.

Cas1

>cas1-P31_assembly_scaffold_59_7|rk-newtermite-P31_Clostridiales_42_22|rk-newtermite-P31 # 7860 # 8840 # 1

MTEQAGAKPELQELPQIKERISFLYLERAVINRSDSAITVTDSDRGVVHVAASLGVILLGPGTKI
THRAVELIGDTGASIVWVGEHGVRYAHGRPLTHSSRLIAQALLISNVSSRAAVARQMYSLRF
PHEDVSNMSLQQLRGREGARIRAVYRKFQKQTGVAVWQGREYDPDDFESSSPVNKALSAAHA
CLYGVVHSVIVAIGCSPGLGFIHSGHERSFVYDIADLYKADITPIAFAEVAAKHDEITDIGSFTRRA
VRDAICDGRILIRAVKDIRFLLMEETEDDINIDIVELWDNKNGFVKSGVAYGIDDETENDVEGYGE
VL*

>cas1-P31_assembly_scaffold_867_4|rk-newtermite-P31_UNK|rk-newtermite-P31 # 1365 #
2291 # -1

MIKRTLYFGNPAYLSLKNAQLVIKLPEVEKNDTLPENFKVESVRTIPVEDIGVVVLDNKQITLTHG
VIEALLENICALITCDNRRMPVGLLLPLEGNTVQNERFRAQIEASLPLKKQLWQQTIAKIENQA
FVLKTCRGTVEGNMKKWANDVKSGDNDNLEGRAAVYYWQNLFEKFIRDREGVPPNMLNYG
YAILRAVVARGLVSSGLLPTLGIHHNRYNAYCLADDIMEPYRPFVDKLVVEIVDSGEDIENLTKE
IKAKLLQIPVLDVKINGQRSPLMVAVTQTASLAKCFLGENRKIVYPSFE*

>cas1-P13_assembly_scaffold_2566_3|rk-newtermite-P13_UNK|rk-newtermite-P13 # 1332 #
2363 # 1

MKLLNNTLFTLDPDVMVNLDTENVCVKRDGEVLLRVPLLNLEGLVLFNYFGATPQLLGECAKRG
ITVSFLSEYGKFLGSFYGETKGNVLLRKEQYRISDDSVRRLEYARNFVFGKLNQKWLRLERGLR
DYALRVNNDLLEQVSQKITKNMKQTLECEDVGLSRAIEGNAAQSYFGAFDELILRNKDGFTFTN
RNRPPPTDPVNAMLSFAYTLLASECRHALESVGLDSYVGFHTDRPGRASLALDLMEELRPYF
ADRFVLSLINRNEIAPSDFTTQGGGAVLLSDGARKNFLAEWQQRKKTISHPFLKEKLEWGIVP
YVQALLSRTIRGDLEQYPPFLWK*

>cas1-P13_assembly_scaffold_6510_3|rk-newtermite-P13_UNK|rk-newtermite-P13 # 793 #
1584 # -1

MLHIGLWSDTGASVLWIGE HGVRYAHGRPLTHSSRLLVAQASLVSNVRSRAAVARRMYAMR
FPGEDVSALSMQQLRGREGARIRAVYRKLKSKETGVPWHGREYDPDDFNAGDSVNKALSSAHA
CLYGLAHSVIVALGCSPGLGFIHTGHERSFVYDIADLYKAQVTPIAFAEIAAKYTEKDDIGSLTRRA
VRDSISKSRILIRSVQDIRALLEKQDDVEVELLQLWDDKNGYIKSGVSYGNVDTKDNDFDLEDG
YGSILED*

>cas1-P12_assembly_scaffold_9077_3|rk-newtermite-P12_UNK|rk-newtermite-P12 # 1348 #
2379 # 1

MKLLNNTLFTLDPDVMVNLDTENVCVKRDGEVLLRVPLLNLEGLVLFNYFGATPQLLGECAKRG
ITVSFLSEYGKFLGSFYGETKGNVLLRKEQYRISDDSVRSLEYARNFVFGKLNQKWLRLERGLR
DYALRVNNDLLEQVSQKITKNMKQALECEDAGSLRAIEGNAAQSYFGAFDELILRNKDGFTFTN
RNRPPPTDPVNAMLSFSYLLASECRHALESVGLDSYVGFYHTDRPGRASLALDLMEELRPYF
ADRFVLSLINRNEIAPSDFTTQGGGAVLLSDGARKSFLAAWQQRKKTISHPFLKEKLEWGIVP
VQALLSRTIRGDLEQYPPFLWK*

>cas1-WG1_assembly_scaffold_5129_2|rk-newtermite-WG1_UNK|rk-newtermite-WG1 # 508 #
1539 # -1

MKKLLNTLFVTLPDVMVNLDTENVCVKRDGEVLLRVPLLNLESLVLFNYFGATPQLLGECACRG
ITVSFLSEYKFLGFSFYGETKGNVLLRKEQYRISDDSVRSLEYARNFVFGKLNQKWVLERGLR
DYALRVNNDLLEQVSQKITKNMKQALECEDAGSLRAIEGNAAQSYFGTFNELILRNKDGFTFTN
RNRPPPTDPVNAMLSFAYTLLASECRHALESVGLDSYVGGFFHTDRPGRASLALDLMEELRPYF
ADRFVLSLINRNEIAPSDFTTQSGGAVLLSDGARKSFLAAWQQRKKTISHPFLKEKIEWGIVPYV
QALLSRTIRGDLEQYPPFLWK*

>cas1-P32_assembly_scaffold_317_7|rk-newtermite-P32_UNK|rk-newtermite-P32 # 8957 #
9964 # -1

MKKLLNTLFVTLPDVMVNLDTENVCVKRDGEVLLRVPLLNLEGLVLFNYFGATPQLLGECACRG
ITVSFLSEYKFLGFSFYGETRGNVLLRKAQYRISDDSVRSLEYARNFVFGKLNQKWVLERGLR
DYALRVNNDLLEQVSQKITKNMKQVLTCECEDAGSLRAIEGNAAQSYFGALEDFIYTSRNRPPPTD
PVNAMLSFAYTLLASECRHALASVGLDSYVGGFFHTDRPGRASLALDLMEELRPYFADRFVLSLI
NRNEIAPSDFTTQSGGAVLLSDGARKNFLAAWQHRKKTISHPFLKEKLEWGIVPYVQALLSR
TIREDLEQYPPFLWK*

>cas1-P32_assembly_scaffold_1520_4|rk-newtermite-P32_Firmicutes_42_16|rk-newtermite-
P32 # 3065 # 4096 # 1

MKKLLNTLFVTLPDVMVNLDTENVCVKRDGEVLLRVPLLNLEGLVLFNYFGATPQLLGECACRG
ITVSFLNEYKFLGFSFYGETKGNVLLRKEQYRISDDLRSLEYARNFVFGKLNQKWVLERGLR
DYALRVNSDLLEKVSQKITMNMKQTLCECEDAGLLRAIEGNAAQSYFSTFNELILRNKKDFTYTS
RSRRPPKDPVNAMLSFAYTLLASECRHALGCVGLDSYVGFYHTDRPGRASLALDLMEELRPYF
ADRFVLSLINRNEIAPSDFTTQSGGAVLLSDGARKSFLAAWQQRKKTISHPFLKEKLEWGIVPY
VQALLSRTIRGDLEQYPPFLWK*

>cas1-P32_assembly_scaffold_8570_4|rk-newtermite-P32_Actinobacteria_53_11|rk-
newtermite-P32 # 2017 # 2946 # 1

MTEQAGAKKPELQELPQIKERISFLYLERAVINRSDSAITVDSRGVVHVAASLGVILLGPGTKII
GEHGVRYAAHGRPLTHSSRLIAQALLISNVSSRAAVARQMYSLRFPHEVSNMSLQQLRGRE
GARIRAVYRKFSKQTGVAWQGREYDPDFESSSPVNKALSAHAACLYGVVHSVIVAIGCSPGL
GFIHSGHERSFVYDIADLYKADITIPAFEVAAKHDEITDIGSFTRRAVRDAICDGRILIRAVKDIRFL
LMEETEDDINIDIVELWDNKNKGFVKSGVAYGIDDETENDVEGYGEVL*

>cas1-WG3_assembly_scaffold_3116_3|rk-newtermite-WG3_UNK|rk-newtermite-WG3 # 1376
2407 # 1

MKKLLNTLFVTLPDVMVNLDTENVCVKRDGEVLLRVPLLNLEGLVLFNYFGATPQLLGECACRG
ITVSFLSEYKFLGFSFYGETKGNVLLRKEQYRISDDSVRSLEYARNFVFGKLNQKWVLERGLR
DYALRVNNDLLEQVSQKITKNMKQALECEDAGLLRAIEGNAAQSYFGAFDELILRNKDGFTFTN
RNRPPPTDPVNAMLSFAYTLLASECRHALESVGLDSYVGFHTDRPGRASLALDLMEELRPYF
ADRFVLSLINRNEIAPSDFTTQSGGAVLLSDGARKSFLAAWQQRKKTISHPFLKEKLEWGIVPY
VQALLSRTIRGDLEQYPPFLWK*

>cas1-WG2_assembly_scaffold_0_93|rk-newtermite-WG2_Firmicutes_43_74|rk-newtermite-
WG2 # 105984 # 106979 # -1

MSQVYVHEHGAVIGVRDGRLEINSEEVRRSIPFESVDGISIFGNSQVSTACMKHLLKHGIATHFY
SSTGTYFGRLISTGHTNAARQKAQILASCDENLCLEFSKSFIEAKLHNQYVLLRRYRGEKDISTE
ASYITMMRGKIAECKSIAEIMGCEGTAARYYYKALGKLVKNEFYFEGRTRRPPLDPFNSMLSLG
YSILLNEVFGAIESVGLNPFYGLLHSEHDKMPALACDLMEEWRPVIVDSTVMALVSGNEISIDEF
YEKEERQGVFLTKEGFSIFIKKLEMKMHSKSKYLDYVEFPISLRTSLNKQALRIVKALEESNPKEY
TGVRIR*

>cas1-WG2_assembly_scaffold_2_23|rk-newtermite-WG2_Firmicutes_43_74|rk-newtermite-
WG2 # 18051 # 19082 # -1

MKLLNNTLFVTLDPVLVNLNENVCIKRDNEVLLRIPLLNLEGLVLFNYFGATARLMAECAKRNIT
ISFLSEYKFLGTFYGETKGNVLLRKEQYRISDDPKKNIVFSRNFIFGKLNQKWWVERGLRDH
SMRVNECVLETASRSITENMKQALECSDVDTLRALEGNAAHYFRAVNELILQNKEDFIFENRS
RRPPTDPVNALLSFAYTLTGEGRHALETAGLDSYVGFMHVDRPGRASMALDLVEEMRPHFA
DRFVLSLINRNEITPDDFVKQESGAVLMSDKARKEFLSAWQQRKKTITHPFLKEKVEWGIVPHI
QAMLLARTIRGDLDEYPPFLWK*

>cas1-WG2_assembly_scaffold_24143_2|rk-newtermite-WG2_UNK|rk-newtermite-WG2 # 654
1661 # -1

MKLLNNTLFVTLDPVDMVNLDTENVCVKRDGEVLLRVPLLNLEGLVLFNYFGATPQLLGECAKRG
ITVSFLNEYKFLGFSFYGETKGNVLLRKEQYRISDDSVRSLEYARNFVFGKLNQKWWLERGLR
DYALRVNSDLLEKVSQQITMNMKQTLCECEDAGLLRAIEGNAAQSYFSTFNELILRNKDGFTFTN
RNRPPPTDPVNAMLSFAYTLTASECRHALASVGLDSYVGGFFHTDRPGRASLALDLMEELRPYF
ADRFVLSLINRNEIAPSDFTTQGGAVLLSDGARKSFLAAWQQRKKTISHPFLCTSTSAFSHY
QGRSRAVS AVFMEIGV*

Cas2

>cas2-P31_assembly_scaffold_867_3|rk-newtermite-P31_UNK|rk-newtermite-P31 # 1031 #
1336 # -1

MWVLFVFDLPTETKKERKAHALFRKLLQDGFMTFQFSIYLRHCASRENADVHIKRVKLLPEK
GHVGIMCITDKQFGNMELFYCKKQTEVETPSQQLELF*

>cas2-P31_assembly_scaffold_24112_1|rk-newtermite-P31_UNK|rk-newtermite-P31 # 149 #
439 # -1

MLVLITYDVNTETPEGRKRLRQVSKTCLNYGQRVQNSVFECVVDAGQLLKVNRLVHLINAETD
SLRFYNLGNRRNRVEHFGAKPSFDVSAPLVV*

>cas2-P13_assembly_scaffold_6510_2|rk-newtermite-P13_UNK|rk-newtermite-P13 # 460 #
771 # -1

MIVLNLTDCPIGLRGDLTKWLLIEIASGVYVGGVQVSARVRDEIWSRVTEHCRNGRAVLVYNTNDE
QRLDFRVHGDWEPIDFDGIKLNALQNGFPPCAHATLTTT*

>cas2-P13_assembly_scaffold_10056_2|rk-newtermite-P13_UNK|rk-newtermite-P13 # 817 #
1137 # 1

MNKFMRLVFFDLVPKTKKERRGATQFRNFLIKDGYHMMQYSLYARVCNGMDAVEKHKRVYL
NVPDNGSVRMLVITERQYESIEILVGNLTECDNAFETEQLSFF*

>cas2-P13_assembly_scaffold_19215_1|rk-newtermite-P13_UNK|rk-newtermite-P13 # 55 #
345 # -1

MLVIIITYDVNTESPDGRRRLRQVAKVCVNYGQRVQNSVFECMADSAQLLTIKDKLVNLIPTED
SLRFYSLGNKYKGVVEHYGIKPSFDVTEPLIL*

>cas2-P12_assembly_scaffold_9077_4|rk-newtermite-P12_UNK|rk-newtermite-P12 # 2391 #
2681 # 1

MLVLVTYDVSTGTPAGRKRLNKVAKRCVAHGQRVQNSVFECNLDWAQYISLQNELKNLIDPTE
DSLRFYNLGNLYEYKIVHVGKSSIQMEGELIV*

>cas2-WG1_assembly_scaffold_14145_2|rk-newtermite-WG1_UNK|rk-newtermite-WG1 # 665
955 # 1

MLVLITYDVNTETPEGRKRLRQVSKTCLNYGQRVQNSVFECVVDAGQLLKVDRLVRLINAETD
SLRFYNLGNNSRNKVEHFGAKPSFDVSAPLVV*

>cas2-P32_assembly_scaffold_317_6|rk-newtermite-P32_UNK|rk-newtermite-P32 # 8655 # 8945 # -1

MLVLITYDVSTRTPVGRKRLSKVAKRCVAHGQRVQNSVFECNLDWAQYISLQNELKNLIDPSED
SLRFYNLGNITYEEKIVHVGTRSSIQLELIV*

>cas2-P32_assembly_scaffold_6718_2|rk-newtermite-P32_Bacteroidia_41_7|rk-newtermite-P32 # 1038 # 1289 # 1

MASDFRKKIMKDGFTQFQFSIYLRHCASRENAEVHIKRVKNHLPGLGQVILCITDKQFGKMEL
FQCKKEKPVTPYQQLF*

>cas2-P32_assembly_scaffold_8049_3|rk-newtermite-P32_UNK|rk-newtermite-P32 # 1310 # 1600 # 1

MMVLVSYDVMVTSPGGKRRLRKVAKACTNYGQRVQYSVFECVIDPAQWVVKLNILEKIIDDKT
DSLRYYYL GANYKRRVEHVGAKPSYDVGPLIV*

>cas2-P32_assembly_scaffold_8570_5|rk-newtermite-P32_Actinobacteria_53_11|rk-newtermite-P32 # 2946 # 3854 # 1

MILQLVSCPPALRGDLTKWLVEIAAGVYVGRVSARVRDRIWSRVVENCKEGRSVLVHSTNNEQ
RFDFRIHGESWEPIDFDGIKMLRPNVERITKSSEINSGKIGGFSNASKIRAARNMSKEVSYPT
DYIVIDVETTGLNPEKDAVIELGAMKIIDRKITETQALILVDITIPAVITKLTGLDVEILQKEGQTLDD
AVAGFIEFCGNQPLIAHNMDFDRSFLNKAMEKCSLPPIQNHTIDTLAMAKRMYRGLKSYKLDL
EAYLKL AHTVDNTPKYNSHRSLGDCNVTHMLYQKLLNYYN*

>cas2-P32_assembly_scaffold_13839_2|rk-newtermite-P32_UNK|rk-newtermite-P32 # 330 # 1247 # 1

MIVLNLTDPCPIGLRGDLTKWLEIASGVYVQVSARVRDEIWSRVKEHCRNGRAVLVYNTNDE
QRLDFRVHGDWEPIDFDGIKMLRPSASRLKRRRTAAHGGRNDGFSNASKKRAVKRFSTMR
TRYPDDYVVVDLETTGLNPETDSIIEIGAVRMISHERVDTFNAFVVERPIPSEVTALTGITDQMI
ESGRELKDVNLQFSLFIGNFPIVGHNIDFDKSFLLNACEKCEIPLIMNRSVDTMSLARRLAKDVE
NYKLETMAMKFSIERKGYHRSINDCETTALLYRKL NLM DIGSESIEI*

>cas2-P32_assembly_scaffold_14287_2|rk-newtermite-P32_UNK|rk-newtermite-P32 # 872 # 1162 # 1

MLVLVTYDVSTETREGKRRLNKVAKKCVANGQRVQNSVFECNLDWSQFVKLQNELKAIINTKT
DSLRFYNLGSAYKEKITHIGAKPSIDLDELIF*

>cas2-P32_assembly_scaffold_42552_3|rk-newtermite-P32_UNK|rk-newtermite-P32 # 1253 # 1543 # 1

MMVLVSYDVMVSSPGGKTRLRKVAKECTNYGQRVQFSVFECVIDPAQWVVKLNLSLEKIINDKT
DSLRYYYL GANYKRRVEHVGAKPSYDVDEPLIV*

>cas2-WG2_assembly_scaffold_0_92|rk-newtermite-WG2_Firmicutes_43_74|rk-newtermite-WG2 # 105658 # 105987 # -1

MISDDYFIDIHNDDFDRGRKYFVLVIYDIVDNKRRSIAKTLKRYAFRVQKSAFEAHLTPSKYDIL
VSILTNLIDDKEDSVRIYKIRGKAAILSMGIGRSPDDEEVIII*

>cas2-WG2_assembly_scaffold_2_22|rk-newtermite-WG2_Firmicutes_43_74|rk-newtermite-WG2 # 17746 # 18036 # -1

MLVLVTYDVSTETPQGRKRLSKVAKKCVANGQRVQNSVFECNLDWSQFVKFRSELEAIIDSSE
DSLRYYNLGNYYKEKIIHIGTKQSIPLELIV*

>cas2-WG2_assembly_scaffold_534_1|rk-newtermite-WG2_UNK|rk-newtermite-WG2 # 417 # 677 # -1

MNKFMRILVFFDLPVKTKEERRGATQFRNFLIKDGYHMMQYSLYARVCNGMDAVEKHKKRVYL
NVPDNGSVRMLVITDFSWKSNRM*

>cas2-WG2_assembly_scaffold_24143_1|rk-newtermite-WG2_UNK|rk-newtermite-WG2 # 360
650 # -1

MLVLITYDVSTRTPAGRKRLSKVAKRCVAHGQRVQNSVFECNLDWAQYISLQNELKNLIDPTED
SLRFYNLGNITYEEKIVHVGTKSSIQLEGELIV*

Cas3

>cas3-P31_assembly_scaffold_59_1|rk-newtermite-P31_Clostridiales_42_22|rk-newtermite-
P31 # 585 # 3335 # 1

MSHHNAKKTSLWAKKSNPDYDKWPLPLHTLNDTAEVACLLWDNWASKSVIKSISKGITADGL
DGYLIAKQVFAFLAYVHDIGKATPVFQVKEAYGKEDLNDFTLNQIKKSGLPINHTYHYKQETRHE
LVGYMILKKGHLDDSVAVIIGSHHGIPPENQINNIKHGIYDVNCGFDNLLWKSQSEFFNNAIEF
ADLNKATVLSLGITRPVQVLLSGLLIMADWIASNEFLFKLINIEDSPLNINSLQRTTNAFQELKFP
PWYIDQEWNLYYTRFGIESPHPVQIVAVEVAKSIIHTPGIFIEAPMGEGKTEAALAVAEIFAGNS
SARGIYFALPSQATSNAMFTRVINWLRTFDNQEISVRLVHGKAEFNEEYNDIANLSRSTEFIDN
DINEENLEIIHEWFKGRKKELLADFAVGTIDHVLMSGLKQKHLMLRHLGLAGKVLIIDECHAYDV
YMESYLLKALNWL GAYGVPVVLSATLPLSKRQNLIEAYLNKKNLGV DSEWAASLAYPMVTYS
DGISVNHPVTVPIAIEKGT DVSIIHSLNECEISDTLYRVLEKGGYAGVIFNTVKRAQEYTTIKITSFE
DAEVDLLHGGFLAIDRVKKEENLLLRLGKSNTIQRSEKYIVVGTQIFEQSLDIDFDVLITQLCPVDL
LLQRIGRLHRHKRDNRPGLKQPRCYVLGANGHDYDEGSAIYSLYLLMRTVEALPQKINLPND
IPKLVAKVYSENDTEASINARIKWEERNEARKDSADIYQINRWKSSKTLINWLDKIRDDSEKRA
EAAVRDGVDSIEVLIQKRQGLL CFLPWINGEALPFTTPTPKLAKIIAGCSVRLPSYFAWRNIAD
RAILEIEEIMMQUERIKSNWYESHWLK GALVMILDENQNVEICDCCINYDEQMGLSFIKKGDSNER
N*

>cas3-P12_assembly_scaffold_1848_1|rk-newtermite-P12_UNK|rk-newtermite-P12 # 97 #
2475 # 1

MMIVCAILSQKVEIIMTYLARFNETREQLLVEHLNNVSELSAHFSAYTSITRILIGILHDLGKATAAF
RNYLEEGVGRGSVVHALQGAFVDDNVGSSNETADILLKEIVSLAIVAHHGSLSDGVSPDGETV
FIQRLADKENEKYNQEAQRNFSSLDLTDSDISDLFLRAKDEVLSLVSRTITQTYIYQESAQFALG
LFVKYIYSCOLIDADRLDAYLFELNEQYQPADMDWDRLTATFEENIQKYECDSEIAKIRRNISDQC
KEAAIKNTGIYRLSVPTGGGKTLSSRFALHHCKKKNKRIIYVIPYLSIIDQTATEFQKILNLQSN
DIILEHHSNIVLPDDEDEREIRKLATSRWDKPIIITTMVQFLETVFSSRGSDLRKFHNMSDSVIIFD
EIQSLPTKTINLNFNEAVSFLAKLCNTTVLLCTATQPLLDKTERGNLLESPELIDCGEAFKSIKRTI
IVAEKEKDVNEFASFVELKAQENENCLAIVNTRKSAREVFERLKGQRDFEVYHLSTSMCGAHR
MDTIKVKAAALDRDRRVCVATQLIEAGIDISFSCVVRVAVAGLDSVAQAAGRCNRDGESESPKE
VYAIPLRDENLSKLPDIECGKEITERLIRENEGADLLAPEILEQFYEHYFYNRRNLMDYPTSDGPT
VYEMLSTNGTGKMNKYNRTGHDSACAIHAHFRSADESFCVIDKNTKQVVVYGEVEELIEAYK
KPPENRITREKLNIVRKLEKFSVSLFSWEMEKLGAHVLDEETGIMILDRHLHYSEDIGVILETDPK
DYIV*

>cas3-P12_assembly_scaffold_10_27|rk-newtermite-P12_Firmicutes_43_106|rk-newtermite-
P12 # 29751 # 32120 # 1

MHQIPNTFYARKASDDRRQNLQNHLSNVADLASKFASDIGTKGITEIAAQFHDIGKATRAFQDYL
VNEGGRRGDVVHSLQGAFVDELCDDESSALVNEIVAMVIAAHHGFLDDGISPDGESIFIEKLTCK
DEEKYYYSELKQNVAIYNTKQTRELITAKAEIEAMISLIMETYQRRKSAEYSAGFALGLFVKYIY
SCLIDADRLDAYLFDVSEQYEPPIVDWTAIINVFENIQNYGNKSDISQIRQAVSTQCKDAATKPT
GIYQLSVPTGGGKTLSSRFALHHCKVANKKRIIYVIPYLSIIETADELRKILDLPQDNDVILEHH
SGIIAQDESDDEREIRKLAASRWDKPLIITTMVQFLETVMSAKGSALRKFHNMSDSVIILDEIQL
PIKTVHLFNETVSFLAKICNSTVLLCTATQPQLHLTERENLLFENNPENLINCADLFLPIKRTSIVAE
PEKDIDAFATFVSEKAQINGNCLAIVNTKKSAREIFDKLNGEKPDGFNIYHLSTSMCPAHREITLK

NIKVALNQRQNIICITTQLIEAGVDISFACVVRATAGLDSITQAAGRCNRNGESDIPKEYIVPLQG
ENLDRLLDIKEGKKITERLVREHKNE DLSEPSIMNQFYDYFYKRNLM DYPTKDNKTLYGMLS
CNKAGRGAYKERTDKDAQCFIAHAFKAASEAFSVIDKRAESV VVAYGEAVSLIEKYRKLPTTLIT
REKLNIIIRKLEKFSVSLFSHEMEQLTKVGAIIVLDDDTNIRILSEHYYSNDVGVIMETNPDLLIL*

Cas5

>cas5-P31_assembly_scaffold_1942_3|rk-newtermite-P31_Firmicutes_45_7|rk-newtermite-P31
664 # 1353 # -1

MAFGFRVEIWGDYACFSRPELKVERVSYEVMTPSAARGIIEAVFWKPAIKYVIDQIDICAPIHFEN
IRRNELSSKIPYSSIQNAEKLLSGDL YLNASTDRTQRAAMVLRNVRYVVTGHFILT DKAGPDDN
EGKFTTILRRRLGKGQNYHTPYLGVREFPANIRLIAEEETLPKPIPESRSLGLMLYDIEYVKSTGK
DGQEVVTSFEPTYFMADMKQGVINLRDMEVFR*

>cas5-P13_assembly_scaffold_5728_2|rk-newtermite-P13_UNK|rk-newtermite-P13 # 796 #
1521 # 1

MQTERKRNSVEFEVYGKYALFSDPVTRVGGEKFSYQVPTYQALKGVLESVYWKPTFIWVIDAV
RVINRIQTEGKGIRPIKMSGGNDLAYTYLKD VRYQVLAHFEWNESRRNLQNDRNEHKHHNVA
RRCIEKGGRRDVLG VRECA YVIPCEFGEGDSFYDKYGEIDFGFILHGIDYPDEADGKFGVRF
WNCKMVDGVITFPSPESDVIKSR TIRDKQKIKAFNLGESLNPVDDEDGEQP*

>cas5-P12_assembly_scaffold_10_28|rk-newtermite-P12_Firmicutes_43_106|rk-newtermite-
P12 # 32133 # 32885 # 1

MQEITKKRNSVEFEVYGKYALFSDPITRVGGEKFSYQVPTYQALKGVLESIYWKPTLIWIIDAVRI
MNKIQTEGKGIRPIKMSGGNDLAYTYLKD VRYQVLAHFEWNEHRPELQNDRNENKHHNIAKR
SIEKGGRRDIFLGTRECA YVAPCEFGSKNSFYDYPEPYGELDFGFMLHGIDYPDETGDGTFG
TRFWHCKMNGV IIFPPDSNEVQRRV IREDQIIKEFVPGENFRMIDDELPGDKGGDE*

>cas5-P33_assembly_scaffold_754_2|rk-newtermite-P33_Actinobacteria_56_10|rk-newtermite-
P33 # 171 # 869 # 1

MQIERKRNSVEFEVYGKYALFSDPVTRVGGEKFSYQVPTYQALKGVLESVYWKPTFVWVIDAV
RVINRIQTEGKGIRPIKMSGGNDLAYTYLKD VRYQVRAHFEWNESRKNLQNDRNEHIEKGGRR
RDVFLGVRECA YVIPCEFGEGDSFYDKYGEIDFGFMLHGIDYPDEADGKFGVRFWNCKLVD
GVITFPSPESDEIKSRIVRDNQKIKAFNLGESLYPVSSEDGEQP*

>cas5-P32_assembly_scaffold_317_11|rk-newtermite-P32_UNK|rk-newtermite-P32 # 13374 #
14099 # -1

MQTERKRNSVEFEVYGKYALFSDPVTRVGGEKFSYQVPTYQALKGVLESVYWKPTFIWVIDAV
RVMNRIQTEGKGIRPIKMSGGNDLAYTYLKD VRYQVRAHFEWNE NRRNLQNDRNEHKHHNV
ARRCIEKGGRRDVLG VRECA YVPCIFGEGDSFYDKYGEIDFGFMLHGIDYPDEADGNFGV
RFWNCKMTDGIITFPPEGEDEIKSR TIRDNQKIKAFNLGESLDPVDSEDGEQP*

>cas5-P32_assembly_scaffold_1749_2|rk-newtermite-P32_UNK|rk-newtermite-P32 # 701 #
1390 # 1

MAFGFRVEIWGDYACFSRPELKVERVSYEVMTPSAARGIIEAVFWKPAIKYVIDQIDVCAPIRFE
NVRRNELSSKIPYSSIQNAEKLLSGDL YLNASTDRTQRAAMVLRNVRYVVTGHFILT DKAGPD
DNEGKFSTILRRRLGKGQNYHTPYLGVREFPANIRLIAEEETLPKPIPESRSLGLMLYDIEYVKST
GKDGQEVVTA FEPTYFMADMKQGVINLRDMEVLR*

>cas5-P32_assembly_scaffold_2174_2|rk-newtermite-P32_Actinobacteria_53_11|rk-
newtermite-P32 # 447 # 1172 # -1

MQTERKRNSVEFEVYGKYALFSDPVTRVGGEKFSYQVPTYQALKGVLESVYWKPTFVWVIDA
RVINRIQTEGKGIRPIKMSGGNDLAYTYLKD VRYQVRAHFEWNESRRNLQNDRNEHKHHNV

ARRCIEKGGRRDVLGVRECQAYVIPCEFGEGDSFYDKYGEIDFGFMLHGIDYPDEADGKFGV
RFWNCKLVDGVITFPSPESDEIKSRIVRDNQKIKAFNLGESLYPVSSEDGEQP*

>cas5-P32_assembly_scaffold_8570_2|rk-newtermite-P32_Actinobacteria_53_11|rk-
newtermite-P32 # 750 # 1460 # 1

MNTLLIRLAAPLQSWGSESKFERRTTQREPTKSGVIGILAAALGRGREDCLNDLTALRFGVRVD
QPGQIMRDYHIARFSKSKEMLERTYVTNRYYISDAVFLVGLGDKHTLEMLESALLSPFYPIFLG
RRSCPPTGRVVMGIRNKPLEETLEDEEWIAGEWYKKREKSCNLTVLVDSKELSANRRRDLPLS
FSQDYRKHKYRYIDDIPDKVQYQYANIANIATEHDPFDVTWESMEV*

>cas5-WG3_assembly_scaffold_4662_4|rk-newtermite-WG3_UNK|rk-newtermite-WG3 # 2461
3186 # -1

MQTERKRNSVEFEVYGKYALFSDPVTRVGGGEKFSYQVPTYQALKGVLESYWKPTFIWVIDAV
RVINRIQTEGKIRPIKMSGGNDLAYTYLKDVRVYQVLAHFEWNESRRNLQNDRNEHKHHNVA
RRCIEKGGRRDVLGVRECQAYVIPCEFGEGDSFYDKYGEIDFGFILHGIDYPDEADGKFGVRF
WNCKMVDGVITFPPEESDVIKSRITIRDKQKIKAFNLGESLNPVDDDEDGEQP*

>cas5-WG3_assembly_scaffold_10604_2|rk-newtermite-WG3_UNK|rk-newtermite-WG3 # 333
1034 # 1

MSTLLMRLAAPLQSWGSDSKFEERRATMREPTKSGVVGMLAAALGRRRDEPIEDLSTLRFGV
RIDQPGNLLKDYHTAKSKDGKVSYVTNRHYLADAVFIVGLEDDTLLNILEKALREPVFPLYLGR
RSCPPAGRISLGISEESLEEALKQVPWQAGEHHRKHVSSEVPQGLTLVLDADGVGNLRRRDL
PLTFSQRHRKHLRYRLVDKPDVAVRVRTKNETEHDAFEAVGKEM*

>cas5-WG2_assembly_scaffold_2_27|rk-newtermite-WG2_Firmicutes_43_74|rk-newtermite-
WG2 # 22555 # 23307 # -1

MQEITKKRNSVEFEVYGRYALFSDPITRVGGGEKFSYQVPTYQALKGVLESYWKPTLIWIIDAVRI
MNKIQTEGKIRPIKMSGGNDLAYTYLKDVKYQVLAHFVWNEHRPELQNDRNENKHHNIAKR
SIEKGGRRDIFLGTRECQAYVAPCEFGSKNSFYDYPEPYGELDFGMFLHGIDYPDETGDGTFG
TRFWHCKMNGVVIIFPPDSNEVQRRVIREDDQIIKEFVPGENFRMIDDELPGDKGGDE*

Cas9

>cas9-P32_assembly_scaffold_916_7|rk-newtermite-P32_UNK|rk-newtermite-P32 # 6981 #
10199 # -1

MRYNIGIDLGIASVGYAVMRLDDNDEPCEVIRLGSRIFNVAENPKDGASLALPRREARGMRRRL
RRKAHRKERLRNLLVSQGILSKEELKNLYKGLTDIYELRKAALDRTLTKELARVLLHIAERRGF
RSNRKATTSKEDGELLKAIDANRTIMNEKGYRTVGEMLYLDEAYS DTKRNKGENYKNTVTRDM
IEDEVIKIFEAQNSGNMCSPDFENAYLTILLAQRSFDEGPGGNSSYSGGIEKMVGDCFTFEEN
EKRAPKASYSFEYFDLLQKINHILDNNGKTLNISNDQRNKMVELLHKSPNKNFADIRKELKLPD
TVRFNFVKYEGKSIEESEKAAKLKHLKYYHEMRKAFDKIGKGRKIDISIADRNEIARIFTLHKNE
DKLRAELQNAGIGCFDIETLIDNLGSFSKFGHLSVKALDKIIPYLEQGLKYNEACEKAGYYFRGH
NHENNSRIVSLKHLADEAENKITSPPVVKRAISQCAKVINAIIREIGTSPVFINIELAREMSKDFKDR
KDAEKTMEENRAYNDRIKDELIEDTKQVNPTGLDIVKLLWKQQDGRCLYSLDPIPREYLFRPGI
ADIDHIIPYSISFDDSYKNKVLVKTEENRQKGNLLPLQYLIGKKRDDFVWVNNNSNLPLNKKRFL
KKEITPEDEEGFRERNLQDTKTMSVFLYNLNDNLAFAPFLTNKKRHVVAVNGSITSMLRKRW
GLSKFREDGDAHHAIDAAVIACCTQCDIKQLTMIDKVSQFYRAHEQRYITIDGKEHKVDKKTGEL
KEMFPRPFDTFNHDVIAKSREVFVSRMP SRKTS GAAHKETIKGASENGLLKKVALSELKDKNT
AEIVDYYPQDDRLLYDALKAQLSKYDGDAAKAFSEPFYKPKSDGSRGPLVKKVKITEKATLSV
PVHAGTGRANNGEMVRVDVFHIEDDGYLVPYIYADTVKSKLPNKAIIVANKPYKEWKEVRDED
FTFSLYPNDLIRIKPKAPMQFRKVNKKSVPETVEKEETFVYYK GASISTGAITVINHDNSYMISS
LGTKTLKAFEKHQVDVLGNVSKVGKEKRQEFAAKGR*

>cas9-WG2_assembly_scaffold_534_3|rk-newtermite-WG2_UNK|rk-newtermite-WG2 # 1582 # 4800 # -1

MRYNIGIDLGIASVGYAVMRLDDNDEPCEVIRLGSRIFNVAENPKDGASLALPRREARGMRRRL
RRKAHRKERLRNLLVSQGILSKEELKNLYKGTLDIYELRKAALDRTLTKELARVLLHIAERRGF
RSNRKATTSKEDGELLKAIDANRTIMNEKGYRTVGEMLYLDEAYSDTKRNGENYKNTVTRDM
IEDEVLKIFTAQRNFGSNTCSPDFENAYLTILLAQRSFDEGPGGNSPYSGGIEKMVGDCFTFEEN
EKRAPKASYSFEYFDLLQKINHILKLDNNGKTLNISNDQRNKMVELLHKSPNKNFADIRKELKLPD
TVRFNFVKYKEYGKSIEESEKAAKLKHLKHYHEMRKAFDKISKGRIKDISIADRNEIARIFTLHKNE
DKLRAELQNAGIGCFDIETLIDNLGFSKFGHLSVKALDKIIPYLEQGLKYNEACEKAGYYFRGH
NHENNSRIVSLKHLADEAENKITSPPVVKRAISQCAKVINAIIREIGTSPVFINIELAREMSKDFKDR
KDAEKSMEENRAYNDRLKDELIKTDRQVNPTGLDIVKLLWKQQDGRCLYSLDPIPREYLFRRPG
IADIDHIIPYSISFDDSYKNKVLVKTEENRQKGNLLPLQYLIGKKRDDFIVWVNSNLPLNKKRFL
KKEITPEDEEGFRERNLQDTKTMSVFLYNLNDNLAFAPFLTNNKRHVAVNGSITSMLRKRW
GLSKFREDGDAHHAIDAAVIACCTQCDIKQLTMIDKVSQFYRAHEQRYITIDGKEHKVDKKTGEL
KEMFPRPFDFTNHDVIAKSREVFVSRMPSRKTSGAAHKETIKGASENGLLKKVALSELKLDKNT
AEIVDYNNPQDDRLLYDALKAQLSKYDGDAAKAFSEPFYKPKSDGSRGPLVKKVKITEKATLSV
PVHAGTGRANNGEMVRVDVFHIEDDGYLVPYITADTVKSELNKAIVANKPYEEWKEVRDED
FTFSLYPNDLIRIKPKAPMQFRKVNKKSVPETVEKEETFVYYKASISTGAITVINHDNSYMISS
LGTKTLKAFEKHQVDVLGNVSKVGKEKRQEFAAKGR*

RuvC

>ruvC-WG1_assembly_scaffold_531_10|rk-newtermite-WG1_UNK|rk-newtermite-WG1 # 9778 # 10287 # -1

MILGIDPGVATVGFGIIEEGGVCKQVRYGVITTPAQMLRALRLKQINNDVSELIQTFKPDIAIVE
ELFFNTNQKTALSVAHGAAVILAGEAQGIPMFEYTPLQVKKAVAGYGHATKKQVMEMVKRLL
LMEQNPKPDDAADALAAIACHARAANSLLSIEGQGISATI*

>ruvC-WG1_assembly_scaffold_15939_2|rk-newtermite-WG1_UNK|rk-newtermite-WG1 # 960 # 1466 # 1

MKILGIDPGTGICGFGVIEIGKSNPKMITAGVISTPPHTPLPDRLLDIYDSFREIINETKPDVVSIE
KLFFVQNITTGISVAHARGVAMLVARQFELPIFEYTPLQIKQTLTGYGRADKKQVQEMVKLHLNL
TKVPKPDDAADALAAAITCSLIERGQQKSPQTPNL*

>ruvC-P32_assembly_scaffold_1732_3|rk-newtermite-P32_Bacteroidia_41_7|rk-newtermite-P32 # 3448 # 4014 # 1

MTKDQIILGIDPGTTIMGYGLLKITNNKPSLLAMGVVDLKKYGDHYLKLQRIFARTLSLIDFKPDC
LAIEAPFYGKNVQSMLKLGRAQGTAMAAALYRDIPITEYMPKIKMAITGNRASKEQVSDMLR
RFLKIPENQMLPQLDATDGLAAAYCHFLQMDNPISDTKYRGWKDYIAKNQEKVKGKR*

>ruvC-P32_assembly_scaffold_7351_2|rk-newtermite-P32_Bacteroidia_41_7|rk-newtermite-P32 # 1137 # 1760 # 1

MERVILGVDPGTVMGFGVIKQVGTIQFIEMDVNLNLSKELDVAKLHRIFECTVALIEKHNP
LAIEAPFFGKNVQSMLKLGRAQGVAAIAAAMSRNVPIEYAPRKIKMAITGKGATASKEQVAKMLFS
LMNINDNMPKHWDATDGLAAAVCHALQNDIPGKSEPSQLPVLKGRTRRSASTKKNQPTKKYS
DWEKFIAQNSARVSS*

>ruvC-P32_assembly_scaffold_8247_3|rk-newtermite-P32_UNK|rk-newtermite-P32 # 982 # 1470 # -1

MRIIGIDPGYAILGYGIVEKKNHFTPCFHAVTTDTKADMTDRLKHYYHTLMNIIESEYEPEVASIE
ELFFNTNNTALLVGQARGVAILACANSGLSICEYTPLQIKQALVGYGRAEKVQVQAMVKTLNL
SETPKPDDTADALAAIACHGNSSGRLSSYK*

Cascade - CasA

>cascade-CasA_P31_assembly_scaffold_59_2|rk-newtermite-P31_Clostridiales_42_22|rk-newtermite-P31 # 3322 # 4926 # 1

MKEIDFNLLDERWIIVTNSYGNEEILSLTDVLIRAHRELRLCLSGEMEAQDVAILRLLLGILYAIYTRTE
DYIEAQEDGDGERCIEIWKNLWERGKFPGDEVTDGYFKQYHDFRFLIHPERPFYQVASLKSQKT
SEYKISKMLGDLVQSGNKKPLFSARTNKSKEKLPYPEASRWLIHLNGFDAAFKTDNKDAGGIS
TGWLGQLGVYIKGNLNFETLLNFPLRKYNEEPWDSSNATWELEPRKERTIIPVPSNGEEL
FTLQSRILLVHENNMVIGSEVIGGDHFLSRDAFNEMPTLWKVKKDRQKDFHLPMEQNTSKQ
FWRDFTALLVENEDAQRPGILKWLDLIFRNIVPHNIKICTLNVMYGSQSSGIKEVWGDVSVN
AGLISTLNIAWSHTIIKLLKLTDTIAKCVGTAAADLELCKGLDEGEKKKNYKKIAGEMKEYPYLQLD
NPFRTWLAANPANKEHDYMNKTMEQWIETAKRIVLKIGKELVDQAGPQAFIGNKDMNAIKAYR
KFRYTIRKILKEGGFV*

Cascade - CasD

>cascade-CasD_P31_assembly_scaffold_59_5|rk-newtermite-P31_Clostridiales_42_22|rk-newtermite-P31 # 6538 # 7248 # 1

MNTLLIRLAAPLQSWGSESKFERRTTQREPTKSGVIGILAAALGRGREDCLNDLTALRFGVRVD
QPGQIMRDYHIARFSKSKEMLERTYVTNRYYISDAVFLVGLGDKHTLEMLLESALLSPFYPIFLG
RRSCPPTGRVVMGIRNKPLEETLEDEEWIAGEWYKKREKSCNLTVLDSKELSANRRRDLPLS
FSQDYRKHKYRYIDDIPDKVVFVQYANIANIATEHDPFDVTWESMEV*

>cascade-CasD_P32_assembly_scaffold_10554_4|rk-newtermite-P32_Actinobacteria_53_11|rk-newtermite-P32 # 1792 # 2493 # -1

MSTLLMRLAAPLQSWGSDSKFEERRATMREPTKSGVVGMLAAALGRRRDEPIEDLSMLRFGV
RIDQPGNLLKDYHTAKSKDGKVSYVTNRHYLADAVFIVGLEGDDTLLHILEKALREPVFPLYLGR
RSCPPTGRISLIGISEESLEKALKQVPWQAGEHLRQKQVSSSEVPQGLTLVLDADGVSNLRRRDL
PLTFSQRHRKHMYRYLVDKPDVAVRVRTKNETEHDAFEAVGKEM*

>cascade-CasD_WG3_assembly_scaffold_10604_2|rk-newtermite-WG3_UNK|rk-newtermite-WG3 # 333 # 1034 # 1

MSTLLMRLAAPLQSWGSDSKFEERRATMREPTKSGVVGMLAAALGRRRDEPIEDLSTLRFGV
RIDQPGNLLKDYHTAKSKDGKVSYVTNRHYLADAVFIVGLEGDDTLLNILEKALREPVFPLYLGR
RSCPPAGRISLIGISEESLEEALKQVPWQAGEHHRKHVSSSEVPQGLTLVLDADGVGNLRRRDL
PLTFSQRHRKHLYRYLVDKPDVAVRVRTKNETEHDAFEAVGKEM*

>cascade-CasD_WG2_assembly_scaffold_2639_3|rk-newtermite-WG2_UNK|rk-newtermite-WG2 # 1370 # 2071 # -1

MSTLLMRLAAPLQSWGSDSKFEERRATMREPTKSGVVGMLAAALGRRRDEPIEDLSTLRFGV
RIDQPGNLLKDYHTAKSKDGKVSYVTNRHYLADAVFIVGLEGDDTLLHILEKALREPVFPLYLGR
RSCPPAGRISLDISEESLEEALKQVPWQAGEHHRKHVSSSEMPQGLTLVLDADGVGNLRRRDL
PLTFSQRHRKHMYRYLVDKPDVAVRVRTKNETEHDAFEAVGKEM*

RAMP

>rampP13_assembly_scaffold_5728_3|rk-newtermite-P13_UNK|rk-newtermite-P13 # 1518 # 1970 # 1

MSLFKNLLETYEKCKGAAGIVQYDANGEANEKKTFLPVFHMTFKSQICITLDSDGGSFVNATRDG
KDITIIIPCTDSSAGRSSGIAAHPLCDQLDYVGGIKNSSVSKRVGCVGETSFRHIKSQVESDSRLC
KHRNYDCRLGEQADIQRLRI*

>rampP33_assembly_scaffold_754_3|rk-newtermite-P33_Actinobacteria_56_10|rk-newtermite-
P33 # 866 # 2815 # 1

MSLFKNLLETYEKCKGAAGIVQPDANGEVNEKKTFLPVFHMTFKSQICITLDSDGGFVNATRDG
KDITIIIPCTDSSAGRSSGVAAHPLCDQLDYVGGISVDKKTAYLKELAAWEKQASGISQAKLKAIH
DYVSTGTMIADLENKQIFKDSEYDAQNTLDNGKVRKIGVRFVEINNDKIPNVWEDNDRKSW
IDYIRPQSSDALFDYLCGARVQIAKQHPKNINSMTGNAKLLSCNDMSGYTFRGRFSSQDDA
VIVDYEQSQKIHQTLRWLINNYGYNIDSQTIIAWAVDTETKPEAEFYKDSYNMFSSMPSAKTDQ
DKLSDAAGEVDANYAKKLNLFQGYGSTEKIKQHARKICIAIFDAATTGRMGLTFYQEMPQDVY
LENIVDWHEDTSYYLTAWIKEKDENGKEKSVPLHYIGAPSYDDILFAVYGKARGGNDAGYNTLK
RKVRKQLLECMFGNFAPKSMVDMAAVRASRPTSFDESDWNRSISITCALARKYEEEEIEITL
LDEARRDRDYLYGRLLAVADRLEQTAMYRAGKQDNRAVRLMAAFVAVKPYHTWGVVHQQ
LVPYINQLNKDRPGSAKYYQSIIDDIMVQLAENYEDNTPLSPLYLLGFSLQRRAFVNKNKKNLE
DNENDTTEQN*

>rampP33_assembly_scaffold_9361_2|rk-newtermite-P33_UNK|rk-newtermite-P33 # 238 # 918
1

MYAKLRITGVLEVATGLHIGASGAFSAIGAVDSPIIKDALES DKPLIPGSSLK GKIRTLARQY
SGKTMGKPENDSDIILRFLGCSKNKDYRTARLIFSDMMLANAVELRKRGVNDL TEVKFENTINR
LTAEANPRQIERAIRGSRFNLDIIEVAGQDESEITEDFALIADGLKLLQCDYLGNGSRGYGKV
VISDLRADVAAGDVNESLVEKVS NLLTEAV*

>rampWG3_assembly_scaffold_4662_2|rk-newtermite-WG3_UNK|rk-newtermite-WG3 # 510 #
1910 # -1

MNVWEDSNLRQSWIDYIRPQGS DALFDYLCGAPVEQIANQHPKNINSMTGNAKLLSCNDMS
GYTFRGRFSNQDDAVIVDYEQSQKMHQTLRWLISNYGYNIDSQTIIAWAVDTETKPEAEFYKDS
YNMFSSMPSAKTDQDKLSDVAGGVDANYAKKLNLFQGYGSADKIKQHAKKICLAVFDAATTG
RMGLTFYQEMPQDVYLENIVDWHEDASYLTAWTKAKGENGKEKSVPIHYIGAPSYDDILFAVY
GKARGGNDAGYNTLKRKRKQLLECMFGNFAPKSMVDMAAVRASRPTSFDESDWNRSISI
TCALARKYKQKKEELTELEDETRDRDYLYGRLLAVADRLEQTAMYRAGKQDNRTTNAVRL
MAAFVAVKPYHTWGVYIYQQLLPYINQLNGAGYYQAIMDEIMVQLAENYENAPLSPLYLLGFSSQ
RRAFVNKNKKNLEDDENDTTEQN*

>rampWG3_assembly_scaffold_11111_3|rk-newtermite-WG3_UNK|rk-newtermite-WG3 # 822 #
1502 # 1

MYAKLRITGVLEVVTGLHIGASGAFSAIGAVDSPIIKDALES DKPLIPGSSLK GKIRTLARQY
SGKTMGKPENDSDIILRFLGCSKNKDYRTARLIFSDMMLANAVELRKRGVNDL TEVKFENTINR
LTAEANPRQIERAIRGSRFNLDIIEVAGQDESEITEDFALIADGLKLLQCDYLGNGSRGYGKV
VISDLRADVAAGEVNESLVEKVS NLLIESV*

>rampWG2_assembly_scaffold_0_97|rk-newtermite-WG2_Firmicutes_43_74|rk-newtermite-
WG2 # 109444 # 110061 # -1

MYTKILYEGEIEVETGLHIGTSGAFSAIGAVDSPVIKDQVTGQPIIPGSSLK GKMRSLARSMSDT
YKTQSHKDDPEEIRFLFGDTEEFHSRLQFFDCFLKNKDELKQVLPTEVKFENTIDRLTSVAN
PRQIERVIPGAKFTFKMTYELADDYKDQLEKDKENIDKAILLENDYLG GGGTRGNGRVVFN
TYD TVFGDDVP*

>rampWG2_assembly_scaffold_2297_4|rk-newtermite-WG2_UNK|rk-newtermite-WG2 # 967 #
2910 # 1

MSLFKNLLETYEKCKDAAGIVKYDANGEANEKKTFLPVFHMTFKSQICITLDSTGRFVNAMRDS
KDTTIIIPCTESGAGRASGIAAHPLCDQLDYVGGINTAKAAAYLKELTAWEKQASGISKAKLKAIH

DYVSTGTMIADLESKQIFKDSEYDTQNKTLDNGKIRKIGVRF AVEINNDKTPNVWEDNDLRKS
WIDYIRPQGGENALFDYLCGMRVEQVANQHPKNINSATGNAKLLSCNTEYGFYVGRFSNQDEA
VVDYEQSQKMHQTLRWLINNHGYNIDSQTIIAWAVDTETNPEVDFHKDSFNMFSMPSAKTTQ
DKLSDAEAGEVDANYAKKLKNLFQGYGYADKIKKHANKICIAIFDAATTGRMGLTFYQEMPQDMY
LENIVDWHEDASYLTAWTKAKGENRKEKSVPIHYIGAPSYDDILFAVYGKARGGNDAGYNTLK
RKVRKQLLECMFGNFAPKSMVDMATVRASRPTSFDESDWNRSISITCALARKYYYQQKKEE
LTLEDETRKDRDYLYGRLLAVADRLEQTAMYKAGKQDNRTTNAVRLMAAFVAVKPYHTWGVY
YQQLLPYINQLNGAGYYQAIMDEIMVQLAENYENNAPLSPLYLLGFSAQRRRAVFNKNKKNLEDD
ENDTTEQN*