



CENTRO UNIVERSITÁRIO DE BRASÍLIA
FACULDADE DE CIÊNCIAS DA EDUCAÇÃO E SAÚDE
GRADUAÇÃO EM BIOMEDICINA

THIFANY VILELA PURCENA

USO DE GRUPOS ORTÓLOGOS E BUSCA POR ILHAS DE PATOGENICIDADE NA
IDENTIFICAÇÃO DE TRÊS ESPÉCIES DE *Bacillus*

Trabalho de conclusão de curso apresentado em formato de artigo científico ao UniCEUB como requisito final para a conclusão do curso em bacharelado para Biomedicina, sob a orientação do professor Paulo Roberto Queiroz.

BRASÍLIA

2020

Dedicatória

Começo dedicando esse trabalho primeiramente a pessoa que me fez pegar gosto pela pesquisa, essa mulher guerreira, mãe, pesquisadora, esposa, filha, irmã e acima de tudo uma orientadora que nunca me deixou na mão! Essa mulher me inspira, ela não tem noção o quanto me ensina é uma tremenda honra ser sua pequena orientanda. Ela me ensinou que lugar de mulher é sim na pesquisa, e que podemos comandar uma equipe inteira e ainda cuidar de uma família. Dedico esse trabalho a minha primeira orientadora Priscila Grynberg.

Em segundo, dedico esse trabalho ao meu orientador, senhor Roberto Coiti Togawa, ao qual ensinou toda a parte da informática e tecnológica, que praticamente me adotou no laboratório, me acalmou em dias de turbulências e me socorreu quando eu precisava.

E ao Paulo Roberto Queiroz, que para sempre será meu eterno professor de genética e biologia molecular, ao qual me ensinou todas as técnicas e teorias que hoje eu domino, e o responsável por me dar dois orientadores incríveis como esses, e por ser esse orientador incrível que é.

Agradecimentos

Começo agradecendo a minha pequena estrelinha, que nunca me abandonou e sempre me deu força, sempre esteve ao meu lado, nos melhores e piores momentos. Agradeço aos meus pais, por todo apoio dado e por não me deixado desistir, mesmo nos dias que eu liguei em desespero e pedindo para voltar para perto deles.

Agradeço as minhas avós, principalmente a minha avó Almira, que sempre me falou que estava a cada dia mais perto do fim, que não me deixou jogar tudo para cima, me deu força, coragem e brigava comigo quando eu precisava.

Meu muito obrigada a todos os meus amigos, que nunca desistiram de mim, me deram um colo quando eu precisei, um abraço, uma palavra, enxugaram minhas lágrimas, me deram forças, tiraram sorrisos e esporros quando eu precisava. São minha família, a família que eu escolhi.

Um agradecimento especial a minha vovó de Brasília, obrigada Dona Regina! Por tudo que me proporcionou, por todo o carinho, todos os conselhos, todo colo que a senhora me deu, todas as conversas que tivemos e todos os momentos. Vai ser minha eterna vovó de Brasília.

Obrigada a todos os meus professores, orientadores, a todos que tiveram paciência comigo nessa jornada, se cheguei até aqui foi com a ajuda de vocês, pois sozinha não seria capaz! Obrigada por todo o ensinamento passado.

“Essa é a minha família, eu achei. Sozinho. Eu achei. É pequena e incompleta. Mas é boa. É, é boa. Ohana quer dizer família, família quer dizer NUNCA abandonar, ou esquecer.”
- Lilo e Stitch

Uso de grupos ortólogos e busca por ilhas de patogenicidade na identificação de três espécies de *Bacillus*

Thifany Vilela Purcena¹
Paulo Roberto Queiroz¹
Priscila Grynberg²

Resumo

B. anthracis, o agente etiológico da doença antraz, *B. cereus*, uma bactéria que pode causar infecções gastrointestinais e *B. thuringiensis*, produtora de cristais tóxicos (Cry) que atuam como inseticidas espécie-específicos naturais são bactérias muito parecidas geneticamente entre si. Diferenciá-las através de métodos moleculares é um grande desafio. O objetivo deste trabalho foi aplicar abordagens de genômica comparativa para classificar estas espécies. Para isso, as proteínas preditas de 45 genomas sendo 10 genomas de cada espécie disponíveis publicamente no NCBI, 14 genomas de *B. thuringiensis* do banco de dados interno da Embrapa e uma espécie *outgroup* foram usadas para a identificação de grupos ortólogos e confecção de uma árvore filogenética, totalizando 268.898 proteínas. A análise filogenética separou, em termos gerais, *B. anthracis* das demais, ainda que as três espécies se misturem entre si em maior ou menor grau. Cepas de *B. thuringiensis* com ação bio-inseticidas de lepidópteros ficaram separadas daquelas que atuam em coleópteros. *B. anthracis* é a espécie que apresenta, em média, o menor número e *B. thuringiensis*, o maior número de ilhas de patogenicidade por genoma. Este trabalho mostrou que através da classificação em grupos ortólogos e análise de ilhas de patogenicidade é possível melhorar a classificação destas três espécies baseado em dados moleculares.

Palavras-Chave: *Bacillus thuringiensis*, *Bacillus cereus*, *Bacillus anthracis*, ortologia, filogenia, genes marcadores

Use of orthologue groups and search for pathogenic islands in the identification of three *Bacillus* species

Abstract

Bacillus anthracis, the etiological agent of Anthrax disease, *B. cereus*, a bacterium which causes gastrointestinal infections and *B. thuringiensis*, a producer of toxic crystals (Cry) proteins that act as natural species-specific insecticides are bacteria very similar to each other. The identification of these species through molecular methods is a big challenge. We aim to test comparative genomics approaches to correctly classify them. The predicted proteins of 45 genomes related to 10 genomes of each species available on the NCBI, 14 genomes of *B. thuringiensis* from Embrapa's internal database and an outgroup species were used to identify orthologous groups and make a phylogenetic tree totaling 268,898 proteins. Phylogenetic analysis generally separated *B. anthracis* from the others, even though the three species are mixed to a greater or lesser degree. Strains of *B. thuringiensis* with bioinsecticidal action by lepidopterans were separated from those that act on coleopterans. *B. anthracis* is the species that presents, on average, the lowest number and *B. thuringiensis*, the highest number of islands of pathogenicity by genome. This work showed that through the classification in orthological groups and analysis of islands of pathogenicity it is possible to improve the classification of these three species based on molecular data.

Keywords: *Bacillus thuringiensis*, *Bacillus cereus*, *Bacillus anthracis*, orthology, phylogeny, molecular markers

¹Acadêmica do curso de Biomedicina do UniCeub

²Professor do UniCeub

³Pesquisadora da Embrapa Recursos Genéticos e Biotecnologia

1. Introdução

O grupo *Bacillus cereus sensu lato* é formado por bactérias do filo Firmicutes, gram-positivas, que possuem baixo conteúdo GC em seus genomas. São encontradas no solo, em vegetais e em águas (ALTAYAR; SUTHERLAND, 2006; EHLING-SCHULZ; LERECLUS; KOEHLER, 2019), podem colonizar insetos e mamíferos (CEUPPENS; BOON; UYTENDAELE, 2013), formam esporos, são aeróbicas ou anaeróbicas facultativas e possuem formato de bastonetes. Ao menos oito espécies fazem parte deste grupo: *Bacillus anthracis*, *Bacillus cereus*, *Bacillus thuringiensis*, *Bacillus mycoides*, *Bacillus pseudomycoides*, *Bacillus weihenstephanensis*, *Bacillus cytotoxicus* e *Bacillus toyonensis*. Há mais quatro espécies que podem fazer parte desse grupo: *Bacillus gaemokensis*, *Bacillus manliponensis*, *Bacillus bingmayongensis* e *Bacillus wiedmannii*. As espécies de maior importância na saúde humana, na indústria de alimentos e na agricultura são *B. anthracis*, *B. cereus* e *B. thuringiensis* (RASKO et al., 2005).

B. cereus ou *B. cereus sensu stricto* (s.s) é um patógeno oportunista associado com intoxicação alimentar, que pode ser emética ou diarreica. Geralmente estas intoxicações alimentares são autolimitantes, ainda que há relatos de mortes em recém-nascidos e pessoas com imunodeficiências (GLASSET et al., 2018). A síndrome emética é o resultado da intoxicação causada pela toxina cereulida depsipeptídea, pré-formada em alimentos contaminados. Já a síndrome diarreica tem sido associada com diferentes enterotoxinas, que possivelmente são produzidas após o consumo de esporos que crescem e acabam sendo ativados no intestino delgado (MENDES et al., 2004).

Os esporos do *B. cereus* são utilizados na indústria de alimentos por serem termorresistentes e, por isso, capazes de sobreviver a processamentos térmicos. Em condições favoráveis eles podem germinar e se multiplicar nos alimentos (MAZIERO, 2011). Porém não são todas as cepas de *B. cereus* que produzem estas toxinas (EHLING-SCHULZ; LERECLUS; KOEHLER, 2019a). Uma cepa de *B. cereus*, por exemplo, não é patogênica e é capaz de produzir várias enzimas de interesse econômico, como a metaloprotease alcalina, de natureza não colagenolítica e não queratinolítica, com possível aplicação na indústria do couro (NILEGAONKAR et al., 2007).

B. anthracis, o agente causador da doença antraz, é mencionado em documentos históricos antes mesmo da era comum, mas começou a ser estudado apenas no século XIX, quando pela primeira vez pesquisadores associaram uma doença a um microorganismo (TURNBULL, 2002). As cepas altamente virulentas carregam, obrigatoriamente, dois grandes plasmídeos: pX01, que codifica as toxinas anthrax e pX02, que codifica as enzimas para a

biossíntese da cápsula (GREEN et al., 1985; MIKESELL et al.; 1983). Ambos são considerados os principais fatores de virulência de *B. anthracis*. A toxina é tripartida, ou seja, composta por três proteínas: antígeno protetor (AP), fator letal (FE) e fator de edema (FL) (BROSSIER; MOCK, 2001). O antígeno protetor facilita a entrada do fator de edema e o fator letal nas células hospedeiras. Juntas essas toxinas desabilitam o sistema imunológico adaptativo do hospedeiro para induzir a doença letal (BHATNAGAR; BATRA, 2001). Estudos identificaram três genes que codificam a cápsula: *capB*, *capC* e *capA*, nesta ordem.

Descoberto em 1911 na província de Thuringia na Alemanha por Ernst Berlin, *B. thuringiensis* passou a ser utilizada como agente de controle de insetos na França em 1938 e na década de 1950 nos Estados Unidos (SÁ, 2015). As estirpes de *B. thuringiensis* produzem inclusões cristalinas parasporais também conhecidas como δ -endotoxinas. As δ -endotoxinas correspondem às proteínas que formam cristais (Cry) e citolíticas (Cyt). Essas proteínas com potencial tóxico são produzidas na fase de esporulação de crescimento da bactéria (PALMA et al., 2014). Algumas cepas de *B. thuringiensis* são capazes de produzir proteínas solúveis que são secretadas no meio extracelular. Essas proteínas correspondem às Proteínas Inseticidas Vegetais (vip) que são produzidas e secretadas durante a fase vegetativa. Convém ressaltar que as proteínas Cry e Vip possuem sequências de aminoácidos diferentes e são produzidas em fases diferentes do crescimento do *B. thuringiensis* em meio de cultura (EHLING-SCHULZ; LERECLUS; KOEHLER, 2019b).

As proteínas Cry despertam interesse biotecnológico devido ao seu potencial de ação tóxica contra várias ordens de insetos (Coleoptera, Diptera, Hemiptera e Lepidoptera) e outros organismos como, por exemplo, nematóides e fungos (Palma et al., 2014). As toxinas vip atuam contra coleópteros e lepidópteros, enquanto as proteínas Cyt apresentam atividade inseticida contra dípteros (CRICKMORE, 2018). Além das proteínas citadas anteriormente, algumas cepas de *B. thuringiensis* podem produzir as toxinas Sip (Proteína Inseticida Secretada) que possuem toxicidade contra larvas de coleóptero. Com relação ao modo de ação dessas toxinas, já é conhecido que as toxinas Cry são responsáveis pela formação de poros na membrana da célula que forma o epitélio intestinal do inseto. O resultado dessa ação é choque osmótico e consequentemente lise celular. As proteínas Cyt possuem mecanismo semelhante de ação às toxinas Cry. Já para as proteínas Vip e Sip o mecanismo de ação ainda não é completamente esclarecido (DONOVAN et al., 2006).

A ação dessa toxina é por meio da interação específica com receptores específicos, (aminopeptidase, fosfatase alcalina e caderina) encontrados no intestino médio do inseto. A ação do pH alcalino no intestino médio do inseto resulta na ativação das toxinas Cry e Cyt consequentemente na morte, devido à formação de edema e ruptura das células (BRAVO;

GILL; SOBERÓN, 2007; PARDO-LÓPEZ, SOBERÓN; BRAVO, 2013; ZHANG et al., 2006). Outros organismos invertebrados, mamíferos, pássaros, répteis e anfíbios não são sensíveis à estas toxinas (BETZ; HAMMOND; FUCHS, 2000).

A diferenciação das espécies pertencentes ao grupo *B. cereus lato sensu* é complexa em função da proximidade filogenética entre elas. Abordagens fenotípicas e genéticas, incluindo métodos moleculares têm sido usadas na tentativa de diferenciar e classificar corretamente sequências de *B. anthracis*, *B. cereus* e *B. thuringiensis*. Um trabalho buscou classificar *B. cereus* e *B. thuringiensis* de acordo com a origem do isolado (alimento, médico ou bio-pesticida) mas nenhuma metodologia foi capaz de separar satisfatoriamente as amostras baseadas em suas origens ambientais (MANZANO et al., 2009). Outra tentativa foi através da amplificação da região espaçadora intergênica polimórfica 16S-23S, mas falhou ao não conseguir diferenciar *B. cereus* de *B. thuringiensis* (ANTOLINOS et al., 2012; VOLOKHOV et al., 2012).

A subunidade 16S do RNA ribossomal também já foi largamente testada. Apesar de ser um ótimo marcador molecular capaz de identificar bactérias em nível de espécie em outros gêneros, este método se mostrou ineficiente em classificar as três espécies de *Bacillus* corretamente (ASH; COLLINS, 1992; VOLOKHOV et al., 2012). Um trabalho chegou a analisar 1007 sequências de 16S rRNA de espécies pertencentes ao grupo *B. cereus* concluindo que a resolução deste gene é muito baixa para discriminar espécies muito próximas entre si (LIU et al., 2015). Assim, buscou-se novos marcadores moleculares (YAMADA et al., 1999). O gene codificador da subunidade beta da RNA polimerase (*rpoB*) também já foi bastante explorado em tentativas de determinar um bom marcador molecular (VOLOKHOV et al., 2012).

Outros trabalhos usaram iniciadores (*primers*) para a amplificação dos genes *Cry*, presentes apenas em *B. thuringiensis* (NOGUERA; IBARRA, 2010). No entanto, essa abordagem é limitante porque a presença confirma que a amostra é *B. thuringiensis*, mas a ausência não confirma que a amostra não é. Seguindo a mesma linha, o gene codificador da toxina cereulide, produzidas por cepas patogênicas de *B. cereus*, foi amplificado para identificar esta espécie. O uso deste gene como marcador molecular não identifica cepas de *B. cereus* não patogênicas. O mesmo ocorre em *B. anthracis* e seus genes codificadores de toxinas que estão presentes no plasmídeo pX01. Se a cepa de *B. anthracis* não possuir este plasmídeo não é possível distingui-lo com os demais *Bacillus*.

A facilidade em se obter genomas completos de bactérias em função dos métodos atuais de sequenciamento abriu espaço a genômica comparativa onde podemos explorar os

grupos de genes ortólogos para realizar inferências filogenéticas e filogenômicas entre amostras. Com a descoberta da molécula de DNA e todo o conhecimento de genética antes e após a esse fato, podemos considerar que genes homólogos são aqueles que compartilham um ancestral comum (SHEIKHIZADEH ANARI et al., 2018). Devido a isso, a homologia não apresenta níveis intermediários pois está relacionada à evolução: é ou não é (REECK et al., 1987).

Existem três classes distintas de genes homólogos: ortólogos, parálogos e os xenólogos. Genes ortólogos são genes homólogos que divergiram através da especiação. Por isso são os objetos ideais para a reconstrução de relações filogenéticas entre espécies e estudar a evolução de genomas. Genes parálogos são aqueles que descendem de um evento de duplicação, independentemente de eventos de especiação. Genes xenólogos são aqueles adquiridos por transferência horizontal, ou seja, o gene em algum momento foi absorvido do meio para dentro do genoma do organismo (através de vírus, por exemplo) (CALDART et al., 2018; DESCORPS-DECLÈRE et al., 2008).

Os plasmídeos exercem um papel fundamental na virulência destas espécies pois eles carregam muitos genes codificadores das toxinas citadas anteriormente. O material genético plasmidial é muito suscetível à transferência horizontal gênica (HGT – *horizontal gene transfer*), que pode ocorrer ainda por outros mecanismos como transposons e classes não canônicas de elementos genéticos móveis (MGE – *mobile genetic elements*) (BELLANGER et al., 2014). Os MGEs que carregam traços adaptativos são comumente chamados de ilhas genômicas. Acredita-se que elas contribuam para a diversificação e adaptação dos microrganismos, tendo um impacto na plasticidade e evolução do genoma, na disseminação de resistência a antibióticos, de genes de virulência, e na formação de vias catabólicas (JUHAS et al., 2009). As ilhas genômicas podem ser subclassificadas em ilhas de resistência, ilhas metabólicas, ilhas simbióticas, ilhas fitness e ilhas de patogenicidade (DOBRINDT et al., 2004; SCHMIDT; HENSEL, 2004).

As ilhas genômicas, em geral, fazem parte do pool genético flexível ou dos conjuntos de genes que codificam características que podem ser favoráveis sob certas condições ambientais. Elas podem variar muito em sequência e composição de uma espécie bacteriana para outra, e mesmo dentro de uma espécie. Essa variabilidade e plasticidade permitem que as bactérias se adaptem a vários ambientes (DOBRINDT ET AL., 2004). Além disso, as ilhas genômicas fazem parte de um grupo maior chamado elementos genéticos móveis, que são segmentos de DNA capazes de se mover dentro de um genoma e, também, de serem transmitidos entre células bacterianas individuais. Novas ilhas genômicas podem integrar-se

no genoma bacteriano central ou em um plasmídeo pré-existente (HALLSTROM; MCCORMICK, 2014).

A diferenciação das espécies pertencentes ao grupo *B. cereus lato sensu* é complexa em função da proximidade filogenética entre elas. Abordagens fenotípicas e genéticas, incluindo métodos moleculares têm sido usadas na tentativa de diferenciar e classificar corretamente sequências de *B. anthracis*, *B. cereus* e *B. thuringiensis*. No passado, as bactérias deste grupo eram classificadas em diferentes espécies de acordo com as sequências de 16S rRNA e características como a presença ou ausência de plasmídeos contendo genes de virulência e características patogênicas (AGATA; OHTA; MORI, 1996; RASHO et al., 2005). Um trabalho buscou classificar *B. cereus* e *B. thuringiensis* de acordo com a origem do isolado (alimento, médico ou bio-pesticida) mas nenhuma metodologia foi capaz de separar satisfatoriamente as amostras baseadas em suas origens ambientais (MANZANO et al., 2009). Outra tentativa foi através da amplificação da região espaçadora intergênica polimórfica 16S-23S, mas falhou ao não conseguir diferenciar *B. cereus* de *B. thuringiensis* (ANTOLINOS et al., 2012; VOLOKHOV et al., 2012).

Ao longo dos últimos seis anos, 14 estirpes de *B. thuringiensis* foram sequenciadas pelo grupo de prospecção de genes tóxicos a insetos em estirpes de *B. thuringiensis*. O Laboratório de Bioinformática da Embrapa Recursos Genéticos e Biotecnologia foi responsável por analisar estes genomas: identificar plasmídeos e ao menos 90 genes codificadores de toxinas, como as Cry (crystal toxin), Cyt (cytotoxin), Vip (vegetative insecticidal protein) e Sip (secreted insecticidal protein), além de outros marcadores moleculares através do desenvolvimento de uma metodologia de mineração de dados. Porém, todo o trabalho foi realizado analisando as estirpes individualmente. Algumas perguntas surgiram nesse caminho que nos levaram a desenvolver este trabalho: há genes específicos das espécies que compõem do grupo *B. cereus sensu lato* que podem ser usados como marcadores moleculares para a correta identificação das espécies? A proximidade filogenética das estirpes de *B. thuringiensis* reflete nos insetos-praga alvos? A presença e organização das ilhas patogênicas podem trazer informações relevantes que auxiliem no desenvolvimento de uma metodologia que aprimore a identificação e classificação destas espécies?

Assim, objetivamos testar três abordagens diferentes e complementares, sendo elas a busca por genes candidatos a marcadores moleculares, a genômica comparativa por grupos ortólogos e a identificação de ilhas de patogenicidade para compreender melhor o potencial destes métodos para a correta classificação molecular destas espécies.

2. Material e Métodos

O trabalho foi desenvolvido no Laboratório de Bioinformática da Embrapa Recursos Genéticos e Biotecnologia.

2.1. Busca por genes candidatos a marcadores moleculares

Os genes 16S rRNA, *gyrB* e *rpoB*, CapA, CapB e CapC, toxina anthrax e Cry foram escolhidos para serem testados como marcadores moleculares para diferenciar *B. cereus*, *B. anthracis* e *B. thuringiensis*. Para isso, sequências nucleotídicas e de proteínas dos oito genes das três espécies foram obtidas no site NCBI para a criação de um banco de dados local, que foi criado com o programa makeblastdb com os seguintes parâmetros: "makeblastdb -in -dbtype [nucl ou prot] -parse_seqids -out -title "Database title". A busca desses genes candidatos nos genomas de interesse foi feita através dos programas blastp e blastn e por meio de linha de comando, com e-value de 10E20 e *output* nos formatos de alinhamento e tabular. O resultado foi analisado em função das métricas de identidade e e-value obtidos.

2.2. Genômica comparativa usando o OrthoFinder

Estudos de genômica comparativa foram feitos com o programa OrthoFinder V.2.3.3 (EMMS E KELLY, 2019) através da busca por ortogrupos e genes ortólogos. Para compreender melhor como as três espécies se relacionam entre si, 10 proteomas cromossomais e plasmidiais de *B. cereus*, 10 de *B. anthracis* e 24 de *B. thuringiensis* (sendo 10 do NCBI e 14 cepas internas da Embrapa) e um genoma *outgroup* (*Vulcanibacillus modesticaldus*) foram incluídas nesta análise (Tabela 1), totalizando 264.060 proteínas. As estipes selecionadas para esta análise foram as que possuíam proteomas preditos (sequências de proteínas na base de dados do NCBI). O programa foi rodado com os seguintes parâmetros: -t 10 -M mas -l 1.8.

O alinhamento múltiplo entre as sequências dos ortogrupos foi feito pelo MAFFT (ROZEWICKI ET AL., 2019). A inferência filogenômica foi feita pelo software IQ-Tree (TRIFINOPOULOS ET AL., 2016), cujo input foi o output obtido pelo MAFFT. O arquivo de saída foi carregado e visualizado na plataforma web iTOL (LETUNIC E BORK, 2019).

Tabela 1: Número de acesso do NCBI dos genomas usados na análise de ortologia

Cepas	Número de acesso no NCBI	Cepas	Número de acesso no NCBI
<i>B. anthracis</i> MCCC 1A01412	CP031643.1	<i>B. thuringiensis s gallerie</i>	CP010089.1
<i>B. anthracis</i> MCCC 1A2161	CP031642.1	<i>B. thuringiensis s kurstaki</i>	CP004870.1
<i>B. anthracis</i> Strain 200031021	CP007618.1	<i>B. thuringiensis</i> HD 1011	CP009335.1
<i>B. anthracis</i> Strain 2002013094	CP009902.1	<i>B. thuringiensis</i> HD771	CP018500.1
<i>B. anthracis</i> Ames	CP003997.3	<i>B. thuringiensis</i> HD12	CP014847.1
<i>B. anthracis</i> Strain CDC 684	CP012581.1	<i>B. thuringiensis</i> HD73	NC_020238.1
<i>B. anthracis</i> Friedrick	CP023001.1	<i>B. thuringiensis s indiana</i>	CP010106.1
<i>B. anthracis</i> SK-102	CP009464.1	<i>B. thuringiensis</i> HD789	CP018508.1
<i>B. anthracis</i> Sterne	CP005945.1	Strain 1930	NCBI (submetido)
<i>B. anthracis</i> Vollum	CP007666.1	Strain 2205	NCBI (submetido)
<i>B. cereus</i> ATCC 14579 INRA	CP004722.1	Strain 2728	NCBI (submetido)
<i>B. cereus</i> ATCC 14579 Zhejiang	CP034551.1	Strain 2795	NCBI (submetido)
<i>B. cereus</i> Strain 03BB87	CP009941.1	Strain 2744	NCBI (submetido)
<i>B. cereus</i> Strain CMCC P0021	CP011151.1	Strain 2193	NCBI (submetido)
<i>B. cereus</i> Strain G1-1	CP042929.1	Strain 1905	NCBI (submetido)
<i>B. cereus</i> GMCC 0460.4	CP011153.1	Strain 1785	NCBI (submetido)
<i>B. cereus</i> Hebei University	CP023245.1	Strain 1287	NCBI (submetido)
<i>B. cereus</i> Strain M13	CP016360.1	Strain 1307	NCBI (submetido)
<i>B. cereus</i> Strain MLY 1	CP024655.1	Strain 601	NCBI (submetido)
<i>B. ceureus</i> Strain ZB201708	CP030982.1	Strain 908	NCBI (submetido)
<i>B. thuringiensis</i> HD 571	CP009600.1	Strain 906	NCBI (submetido)
<i>B. thuringiensis</i> HD 682	CP009720.1	Strain 907	NCBI (submetido)

2.3. Predição de ilhas patogênicas

A predição de ilhas patogênicas foi feita com o programa web IslandViewer v.4 (BERTELLI et al., 2017). Este programa integra quatro métodos diferentes de predição de ilhas patogênicas: IslandPick, IslandPath-DIMOB, SIGI-HMM e Islander. Todos os genomas completos de bactérias e arqueobactérias disponíveis para download no NCBI já são pré-computados nesta plataforma. Assim, para as estirpes usadas neste trabalho que foram baixadas do NCBI, fizemos o download direto dos resultados.

Os genomas das estirpes da Embrapa foram carregados na plataforma. O programa exige a escolha de um genoma de referência para fazer a predição. Assim, para cada estirpe interna escolhemos como referência a mais próxima filogeneticamente de acordo com o resultado do OrthoFinder (Tabela 2).

Tabela 2: Genoma de referência escolhido para a predição de ilhas patogênicas pelo programa IslandViewer 4

Estirpe Embrapa	Genoma de referência
Strain 1930	<i>B. thuringiensis</i> kurstaki HD1
Strain 2205	<i>B. cereus</i> strain 03BB81
Strain 2728	<i>B. thuringiensis</i> kurstaki HD1
Strain 2195	<i>B. thuringiensis</i> kurstaki HD1
Strain 2744	<i>B. thuringiensis</i> kurstaki HD1
Strain 2193	<i>B. thuringiensis</i> s <i>galleriae</i>
Strain 1905	<i>B. thuringiensis</i> kurstaki HD1
Strain 1785	<i>B. thuringiensis</i> HD 789
Strain 1287	<i>B. thuringiensis</i> HD 12
Strain 601	<i>B. thuringiensis</i> kurstaki HD1
Strain 908	<i>B. thuringiensis</i> kurstaki HD1
Strain 906	<i>B. thuringiensis</i> kurstaki HD1
Strain 907	<i>B. thuringiensis</i> kurstaki HD1
Strain 1307	<i>B. thuringiensis</i> kurstaki HD1

O programa gera uma tabela com as localizações genômicas de cada ilha, assim como os genes presentes em cada uma delas.

3. RESULTADO

3.1. Genes candidatos a marcadores para auxiliar a diferenciação das três espécies por métodos moleculares

Duas classes gênicas foram escolhidas para serem testadas com o objetivo de diferenciar as três espécies entre si, e em especial, identificar corretamente cepas de *B. thuringiensis*: genes housekeeping, ou seja, genes conservados e presentes em todas as espécies de bactérias e genes codificadores de toxinas, de expressão mais específica. Os genes housekeeping selecionados foram a subunidade 16S do RNA ribossomal (*16S rRNA*), a subunidade beta da girase (*gyrB*) e a subunidade B da RNA polimerase (*rpoB*). Os resultados mostraram que estes genes não são apropriados para serem usados como marcadores moleculares, pois possuem uma resolução muito baixa. Encontramos identidades de entre 99% e 100% entre genes de espécies diferentes, ou seja, são muito conservados entre si (Tabela 3).

Tabela 3: Identidade (%) encontrada nos alinhamentos dos genes housekeeping em cada espécie

Gene	<i>B.thuringiensis</i>	<i>B. cereus</i>	<i>B. anthracis</i>
<i>B. anthracis</i>			
<i>gyrB</i>	99%	99%	100%
<i>rpoB</i>	100%	100%	100%
16S rRNA	99%	99%	100%
<i>B. cereus</i>			
<i>gyrB</i>	99%	100%	99%
<i>rpoB</i>	100%	100%	100%
16S rRNA	100%	100%	99%
<i>B. thuringiensis</i>			
<i>gyrB</i>	100%	99%	99%
<i>rpoB</i>	100%	100%	100%
16S rRNA	100%	100%	99%

Em face dos resultados pouco esclarecedores, exploramos a capacidade discriminatória dos genes espécie-específicos *capA*, *capB*, *capC*, toxina antrax e os genes *cry*. O gene da *CapA* é conservado em todas as cepas de *B. anthracis* com identidades entre 92% e 96%. Já nas demais espécies este gene pode ser ausente, ser pouco ou bastante conservado (Tabela 4). O gene *CapB*, por outro lado, foi encontrado em apenas três cepas de *B. anthracis* com 100% de identidade. Esse gene curiosamente foi encontrado em uma cepa da Embrapa (Strain 1338) com 65% de identidade. O gene *CapC* foi o menos conservado, com identidades entre 88 e 89% em *B. anthracis*. Já a cepa da Embrapa Strain 1338, esse gene teve 65% de identidade.

Conforme era esperado, os genes *cry* são conservados nas *B. thuringiensis*, presentes em quase todas as cepas anotadas como tal incluindo as cepas da Embrapa, com valores entre 99 e 100% de identidade. No entanto, algumas cepas anotadas como *B. thuringiensis* apresentaram proteínas pouco conservadas (36-37% de identidade), ou não foi encontrado nenhuma proteína similar com toxinas *cry* (0%) (Tabela 4). O gene *Cry* é um marco fenotípico dessa bactéria, e logo deveria estar presente em todas as cepas classificadas como *B. thuringiensis*. Algumas cepas de *B. cereus* e *B. anthracis* possuem traços de proteínas *cry*, com baixa identidade (35% e 36%).

A toxina Anthrax é específico da *B. anthracis*. No entanto as nossas análises mostraram que nem todas as cepas a possuem como por exemplo *B. anthracis* MCCC 1A01412 e *B. anthracis* Sterne. Nas demais o gene foi encontrado com identidades entre 92% e 96%. Curiosamente, a cepa de *B. cereus* Strain 03BB87 possui este gene, com 95% de identidade. Outras cepas, tanto de *B. cereus* quanto de *B. thuringiensis* apresentaram traços desta proteína (Tabela 4).

Tabela 4: Identidade (%) dos genes *capA*, *capB*, *capC*, *Cry* e toxina Antraz encontrados

Cepas	CapA	CapB	CapC	Cry	Toxina Antraz
<i>B. anthracis</i> MCCC 1A01412	96%	0%	0%	36%	0%
<i>B. anthracis</i> MCCC 1A2161	96%	0%	0%	33%	0%
<i>B. anthracis</i> Strain 200031021	92%	100%	88%	36%	0%
<i>B. anthracis</i> Strain 2002013094	92%	100%	88%	36%	96%
<i>B. anthracis</i> Ames	96%	0%	0%	36%	0%
<i>B. anthracis</i> Strain CDC 684	96%	100%	89%	36%	96%
<i>B. anthracis</i> Friedrick	96%	100%	89%	36%	92%
<i>B. anthracis</i> SK-102	96%	100%	89%	36%	92%
<i>B. anthracis</i> Sterne	96%	0%	0%	36%	0%
<i>B. anthracis</i> Vollum	96%	100%	89%	36%	92%
<i>B. cereus</i> ATCC 14579 INRA	90%	0%	0%	0%	0%
<i>B. cereus</i> ATCC 14579 Zhejiang	90%	0%	0%	0%	0%
<i>B. cereus</i> Strain 03BB87	0%	0%	0%	34%	95%
<i>B. cereus</i> Strain CMCC P0021	90%	0%	0%	35%	0%
<i>B. cereus</i> Strain G1-1	89%	0%	0%	0%	0%
<i>B. cereus</i> GMCC 0460.4	90%	0%	0%	35%	0%
<i>B. cereus</i> Hebei University	90%	0%	0%	0%	0%
<i>B. cereus</i> Strain M13	90%	0%	0%	0%	0%
<i>B. cereus</i> Strain MLY 1	27%	0%	0%	0%	32%
<i>B. cereus</i> Strain ZB201708	90%	0%	0%	35%	0%
<i>B. thuringiensis</i> HD 571	92%	0%	0%	36%	0%
<i>B. thuringiensis</i> HD 682	96%	0%	0%	0%	0%
<i>B. thuringiensis</i> s <i>gallerie</i>	89%	0%	0%	100%	0%
<i>B. thuringiensis</i> s <i>kurstaki</i>	90%	0%	0%	99%	0%
<i>B. thuringiensis</i> HD 1011	96%	0%	0%	36%	0%
<i>B. thuringiensis</i> HD 789	91%	0%	0%	100%	0%
<i>B. thuringiensis</i> HD 771	27%	0%	0%	100%	0%
<i>B. thuringiensis</i> HD 12	90%	0%	0%	99%	33%
<i>B. thuringiensis</i> 73	90%	0%	0%	100%	0%
<i>B. thuringiensis</i> s <i>indiana</i>	88%	0%	0%	100%	0%
Strain 601	33%	0%	0%	100%	33%
Strain 1785	0%	0%	0%	100%	0%
Strain 2193	0%	0%	0%	100%	0%
Strain 2205	0%	0%	0%	0%	0%
Strain 908	90%	0%	0%	100%	36%
Strain 906	90%	0%	0%	100%	35%
Strain 907	90%	0%	0%	100%	29%
Strain 1287	27%	0%	0%	100%	31%
Strain 2195	27%	0%	0%	99%	0%
Strain 1905	27%	0%	0%	100%	0%
Strain 1930	92%	0%	0%	29%	0%
Strain 1338	33%	65%	65%	0%	0%
Strain 1307	0%	0%	0%	100%	0%

Strain 2728	93%	0%	0%	37%	0%
Strain 2744	27%	0%	0%	37%	0%

3.2. Análise de ortologia em uma abordagem de genômica comparativa

Com o objetivo de compreendermos melhor as relações filogenéticas entre *B. anthracis*, *B. cereus* e *B. thuringiensis*, foi realizado uma análise de genômica comparativa utilizando o software OrthoFinder (EMMS; KELLY, 2019). Para isso 45 proteomas, totalizando 268.898 proteínas, foram utilizados sendo um de uma espécie distante filogeneticamente (*outgroup*) para auxiliar na topologia da árvore. Destas, 98,9% (265.880) foram incluídas em um dos 11.839 ortogrupos definidos pelo programa. Um total de 516 proteínas estão inseridas em 198 ortogrupos espécie-específicos, ou seja, compostos por proteínas de uma única espécie. Há 914 ortogrupos formados por uma única cópia gênica (Tabela 5).

O total de 1.197 ortogrupos contém proteínas de todas as espécies analisadas, incluindo o *outgroup*. Na ausência da espécie *V. modesticaldus*, esse número sobe para 1.705. Estes ortogrupos contém o *core genes*, ou pan-genoma, que seria o genoma formado por todos os genes de todas as cepas de um clado. Por fim, 3.018 genes não foram classificados em nenhum ortogrupo (Tabela 5). Os números mínimos e máximos encontrados ao analisar as espécies individualmente encontram-se na Tabela 5.

Tabela 5: Informações gerais sobre o resultado da análise de ortologia

Informação	Número total (%)	Mín-Max
Espécies	45	
Genes	268898	5319-7096
Genes em ortogrupos	265880 (98,9%)	5270-7089
Genes sem classificação	3018 (1,1%)	5-263
Ortogrupos	11839	5056-6784
Ortogrupos espécie-específico	198	0-39
Genes em ortogrupos espécie-específico	516 (0,2%)	0-121
Ortogrupos com todas as espécies	1197	
Ortogrupos com todas as espécies exceto outgroup	1705	
Ortogrupos com um único gene	914	

As sequências proteicas incluídas nesta análise foram alinhadas e as inferências filogenômicas foram visualizadas em uma árvore filogenética (Figura 1). A árvore foi enraizada na espécie *V. modesticaldus*. Os valores de *bootstrap* foram, no geral, altos (acima de 90%), o que suporta a topologia encontrada. As cepas foram classificadas, principalmente, em dois grandes clados. O primeiro clado contém todas as cepas de *B. anthracis* juntamente com *B.*

thuringiensis HD571, HD682, HD1011, formando um sub-clado (linhas laranjas) e *B. cereus* 03BB87 e a cepa da Embrapa 2205, formando o segundo sub-clado (linhas verdes).

Nós então analisamos os nossos resultados prévios da busca por genes marcadores moleculares dentro na árvore filogenética, e obtivemos informações interessantes. O gene *capA* foi encontrado em todas as cepas do subclado laranja, além de outras cepas dos demais clados e sub-clados, indicando que este gene provavelmente está presente no ancestral comum a todas as espécies. Os genes codificadores da cápsula *capB* e *capC*, e o gene codificador da toxina antrax estão presentes apenas em cepas de *B. anthracis*.

A *B. anthracis* 2000031021 não apresentou o gene da toxina. Isso pode indicar ausência do plasmídeo pX01. A *B. anthracis* Ames possui genes pouco conservados da toxina antrax, *capB* e *capC*. As cepas *B. anthracis* Steme, MCCC 1A02161 e MCCC 1A01412 não possuem o gene da toxina nem *capB* e *capC*, indicando ausência de ambos os plasmídeos. Junto com as cepas de *B. anthracis* encontramos três cepas de *B. thuringiensis*. Duas cepas possuem genes pouco conservados para as toxinas cry, e a terceira não possui nenhuma sequência minimamente similar. Hipotizamos que estas cepas são *B. thuringiensis*-like ou outra espécie. No outro sub-clado encontramos a toxina antrax conservada em uma cepa de *B. cereus*. As cepas 1930 e 2205, sequenciadas pelo nosso grupo, em função da posição na árvore e pelos genes conservados presentes ou não, possivelmente são outra espécie do grupo *B. cereus*.

O segundo clado pode ser sub-dividido em três sub-clados. O primeiro sub-clado, com linhas roxas, é formado apenas com cepas de *B. cereus*. Todas apresentam o gene *capA* conservado e nenhum resquício dos demais genes analisados. O segundo sub-clado (linhas vermelhas) foi formado por cepas de *B. thuringiensis*, sendo sete sequenciadas pelo nosso grupo. Todas as cepas da Embrapa, além da HD73 e Kurstaki HD1 que foram retiradas do NCBI, possuem toxinas cry que atuam como bio-inseticidas de lepidópteros. Podemos supor que as demais cepas também podem ter a mesma ação bio-inseticida contra esta ordem de insetos. Todas as cepas deste sub-clado, com exceção da 2728 e 2744, possuem genes cry conservados. As cepas 2728 e 2744, diferentemente das demais, apresentou genes cry pouco conservados, e atividade contra lepdópteros (2744) e coleópteros (2728 e 2744). Por isso elas foram classificadas como *B. thuringiensis*-like porque há questões que ainda precisam ser compreendidas.

Por fim, o terceiro sub-clado (linhas azuis) foi formado por cepas de *B. thuringiensis*, incluindo sete sequenciadas pela Embrapa e quatro cepas de *B. cereus*. Todas as *B. thuringiensis* apresentaram genes cry conservados, enquanto de três das quatro *B. cereus*

apresentaram genes cry pouco conservados. As cepas da Embrapa apresentaram ação bioinseticida contra coleópteros. É importante ressaltar que nenhuma cepa fora do sub-clado das *B. thuringiensis* e nenhuma *B. cereus* dentro deste sub-clado apresentou genes cry conservados.

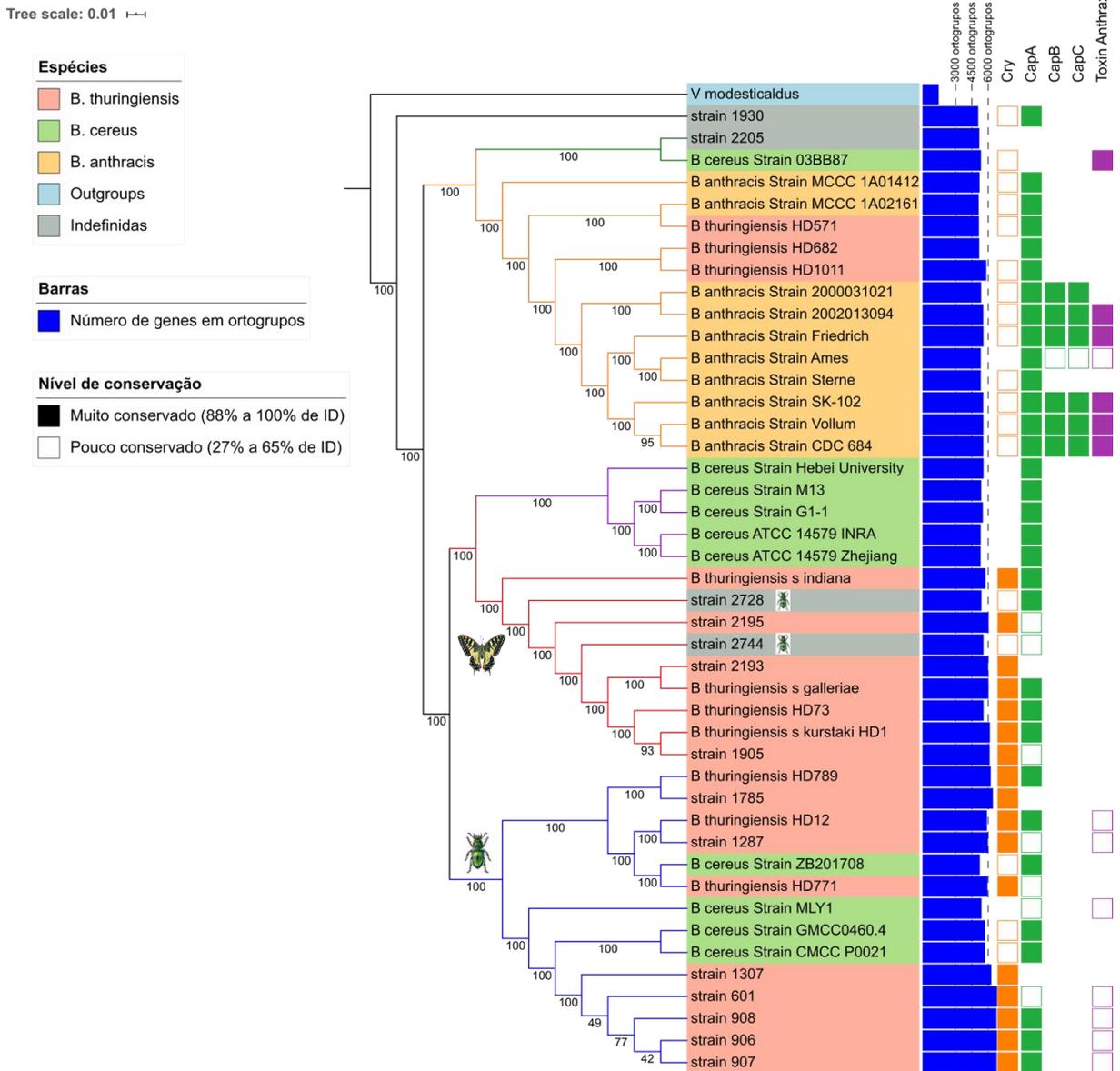


Figura 1: Árvore filogenética de 44 cepas de *B. anthracis*, *B. cereus* e *B. thuringiensis* com valores de bootstrap. As espécies estão preenchidas com cores distintas. As barras azuis representam o número de ortogrupos encontrados para cada cepa. Os quadrados representam a presença conservada (cores preenchidas), presença pouco conservada (quadrados vazios) ou ausência absoluta de genes codificadores para toxinas.

3.3. Identificação de ilhas genômicas contendo genes associados à patogenicidade

Para verificar a hipótese de que os genes codificadores para toxinas se localizam em ilhas genômicas, e, por conseguinte, são produtos de transferência horizontal de genes, identificamos as ilhas de patogenicidade em *B. anthracis*, *B. cereus* e *B. thuringiensis* e verificamos os genes de interesse. A análise das ilhas de patogenicidade trouxe informações interessantes. Observamos que há diferenças tanto na quantidade de ilhas patogênicas, quanto na quantidade de genes de interesse entre as próprias espécies (exemplo, *B. anthracis* Strain SK-102 e *B. anthracis* Strain Vollum) e entre espécies diferentes (*B. anthracis* Strain Vollum e *B. thuringiensis* HD57).

Os dados mostraram que *B. anthracis* é a espécie que apresenta, em média, o menor número de ilhas de patogenicidade (19 por genoma) com média aproximada de dois genes de cada categoria por ilha (Figura 2 e Tabela 6). Os genomas de *B. cereus* possuem em média 26 ilhas por genoma, sendo que cada ilha apresenta em média cinco genes de esporulação e apenas um das demais categorias. *B. thuringiensis* foi a espécie que apresentou o maior número médio de ilhas de patogenicidade por genoma (33), e como era esperado, o maior número médio de genes de toxina (quatro por ilha). Esta espécie também apresentou em média 4,5 genes de esporulação por ilha. Esses resultados mostraram que cada espécie possui um padrão muito diferente de ilhas de patogenicidade, fato que pode ser mais um fator a ser considerado para classificá-las.

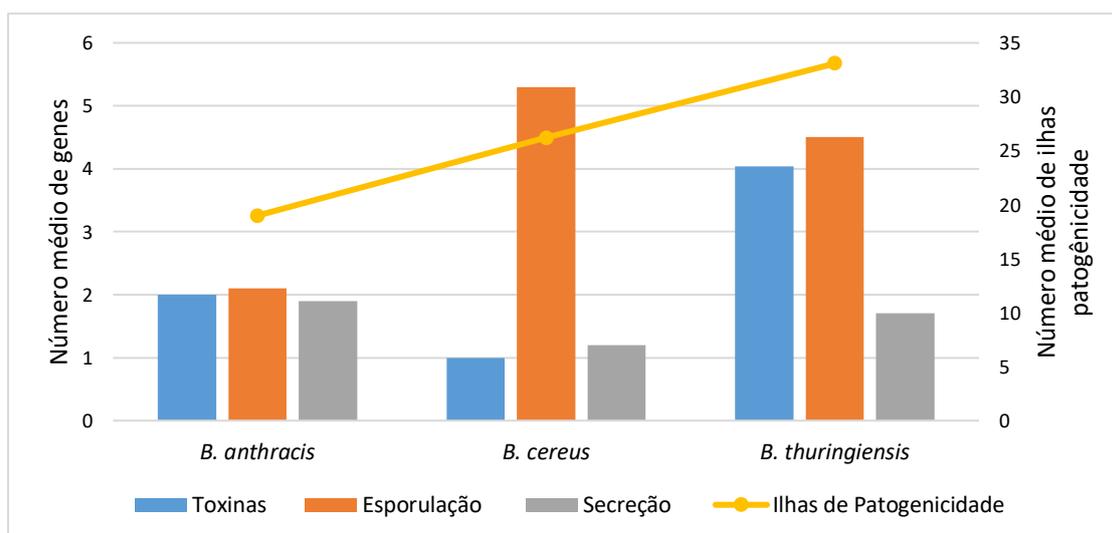


Figura 2: Número médio de genes de toxinas, esporulação e secreção por ilha de patogenicidade e número médio de ilhas de patogenicidade encontrados em cada espécie

Tabela 6: Quantidade de ilhas de patogenicidade e de genes de toxina, esporulação e secreção, assim como o número de genes totais encontradas nas cepas de *B. anthracis*, *B. cereus* e *B. thuringiensis*

Cepas	Ilhas de Patogenicidade	Toxina	Esporulação	Secreção	Total de genes
<i>B. anthracis</i> Strain MCCC 1A01412	30	0	2	1	298
<i>B. anthracis</i> Strain MCCC 1A02161	8	1	0	0	130
<i>B. anthracis</i> Strain 2000031021	18	1	3	4	263
<i>B. anthracis</i> Strain 2002013094	22	1	1	1	347
<i>B. anthracis</i> Strain Friedrich	22	1	1	3	264
<i>B. anthracis</i> Strain Ames	14	1	0	2	172
<i>B. anthracis</i> Strain Sternes	14	0	0	0	188
<i>B. anthracis</i> Strain SK-102	20	4	5	2	341
<i>B. anthracis</i> Strain Vollum	26	4	4	3	389
<i>B. anthracis</i> Strain CDC 684	16	7	5	3	347
<i>B. cereus</i> Strain 03BB87	24	2	7	2	388
<i>B. cereus</i> Strain Hebei University	20	2	5	1	303
<i>B. cereus</i> Strain M13	26	3	4	0	358
<i>B. cereus</i> Strain G1-1	40	2	5	1	405
<i>B. cereus</i> Strain ATCC 14579 INRA	20	0	3	0	207
<i>B. cereus</i> Strain ATCC 14579 Zhejiang	18	0	3	0	225
<i>B. cereus</i> Strain ZB201708	18	0	3	4	276
<i>B. cereus</i> Strain MLY1-1	38	1	11	1	467
<i>B. cereus</i> Strain GMCC0460.4	30	0	6	1	318
<i>B. cereus</i> Strain CMCC P0021	28	0	6	2	423
<i>B. thuringiensis</i> HD571	20	2	3	3	194
<i>B. thuringiensis</i> HD682	12	1	1	1	174
<i>B. thuringiensis</i> HD1011	4	0	0	0	467
<i>B. thuringiensis</i> s <i>indiana</i>	32	3	1	1	437
<i>B. thuringiensis</i> s <i>galleriae</i>	50	4	7	2	575
<i>B. thuringiensis</i> HD73	42	1	7	3	407
<i>B. thuringiensis</i> s <i>Kurstaki</i> HD1	50	6	3	2	619
<i>B. thuringiensis</i> HD789	20	1	4	1	680
<i>B. thuringiensis</i> HD12	44	5	8	4	652
<i>B. thuringiensis</i> HD771	36	2	8	0	406
Strains 1930	34	4	6	0	515
Strains 2205	34	2	11	3	399
Strains 2728	30	3	7	2	390
Strains 2195	0	0	0	0	0
Strains 2744	30	1	5	0	415
Strains 2193	58	10	3	1	775
Strains 1905	50	4	6	2	700
Strains 1785	28	0	4	2	390
Strains 1287	42	10	0	1	557
Strains 1307	50	11	6	2	932
Strains 601	32	4	5	3	682
Strains 908	32	8	5	3	765

Strains 906	32	9	4	2	715
Strains 907	32	6	4	3	778

Alguns resultados nos chamaram a atenção, como a *Strain* 2195 que não apresentou nenhuma ilha patogênica e *B. thuringiensis* HD1011, que possui quatro ilhas (enquanto a média foi de quase 35) e *B. anthracis* Strain Sternes que possui 14 ilhas (enquanto a média para esta espécie foi de aproximadamente 3), mas ambas não apresentaram nenhum gene de interesse.

4. DISCUSSÃO

Os resultados mostraram que a tentativa de classificar as três espécies baseado em genes candidatos a marcadores moleculares não foi efetivo porque eles mostraram-se muito conservados intra e entre espécies. Os nossos resultados corroboram com diversos trabalhos já publicados. Liu e colaboradores viram que os níveis de divergência do gene 16S rRNA se sobrepõe entre as comparações inter (97,34%-100%) e intra-espécies (99,14-100%), ao comparar 1007 sequências de 223 cepas do grupo *B. cereus*, indicando que a resolução deste gene foi muito baixa para ser usado como marcador molecular (LIU et al., 2015). De fato, desde a década de 90 que este gene já estava na berlinda como possível marcador (FOX; WISOTZKEY; JURTSCHUK, 1992). Resultados semelhantes foram encontrados em outros gêneros de bactérias, como *Aeromonas*, *Pseudomonas* e *Vibrio* (ARIAS, OLIVARES-FUSTER E GORIS, 2010; BODILIS ET AL., 2012; ROGER ET AL., 2012). Os genes *gyrB* e *rpoB* mostraram-se ser bons em separar corretamente *B. anthracis* das demais, mas falharam em separar *B. cereus* de *B. thuringiensis* (BAVYKIN et al., 2004; KI, ZHANG E QIAN, 2009).

Os resultados dos genes espécie-específicos *capA*, *capB*, *capC* e toxina antrax de *B. anthracis*, e os genes *cry* de *B. thuringiensis* nos fizeram pensar na possibilidade de aquisição via transferência horizontal gênica, com a aquisição do plasmídeo pXO1-like. É sabido que os três genes da biossíntese da cápsula se localizam no plasmídeo pXO2 e o gene da toxina, no pXO1, que algumas cepas de *B. anthracis* não apresentam ambos os plasmídeos, e isso afeta diretamente a virulência desta espécie. O uso destes genes como marcadores moleculares teriam uma alta especificidade, porém a sensibilidade seria limitada. Curiosamente, a cepa *B. cereus* 03BB87 possui a toxina antrax conservada, mas nenhum gene codificador da cápsula. Uma explicação seria que esta cepa adquiriu o plasmídeo apenas pXO1. Ainda questiona-se se ela é de fato *B. cereus*, e não outra espécie do grupo *B. cereus*. Erros de identificação são comuns de ocorrerem. Foi visto que 149 de 224 genomas depositados no GenBank podem estar com erros de identificação de acordo com uma análise de dDDH (digital DNA: DNA

hybridization), uma quantidade surpreendente e assustadora (LIU ET AL., 2015). Esse resultado é um forte indicativo de um ganho mais recente destes genes em termos de evolução.

O uso de metodologia de análise de grupos ortólogos em uma estratégia de genômica comparativa trouxe resultados muito interessantes, ainda que não tenha resolvido totalmente a classificação de *B. anthracis*, *B. cereus* e *B. thuringiensis*. A topologia da árvore filogenética encontrada por nós está de acordo com outros trabalhos que usaram desde genes individuais a abordagens de pan-genomas. Em todos os estudos, as três espécies se misturam entre si em maior ou menor grau (BAZINET, 2017; LIU ET AL., 2015), mostrando a enorme complexidade em diferenciá-las. De fato, há alguns grupos que defendem que as três espécies deveriam ser consideradas apenas uma em função da grande dificuldade de diferenciá-las pelos mais diversos métodos laboratoriais (HELGASON ET AL., 2000).

As análises de ilhas genômicas também revelaram informações importantes e úteis que podem colaborar com a classificação corretas das espécies de *Bacillus* de interesse deste trabalho. O alto número médio de ilhas de patogenicidade encontradas em *B. thuringiensis* pode estar relacionado com o fato de esta espécie ser a que apresenta a maior quantidade média de plasmídeos em seus genomas: 6,4 (11,2% de DNA plasmidial) contra 2 (4,2% de DNA plasmidial) nas demais espécies. O *B. thuringiensis* também apresenta a maior quantidade de transposases da família 3 (IS3), com média de 14,3 elementos por genoma contra 5,5 em *B. cereus*. Sabe-se que elementos transponíveis estão associados com fatores de virulência (MAHILLON ET AL., 1985), e isso pode explicar os achados em *B. thuringiensis*. De fato, um fator de correlação de 0,91 foi encontrado entre a porcentagem de elementos transponíveis em genomas e a quantidade média de plasmídeos encontrados, indicando que os elementos ISs e Tn3-like possui um papel importante na plasticidade plasmidial (FAYAD, KALLASSY AWAD E MAHILLON, 2019).

Isso tem uma importância grande no caso de *B. thuringiensis* onde seu estilo de vida particular depende da presença, diversidade, combinação e do embaralhamento de diferentes genes de entomotoxinas (FAYAD, KALLASSY AWAD E MAHILLON, 2019). Com isso, acreditamos que *B. thuringiensis* é mais permissiva a transferências horizontais gênicas do que as demais. O que fez com que esta espécie adquirisse mais ilhas patogênicas como forma de adaptação devido ao estresse que sofreu no ambiente em que vive.

5. CONCLUSÕES

Neste trabalho buscamos identificar genes marcadores conservados ou espécie específicos que pudessem ser usados para auxiliar na identificação das três espécies de *Bacillus*. Os genes housekeeping possuem baixa resolução, não sendo interessantes para serem usados como método de diferenciação entre as espécies. Os genes codificadores de toxinas foram mais promissores, mas vimos que entre todas as espécies há exceções. A análise de genômica comparativa através dos grupos ortólogos foi eficiente em separar *B. anthracis* das outras duas espécies. *B. cereus* e *B. thuringiensis* não formaram grupos independentes. No entanto, esta análise evidenciou que as cepas de *B. thuringiensis* tóxicas a coleópteros foram um sub-clado diferente e independente das cepas tóxicas a lepdópteros, com exceção de duas estirpes anômalas (ausência de genes cry). As análises das ilhas patogênicas foram bem elucidativas com relação à diferença entre as espécies com relação ao número e configuração das ilhas de patogenicidade, mas sempre há casos que não fogem à regra. Mas os nossos resultados foram bastante clarificantes com relação ao papel dos eventos de transferência horizontal gênica na diversidade de genes de toxinas.

6. REFERENCIAS

AGATA, N.; OHTA, M.; MORI, M. Production of an Emetic Toxin, Cereulide, is Associated with a Specific Class of *Bacillus cereus*. **Current Microbiology**, v.33, n.1, t, 67-69, 1996. Disponível em: <http://search.ebscohost.com/login.aspx?direct=true&db=edsbl&AN=RN009626669&lang=pt-br&site=eds-live>. Acesso em: 10 nov. 2020.

ALTAYAR, M.; SUTHERLAND, A. D. *Bacillus cereus* is common in the environment but emetic toxin producing isolates are rare. **JOURNAL OF APPLIED MICROBIOLOGY**, n.1, p.7, 2006. Disponível em: <http://search.ebscohost.com/login.aspx?direct=true&db=edsbl&AN=RN180389050&lang=pt-br&site=eds-live>. Acesso em: 10 nov. 2020.

ANTOLINOS, VERA et al. Development of a high-resolution melting-based approach for efficient differentiation among *Bacillus cereus* group isolates. **FOODBORNE PATHOGENS AND DISEASE**, v.9, n.9, p.777-785, 2012. Disponível em: <https://www.liebertpub.com/doi/abs/10.1089/fpd.2012.1148>. Acesso em: 10 nov. 2020

ARIAS, C. R.; OLIVARES-FUSTER, O.; GORIS, J. High intragenomic heterogeneity of 16S rRNA genes in a subset of *Vibrio vulnificus* strains from the western Mediterranean coast. **International Microbiology**, v. 13, n. 4, p. 179–188. DOI: 10.2436/20.1501.01.124.

ASH, C.; COLLINS, M. D. Comparative analysis of 23S ribosomal RNA gene sequences of *Bacillus anthracis* and emetic *Bacillus cereus* determined by PCR-direct sequencing. **FEMS Microbiology Letters**, v.94, n.1–2, p.75–80, Disponível em: <http://search.ebscohost.com/login.aspx?direct=true&db=edselc&AN=edselc.2-52.0-0026891169&lang=pt-br&site=eds-live>. Acesso em: 10 nov. 2020.

BAVYKIN, S. G. *et al.* Use of 16S rRNA, 23S rRNA, and gyrB Gene Sequence Analysis To Determine Phylogenetic Relationships of Bacillus cereus Group Microorganisms. **JOURNAL OF CLINICAL MICROBIOLOGY**, n.8, p.3711, 2004. Disponível em: <http://search.ebscohost.com/login.aspx?direct=true&db=edsbl&AN=RN154603413&lang=pt-br&site=eds-live>. Acesso em: 10 nov. 2020.

BAZINET, A. L. Pan-genome and phylogeny of Bacillus cereus sensu lato. **BMC Evolutionary Biology**, n.1, 2017. Disponível em: <http://search.ebscohost.com/login.aspx?direct=true&db=edsgao&AN=edsgcl.509101465&lang=pt-br&site=eds-live>. Acesso em: 10 nov. 2020.

BELLANGER, X. *et al.* Conjugative and mobilizable genomic islands in bacteria: evolution and diversity. **FEMS microbiology reviews**, v.38, n 4, p.720–760, 2014. Disponível em: <http://search.ebscohost.com/login.aspx?direct=true&db=mdc&AN=24372381&lang=pt-br&site=eds-live>. Acesso em: 10 nov. 2020.

BERTELLI, C. *et al.* IslandViewer 4: expanded prediction of genomic islands for larger-scale datasets. **Nucleic acids research**, v.45, n.W1, p. W30–W35, 2017. Disponível em: <http://search.ebscohost.com/login.aspx?direct=true&db=mdc&AN=28472413&lang=pt-br&site=eds-live>. Acesso em: 10 nov. 2020.

BETZ, F. S.; HAMMOND, B. G.; FUCHS, R. L. Safety and Advantages of Bacillus thuringiensis-Protected Plants to Control Insect Pests. **Regulatory Toxicology and Pharmacology**, v.32, n.2, p.156-173, 2000. Disponível em: <http://search.ebscohost.com/login.aspx?direct=true&db=edsbas&AN=edsbas.A8D509F3&lang=pt-br&site=eds-live>. Acesso em: 10 nov. 2020.

BHATNAGAR, R.; BATRA, S. Anthrax toxin. **Anthrax Toxin Critical Reviews in Microbiology**, v.27, n.3 p.167-200, 2001. Disponível em: <http://search.ebscohost.com/login.aspx?direct=true&db=edsbas&AN=edsbas.798666D3&lang=pt-br&site=eds-live>. Acesso em: 10 nov. 2020.

BODILIS, J. *et al.* Variable Copy Number, Intra-Genomic Heterogeneities and Lateral Transfers of the 16S rRNA Gene in Pseudomonas. **PLoS ONE**,v.7, n. 4, p. e35647, 2012. Disponível em: <http://search.ebscohost.com/login.aspx?direct=true&db=edsgao&AN=edsgcl.477046342>. Acesso em: 10 nov. 2020.

BRAVO, A.; GILL, S. S.; SOBERÓN, M. Mode of action of Bacillus thuringiensis Cry and Cyt toxins and their potential for insect control. **Toxicon**, v.49 n.4 p.423-435, 2007. Disponível em: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC1857359/>. Acesso em: 10 nov. 2020.

BROSSIER, F.; MOCK, M. Toxins of Bacillus anthracis. **Toxicon**, v. 39, n. 11, p. 1747–1755, 2001. Disponível em: https://www.sciencedirect.com/science/article/pii/S00410101001611?casa_token=XpfDsHCA9IYAAAAA:6vWIA0qysHrckQVT10QMmDhVWL4Gs8wK-QKG-zQeiVTd1eN9VZshZJlt0tKkDoSKJhRFA8pUPnU. Acesso em: 10 nov. 2020.

CALDART, E. T. *et al.* Phylogenetic Analysis: Basic Concepts and Its Use as a Tool for Virology and Molecular Epidemiology. **Acta Scientiae Veterinariae**, v. 44, n. 1, p. 20,

2016. Disponível em:
<http://search.ebscohost.com/login.aspx?direct=true&db=edsbas&AN=edsbas.E3181A40&lang=pt-br&site=eds-live>. Acesso em: 11 nov. 2020.

CEUPPENS, S.; BOON, N.; UYTENDAELE, M. Diversity of *Bacillus cereus* group strains is reflected in their broad range of pathogenicity and diverse ecological lifestyles. **FEMS microbiology ecology**, v. 84, n. 3, p. 433–450, 2013. Disponível em:
<http://search.ebscohost.com/login.aspx?direct=true&db=mdc&AN=23488744&lang=pt-br&site=eds-live>. Acesso em: 11 nov. 2020.

DESCORPS-DECLÈRE, S. *et al.* The multiple facets of homology and their use in comparative genomics to study the evolution of genes, genomes, and species. **Biochimie**, v.90, n.4, p.595–608, 2008. Disponível em:
<http://search.ebscohost.com/login.aspx?direct=true&db=edsbas&AN=edsbas.6E9FA390&lang=pt-br&site=eds-live>. Acesso em: 11 nov. 2020.

DOBRINDT, U. *et al.* Genomic islands in pathogenic and environmental microorganisms. **Nature Reviews Microbiology**, v.2, n.5, p.414, 2004. Disponível em:
<http://search.ebscohost.com/login.aspx?direct=true&db=edsgao&AN=edsgcl.188965963&lang=pt-br&site=eds-live>. Acesso em: 11 nov. 2020.

DONOVAN, W. P. *et al.* Discovery and characterization of Sip1A: A novel secreted protein from *Bacillus thuringiensis* with activity against coleopteran larvae. **Applied Microbiology and Biotechnology**, v.72, n.4, p.713–719, 2006. Disponível em:
<http://search.ebscohost.com/login.aspx?direct=true&db=edselc&AN=edselc.2-52.0-33749002224&lang=pt-br&site=eds-live>. Acesso em: 11 nov. 2020.

EHLING-SCHULZ, M.; LERECLUS, D.; KOEHLER, T. M. The *Bacillus cereus* Group: *Bacillus* Species with Pathogenic Potential. **Microbiology Spectrum**, v. 7, n. 3, p. 1–35, 2019a. Disponível em:
<http://search.ebscohost.com/login.aspx?direct=true&db=edsbas&AN=edsbas.B9EAC5D9&lang=pt-br&site=eds-live>. Acesso em: 11 nov. 2020.

EMMS, D. M.; KELLY, S. OrthoFinder: Phylogenetic orthology inference for comparative genomics. **Genome Biology (Online Edition)**, v. 20, n. 1, p. 1-14, 2019. Disponível em:
<http://search.ebscohost.com/login.aspx?direct=true&db=edsgao&AN=edsgcl.607365821&lang=pt-br&site=eds-live>. Acesso em: 11 nov. 2020.

FAYAD, N.; KALLASSY AWAD, M.; MAHILLON, J. Diversity of *Bacillus cereus* sensu lato mobilome. **BMC Genomics**, v. 20, n. 1, p. 1–11, 2019. Disponível em:
<http://search.ebscohost.com/login.aspx?direct=true&db=edsgao&AN=edsgcl.590705305&lang=pt-br&site=eds-live>. Acesso em: 11 nov. 2020.

FOX, G. E.; WISOTZKEY, J. D.; JURTSCHUK, P. How close is close: 16S rRNA sequence identity may not be sufficient to guarantee species identity. **International Journal of Systematic Bacteriology**, v.42, n.1, p.166-170, 1992. Disponível em:
<http://search.ebscohost.com/login.aspx?direct=true&db=edsnas&AN=edsnas.20050000803&lang=pt-br&site=eds-live>. Acesso em: 11 nov. 2020.

GLASSET, B. *et al.* *Bacillus cereus*, a serious cause of nosocomial infections: Epidemiologic and genetic survey. **PLoS ONE**, v.13, n.5, p.e0194346., 2018. Disponível em:

<http://search.ebscohost.com/login.aspx?direct=true&db=edsgea&AN=edsgcl.540019393&lang=pt-br&site=eds-live>. Acesso em: 11 nov. 2020.

GREEN, B. D. *et al.* Demonstration of a capsule plasmid in *Bacillus anthracis*. **Infection and Immunity**, v.49, n.2, p.291-297, 1985. Disponível em: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC262013/pdf/iai00113-0043>. Acesso em: 11 nov. 2020.

GRIFFITHS, M. W.; SCHRAFT, H. *Bacillus cereus* Food Poisoning. *In: Foodborne Diseases: Third Edition*, Academic Press, 2017. p. 395-405. Disponível em: <https://www.sciencedirect.com/science/article/pii/B9780123850072000206> Acesso em: 11 nov. 2020.

HALLSTROM, K. N.; MCCORMICK, B. A. Pathogenicity Islands: Origins, Structure, and Roles in Bacterial Pathogenesis. *In: Molecular Medical Microbiology: Second Edition*. Academic Press, p.303-314, 2015. Disponível em: <https://www.sciencedirect.com/book/9780123971692/molecular-medical-microbiology> Acesso em: 11 nov. 2020.

HELGASON, E. *et al.* *Bacillus anthracis*, *Bacillus cereus*, and *Bacillus thuringiensis* - One species on the basis of genetic evidence. **Applied and Environmental Microbiology**, v. 66, n.6, p.2627-2630, 2000. Disponível em: <http://search.ebscohost.com/login.aspx?direct=true&db=edsbas&AN=edsbas.FCFB4D3A&lang=pt-br&site=eds-live>. Acesso em: 16 nov. 2020.

JUHAS, M. *et al.* **Genomic islands: Tools of bacterial horizontal gene transfer and evolution** **FEMS Microbiology Reviews**, v.33, n.2, p.376-393, 2009. Disponível em: <http://search.ebscohost.com/login.aspx?direct=true&db=edsbas&AN=edsbas.EE59F88F&lang=pt-br&site=eds-live>. Acesso em: 17 nov. 2020.

KI, J. S.; ZHANG, W.; QIAN, P. Y. Discovery of marine *Bacillus* species by 16S rRNA and rpoB comparisons and their usefulness for species identification. **Journal of Microbiological Methods**, v.77, n.1, p.48-57, 2009. Disponível em: <http://search.ebscohost.com/login.aspx?direct=true&db=edsbas&AN=edsbas.7B9B0101&lang=pt-br&site=eds-live>. Acesso em: 17 nov. 2020.

LETUNIC, I.; BORK, P. Interactive Tree of Life (iTOL) v4: Recent updates and new developments. **Nucleic Acids Research**, v. 47, n. W1, p. W256-W259, 2019. Disponível em: <https://academic.oup.com/nar/article/47/W1/W256/5424068>. Acesso em: 17 nov. 2020.

LIU, Y. *et al.* Genomic insights into the taxonomic status of the *Bacillus cereus* group. **Scientific Reports**, v.5, n.1 p.1-11, set. 2015. Disponível em: <http://search.ebscohost.com/login.aspx?direct=true&db=edsbas&AN=edsbas.D87AD8FE&lang=pt-br&site=eds-live>. Acesso em: 17 nov. 2020.

MAHILLON, J. *et al.* Nucleotide sequence and structural organization of an insertion sequence element (IS231) from *Bacillus thuringiensis* strain berliner 1715. **The EMBO journal**, v.4, no.13B, p.3895-3899, 1985. Disponível em: <http://search.ebscohost.com/login.aspx?direct=true&db=edsbas&AN=edsbas.A5AF691E&lang=pt-br&site=eds-live>. Acesso em: 26 nov. 2020.

MAKINO, S. I. *et al.* Molecular characterization and protein analysis of the cap region, which is essential for encapsulation in *Bacillus anthracis*. **Journal of Bacteriology**, v.171, n.2, p. 722–730, 1989. Disponível em: <http://search.ebscohost.com/login.aspx?direct=true&db=edsbas&AN=edsbas.A903D58&lang=pt-br&site=eds-live>. Acesso em: 26 nov. 2020.

MANZANO, M. *et al.* Molecular methods to evaluate biodiversity in *Bacillus cereus* and *Bacillus thuringiensis* strains from different origins. **Food Microbiology**, v.26, n.3, p. 259–264, 2009. Disponível em: <http://search.ebscohost.com/login.aspx?direct=true&db=edsbas&AN=edsbas.63E6E026&lang=pt-br&site=eds-live>. Acesso em: 26 nov. 2020.

MAIKE TAÍS MAZIERO; LUCIANO DOS SANTOS BERSOT. *Bacillus Cereus* Em Produtos Lácteos - Uma Revisão. **Revista do Instituto de Laticínios Cândido Tostes**, v.66, n.381, p.5–10, 2011. Disponível em: <http://search.ebscohost.com/login.aspx?direct=true&db=edsair&AN=edsair.doajarticles..c269521b50c55c0aacf453fbe489e6ac&lang=pt-br&site=eds-live>. Acesso em: 30 nov. 2020.

MENDES, R. A. *et al.* Contaminação ambiental por *Bacillus cereus* em unidade de alimentação e nutrição. **Revista de Nutrição**, v. 17, n. 2, p. 255-261, 2004. Disponível em: https://www.scielo.br/scielo.php?script=sci_arttext&pid=S1415-52732004000200012. Acesso em: 26 nov. 2020.

MIKESELL, P. *et al.* Evidence for plasmid-mediated toxin production in *Bacillus anthracis*. **Infection and Immunity**, v.39, p. 371-376, 1983. Disponível em: <http://search.ebscohost.com/login.aspx?direct=true&db=edsbas&AN=edsbas.566B54E3&lang=pt-br&site=eds-live>. Acesso em: 26 nov. 2020.

NILEGAONKAR, S. S. *et al.* Production and partial characterization of dehairing protease from *Bacillus cereus* MCM B-326. **Bioresource Technology**, v. 98, n. 6, p. 1238–1245, abr. 2007. Disponível em: <http://search.ebscohost.com/login.aspx?direct=true&db=edsbas&AN=edsbas.E88B9690&lang=pt-br&site=eds-live>. Acesso em: 26 nov. 2020.

NOGUERA, P. A.; IBARRA, J. E. Detection of new cry genes of *Bacillus thuringiensis* by Use of a Novel PCR primer system. **Applied and Environmental Microbiology**, v.79, n.18, p.6150-6155, 2010. Disponível em: <http://search.ebscohost.com/login.aspx?direct=true&db=edsbas&AN=edsbas.5B319F94&lang=pt-br&site=eds-live>. Acesso em: 26 nov. 2020..

PALMA, L. *et al.* **Bacillus thuringiensis** toxins: An overview of their biocidal activity. **ToxinsMDPI AG**, v.6, n.12, p.3296–3325, 2014. Disponível em: <http://search.ebscohost.com/login.aspx?direct=true&db=edsdoj&AN=edsdoj.fa6a3ac225b9470c8182fa0a43457f40&lang=pt-br&site=eds-live>. Acesso em: 26

PARDO-LÓPEZ, L.; SOBERÓN, M.; BRAVO, A. *Bacillus thuringiensis* insecticidal three-domain Cry toxins: Mode of action, insect resistance and consequences for crop protection. **FEMS Microbiology Reviews**, v.37, n.1, p.3-22, 2013. Disponível em:

<http://search.ebscohost.com/login.aspx?direct=true&db=edsbas&AN=edsbas.D2FBF859&lang=pt-br&site=eds-live>. Acesso em: 26 nov. 2020.

RASKO, D. A. *et al.* Genomics of the *Bacillus cereus* group of organisms. **FEMS Microbiology Reviews**, v.29, n.2, p.303-329, 2005. Disponível em: <https://academic.oup.com/femsre/article/29/2/303/572138>. Acesso em: 26 nov. 2005.

REECK, G. R. *et al.* "Homology" in proteins and nucleic acids: A terminology muddle and a way out of it. **Cell**, v.50, n.5, p.667, 1987. Disponível em: [https://www.cell.com/cell/pdf/0092-8674\(87\)90322-9.pdf](https://www.cell.com/cell/pdf/0092-8674(87)90322-9.pdf). Acesso em: 26 nov. 2020.
1987.

ROGER, F. *et al.* Ribosomal Multi-Operon Diversity: An Original Perspective on the Genus *Aeromonas*. **PLoS ONE**, v.7, n.9, p.e46268, 2012. Disponível em: <http://search.ebscohost.com/login.aspx?direct=true&db=edsair&AN=edsair.od.....166..593064b21a6a86dee6f11e9c0f03df48&lang=pt-br&site=eds-live>. Acesso em: 27 nov. 2020

ROZEWICKI, J. *et al.* MAFFT-DASH: Integrated protein sequence and structural alignment. **Nucleic Acids Research**, v.47, n.W1, p.W5-W10, 2019. Disponível em: <https://academic.oup.com/nar/article/47/W1/W5/5486273>. Acesso em: 27 nov. 2020

SÁ, D. M. Modelos Caracterizando a Interação Entre As Toxinas Da Família Cry1a De *Bacillus Thuringiensis* E O Receptor Bt-R1 De *Manduca Sexta*. 2015. Disponível em: <https://repositorio.unb.br/handle/10482/19647>. Acesso em 27 nov. 2020

SCHMIDT, H.; HENSEL, M. Pathogenicity Islands in Bacterial Pathogenesis. **Clinical Microbiology Reviews**, v.19, n.1, p.257-257, 2006. Disponível em: <https://cmr.asm.org/content/cmr/19/1/257.full.pdf>. Acesso em 27 nov. 2020

SHEIKHIZADEH ANARI, S. *et al.* Efficient inference of homologs in large eukaryotic pan-proteomes. **BMC bioinformatics**, v.19, n.1, p.340, 2018. Disponível em: <https://link.springer.com/article/10.1186/s12859-018-2362-4>. Acesso em 27 nov. 2020

TRIFINOPOULOS, J. *et al.* W-IQ-TREE: a fast online phylogenetic tool for maximum likelihood analysis. **Nucleic acids research**, v. 44, n. W1, p. W232-W235, 2016. Disponível em: <https://academic.oup.com/nar/article/44/W1/W232/2499307>. Acesso em 27 nov. 2020.

TURNBULL, P. C. B. Introduction: anthrax history, disease and ecology. In: **Anthrax**. Springer, Berlin, Heidelberg, 2002. p.1-19. Disponível em: https://link.springer.com/chapter/10.1007/978-3-662-05767-4_1. Acesso em 27 nov. 2020

VOLOKHOV, D. V. *et al.* RNA polymerase beta subunit (*rpoB*) gene and the 16S-23S rRNA intergenic transcribed spacer region (ITS) as complementary molecular markers in addition to the 16S rRNA gene for phylogenetic analysis and identification of the species of the family *Mycoplasmataceae*. **Molecular phylogenetics and evolution**, v. 62, n. 1, p. 515-528, 2012. Disponível em: https://www.sciencedirect.com/science/article/pii/S1055790311004611?casa_token=JXW6ab

KxgDgAAAAA:k_lbaWEtvkRG-oQtOhNtP1d--
3bhSbZPmTgdb7Lyz5iOurTUryZ5JTSx2jdwttqzZ7vC7fDn8s4. Acesso em 27 nov. 2020

YAMADA, S. *et al.* Cloning and nucleotide sequence analysis of *gyrB* of *Bacillus cereus*, *B. thuringiensis*, *B. mycoides*, and *B. anthracis* and their application to the detection of *B. cereus* in rice. **Applied and Environmental Microbiology**, v. 65, n. 4, p. 1483-1490, 1999. Disponível em: <https://aem.asm.org/content/65/4/1483.short>. Acesso em: Acesso em 27 nov. 2020

ZHANG, X. *et al.* A mechanism of cell death involving an adenylyl cyclase/PKA signaling pathway is induced by the Cry1Ab toxin of *Bacillus thuringiensis*. **Proceedings of the National Academy of Sciences of the United States of America**, v. 103, n. 26, p. 9897-9902, 2006. Disponível em: <https://www.pnas.org/content/103/26/9897/>. Acesso em 27 nov. 2020