



UNIVERSIDADE
ESTADUAL DE LONDRINA

PRISCILLA DE FREITAS CARDOSO

**ANÁLISE *IN SILICO* DE GENES *rap-phr* EM *Bacillus cereus*
*lato sensu***

Londrina
2015

PRISCILLA DE FREITAS CARDOSO

**ANÁLISE *IN SILICO* DE GENES *rap-phr* EM *Bacillus cereus*
*lato sensu***

Defesa apresentada ao Programa de Pós-Graduação em Genética e Biologia Molecular, da Universidade Estadual de Londrina, como requisito para a obtenção do título de Mestre.

Orientadora: Prof^a. Dr^a. Gislayne Trindade Vilas-Bôas.

Londrina
2015

**Catálogo elaborado pela Divisão de Processos Técnicos da Biblioteca Central da
Universidade Estadual de Londrina.**

Dados Internacionais de Catalogação-na-Publicação (CIP)

C268a Cardoso, Priscilla de Freitas.

Análise *in silico* de genes *rap-phr* em *Bacillus cereus lato sensu* / Priscilla de Freitas Cardoso. – Londrina, 2015.
103 f. : il.

Orientador: Gislayne Fernandes Lemes Trindade Vilas-Bôas.

Dissertação (Mestrado em Genética e Biologia Molecular) – Universidade Estadual de Londrina, Centro de Ciências Biológicas, Programa de Pós-Graduação em Genética e Biologia Molecular, 2015.

Inclui bibliografia.

1. Bacillus (Bactéria) – Genética – Teses. 2. Proteínas de bactérias – Teses.
3. Esporos bacterianos – Teses. 4. Plasmídeos – Teses. I. Vilas-Bôas, Gislayne Fernandes Lemes Trindade. II. Universidade Estadual de Londrina. Centro de Ciências Biológicas. Programa de Pós-Graduação em Genética e Biologia Molecular. III. Título.

CDU 579.852.1

PRISCILLA DE FREITAS CARDOSO

ANÁLISE *IN SILICO* DE GENES *rap-phr* EM *Bacillus cereus lato sensu*

Defesa apresentada ao Programa de Pós-Graduação em Genética e Biologia Molecular, da Universidade Estadual de Londrina, como requisito para a obtenção do título de Mestre.

BANCA EXAMINADORA

Orientadora: Prof^a. Dr^a. Gislayne Trindade
Vilas-Bôas
Universidade Estadual de Londrina - UEL

Prof.^a Dr.^a Olívia Marcia Nagy Arantes
Universidade Estadual de Londrina - UEL

Prof. Dr. Carlos Roberto Maximiano da Silva
Universidade Estadual de Londrina - UEL

Londrina, 12 de fevereiro de 2015.

AGRADECIMENTOS

À Deus, por minha vida, por guiar e iluminar meus caminhos e por tudo que sou hoje;

À minha família que sempre foi meu porto seguro: meus pais Lourdes e Romulo e minha irmã Larissa, pelo exemplo, incentivo, confiança, carinho e apoio irrestrito;

À minha orientadora, Prof^a. Dr^a. Gislayne Trindade Vilas-Bôas, pelas orientações, correções, ensinamentos, suporte, disposição, paciência e confiança;

Ao Prof. Dr. Laurival Antônio Vilas Bôas, pelas ideias, ajudas, suporte na parte da bioinfo e os momentos descontração;

À Prof^a. Dr^a. Olívia Marcia Nagy Arantes, pela disponibilidade em avaliar este trabalho e por suas contribuições;

Ao Prof. Dr. Carlos Roberto Maximiano da Silva, por suas correções na qualificação e, novamente, pela disponibilidade e atenção a este trabalho;

À Prof^a. Dr^a. Daniele Sartori pelo apoio em geral e contribuições na qualificação;

Ao pessoal do Laboratório de Genética e Taxonomia de Bactérias: Fernanda, Ana Paula, Amanda, Fernanda Henrique, Carla, Josiane, Kelly, Janaína, André (Deco), André (Nadal) e Ivan pelo companheirismo, conversas, incentivo, ideias e opiniões. Em especial, agradeço à Fernanda Fazon, por sua ajuda e contribuição na realização do trabalho e na discussão dos 'rap' e ao Ivan, pelo auxílio e dúvidas sanadas na área da bioinformática;

À Pós-Graduação em Genética e Biologia Molecular, especialmente à coordenadora, Prof^a. Dr^a Ana Lúcia Dias e à secretária Sueli, por todo suporte, disposição e lembretes;

Aos professores da Pós-Graduação em Genética e Biologia Molecular por compartilharem seus conhecimentos, pela formação acadêmica e contribuições diretas ou indiretas a este trabalho;

À Universidade Estadual de Londrina por tantos conhecimentos e pelos tantos anos em que tem sido minha segunda casa;

À CAPES, pelo suporte financeiro à pesquisa.

A todos aqueles que apesar de não terem sido especialmente citados, de alguma forma contribuíram direta ou indiretamente para a realização deste trabalho.

Muito Obrigada!!!

CARDOSO, Priscilla de Freitas. **Análise *in silico* de genes *rap-phr* em *Bacillus cereus lato sensu***. 2015. 103f. Dissertação (Programa de Pós-Graduação em Genética e Biologia Molecular) – Universidade Estadual de Londrina, Londrina, 2015.

RESUMO

Espécies do gênero *Bacillus* produzem proteínas Rap (regulador de resposta aspartato fosfatase) que regulam vias essenciais para a bactéria, como esporulação, competência e formação do biofilme. Rap é inibida pelo peptídeo Phr, que é secretado, processado e internalizado para cumprir sua função, atuando assim como sensor de *quorum*. Essas proteínas têm sido investigadas em *B. subtilis* e com poucos estudos em *Bacillus cereus lato sensu*. Um grupo bacteriano ao qual pertencem *B. cereus stricto sensu*, *B. thuringiensis* e *B. anthracis*, bactérias fenotipicamente diferentes, mas geneticamente muito similares. Atualmente não há consenso se estas bactérias devem ser consideradas espécies distintas ou variantes especializadas de uma única espécie, inclusive porque suas principais características fenotípicas são codificadas em plasmídeos. O objetivo deste trabalho foi identificar e caracterizar os sistemas *rap-phr* nos genomas destas espécies. Foram identificados 189 genes *rap* e 147 genes *phr* a partir do genoma de 34 linhagens do grupo com sequências genômicas completas. Dos genes obtidos, 23% são plasmidiais, uma frequência alta considerando um gene essencial para o metabolismo da célula. As proteínas Rap e Phr foram comparadas com proteínas ortólogas de *B. subtilis* e entre si. Foi possível notar uma grande variabilidade das sequências e em sua distribuição e na sintonia entre os genes de *B. anthracis*, enquanto que os genes de *B. cereus* e *B. thuringiensis* estão mais distribuídos e intercalados. Não foi possível identificar marcadores quanto a espécie nem em relação a sua localização no genoma. Entretanto, a grande quantidade dos genes plasmidiais nas linhagens de *B. thuringiensis*, em conjunto com a maior plasticidade dos genomas dessa espécie, pode permitir responder melhor às modificações ambientais.

Palavras-chave: Esporulação. Plasmídeo. *B. cereus stricto sensu*. *B. thuringiensis*. *B. anthracis*.

CARDOSO, Priscilla de Freitas. *In silico analysis of rap-phr genes in bacillus cereus lato sensu*. 2015. 103p. Dissertação (Programa de Pós-Graduação em Genética e Biologia Molecular) – Universidade Estadual de Londrina, Londrina, 2015.

ABSTRACT

Species of *Bacillus* genus produces Rap proteins (response regulator aspartate phosphatase) that regulates bacterial essential pathways, such as sporulation, competence for DNA transformation and biofilm formation. Rap is inhibited by Phr peptide that is secreted, processed and internalized to fulfill its function, thus acting as a *quorum* sensor. These proteins have been investigated in *B. subtilis* and there are few studies on *B. cereus lato sensu*. A bacterial group which belongs *B. thuringiensis*, *B. cereus stricto sensu* and *B. anthracis*, bacteria phenotypically different although highly similar genetically. Currently, there is no consensus on whether these bacteria should be considered distinct species or specialized variants of a single species, especially because its phenotypic characteristics are encoded by plasmid genes. The objective of this study was to identify and characterize in various aspects the *rap-phr* systems in these three species genomes. 189 *rap* and 147 *phr* genes were identified from the genome of 34 strains with complete genomic sequence. Of these, 23% are located in plasmids, a high frequency considering an essential gene for cell metabolism. Rap and Phr proteins are compared with *B. subtilis* orthologous proteins and among themselves. It was possible to notice a high sequence variability and distribution, the genes synteny in *B. anthracis*, while in *B. cereus* and *B. thuringiensis* they are more interspersed and distributed. It was not possible to identify species-specific markers, neither markers related with genome localization (chromosomal or plasmid). However, the high amount of plasmid genes in *B. thuringiensis* strains, together with the higher plasticity of the genome of this species, can enable them to provide a better response to environmental changes.

Keywords: Sporulation. Plasmid. *B. cereus sensu stricto sensu*. *B. thuringiensis*. *B. anthracis*.

LISTA DE ILUSTRAÇÕES

Figura 1 – Representação esquemática das vias de sinalização reguladas por módulos Rap-Phr	14
Figura 2 – Representação esquemática dos operons <i>rap-phr</i>	15
Figura 3 – Representação esquemática da organização estrutural das proteínas da família RNPP.....	16
Figura 4 – Estrutura espacial da proteína RapJ ligada ao peptídeo PhrC de <i>B. subtilis</i>	18

Artigo

Figura 1 – Árvore filogenética de sequências dos cinco esquemas de MLST das espécies de <i>B. cereus lato sensu</i>	43
Figura 2 – Número de genes <i>rap</i> por linhagem.....	45
Figura 3 – Representação gráfica da localização dos genes <i>rap</i> no cromossomo das linhagens de <i>B. anthracis</i> (Ba).....	47
Figura 4 – Representação gráfica dos alinhamentos entre blocos gênicos de linhagens de <i>B. anthracis</i> (Ba) gerado pelo Genomes Match	50
Figura 5 – Representação gráfica da localização dos genes <i>rap</i> no cromossomo das linhagens de <i>B. cereus</i> (Bc).....	49
Figura 6 – Representação gráfica da localização dos genes <i>rap</i> no cromossomo das linhagens de <i>B. thuringiensis</i> (Bt).	51
Figura 7 – Representação gráfica da localização dos genes <i>rap</i> nos plasmídeos das linhagens de <i>B. cereus</i> estudadas	52
Figura 8 – Representação gráfica da localização dos genes <i>rap</i> nos plasmídeos das linhagens de <i>B. thuringiensis</i> ; os genes <i>cry</i> , quando presentes também estão identificados.....	54
Figura 9 – Representação gráfica do alinhamento entre as sequências proteicas de genes anotados como <i>rapA</i>	65
Figura 10 – Representação gráfica do alinhamento entre as sequências proteicas de genes anotados como <i>rapD</i>	66
Figura 11 – Árvore filogenética das proteínas Rap de <i>B. cereus lato sensu</i> construída pelo algoritmo <i>Neighbour-Joining</i>	70

Figura 12 – Árvore filogenética dos peptídeos Phr de *B. cereus lato sensu* construída pelo algoritmo *Neighbour-Joining*.....71

LISTA DE TABELAS E QUADROS

Artigo

Tabela 1 – Dados gerais dos genomas de <i>Bacillus cereus lato sensu</i> empregados neste estudo.....	79
Tabela 2 – Relação dos genes <i>rap-phr</i> encontrados, juntamente com os genes localizados mais proximamente ao lado.....	82
Tabela 3 – Relação dos plasmídeos contidos nas linhagens de <i>B. cereus</i> (Bc), <i>B. thuringiensis</i> (Bt) e <i>B. anthracis</i> (Ba) empregadas neste estudo, bem como a localização do gene <i>rap</i> dentro do plasmídeo.	89
Tabela 4 – Distribuição dos genes <i>rap</i> nas linhagens de <i>B. subtilis</i> (Bs).....	57
Tabela 5 – Distribuição dos genes <i>rap</i> segundo as classes nas linhagens de <i>B. subtilis</i>	92
Tabela 6 – Características principais dos genes <i>rap</i> curtos de <i>B. cereus lato sensu</i>	60
Tabela 7 – Relação dos domínios encontrados nos três bancos de dados para cada um dos genes <i>rap</i> identificados.....	95
Quadro 1 – Resultado do BLASTp entre as proteínas Rap de <i>Bacillus subtilis</i> str.168	98
Quadro 2 – Resultado do BLASTp entre as proteínas Rap de <i>Bacillus cereus</i> ATCC 14579.....	99
Quadro 3 – Resultado do BLASTp entre as proteínas Rap de <i>Bacillus subtilis</i> kurstaki HD-73	99
Quadro 4 – Resultado do BLASTp entre os peptídeos Phr de <i>B. subtilis</i> str. 168.....	101
Quadro 5 – Resultado do BLASTp entre os peptídeos Phr de <i>B. thuringiensis</i> HD-73.....	102
Quadro 6 – Agrupamentos das proteínas Rap por similaridade	103

SUMÁRIO

1	FUNDAMENTAÇÃO TEÓRICA	12
1.1	REGULADORES DE RESPOSTA ASPARTATO-FOSFATASES (RAP) E SEUS INIBIDORES PHR	12
1.2	<i>BACILLUS CEREUS LATO SENSU</i>	19
1.2.1	Plasmídeos em <i>B. cereus lato sensu</i>	23
2	OBJETIVOS	26
2.1	OBJETIVO GERAL	26
2.2	OBJETIVOS ESPECÍFICOS	26
	REFERÊNCIAS	27
3	ARTIGO	30
3.1	RESUMO	31
3.2	INTRODUÇÃO	32
3.3	METODOLOGIA	34
3.3.1	Linhagens bacterianas	34
3.3.2	Levantamento dos genes	34
3.3.2.1	Genes <i>rap</i>	35
3.3.2.1	Genes <i>phr</i>	35
3.3.3	Caracterização dos genes/proteínas	35
3.3.3.1	Genes <i>rap</i>	35
3.3.3.2	Genes <i>phr</i>	36
3.3.4	Localização dos genes <i>rap-phr</i>	36
3.3.5	Alinhamento cromossômico	36
3.3.6	Análises de similaridade entre as sequências	37
3.3.6.1	Similaridade com proteínas de <i>B. subtilis</i>	37
3.3.6.2	Análises de similaridade entre proteínas Rap ou Phr - De uma mesma linhagem	37
3.3.6.3	Análises de similaridade entre proteínas Rap – Entre linhagens diferentes	37
3.3.6.4	Análises de similaridade entre proteínas Rap – Agrupamento	38
3.3.7	Análises filogenéticas	38

3.4	RESULTADOS E DISCUSSÃO	38
3.4.1	Linhagens bacterianas	38
3.4.2	Levantamento e localização dos genes <i>rap-phr</i>	44
3.4.2.1	Levantamento dos genes <i>rap-phr</i>	44
3.4.2.2	Localização dos genes <i>rap-phr</i>	46
3.4.2.3	Genes plasmidiais	51
3.4.2.3	Genes localizados à montante e à jusante ao cassete <i>rap-phr</i>	56
3.4.3	Caracterização dos genes <i>rap-phr</i>	57
3.4.3.1	Comparação com genes/proteínas de <i>B. subtilis</i>	57
3.4.3.2	Tamanho dos genes <i>rap</i> e <i>phr</i>	59
3.4.3.3	Genes <i>rap</i> curtos	60
3.4.3.4	Caracterização das sequências	62
3.4.4	Análise de similaridade de proteínas Rap	63
3.5	CONCLUSÃO	72
	REFERÊNCIAS	74
APÊNDICES	78

1 FUNDAMENTAÇÃO TEÓRICA

1.1 REGULADORES DE RESPOSTA ASPARTATO-FOSFATASES (RAP) E SEUS INIBIDORES PHR

A interação entre bactérias e meio ambiente passa pela ocorrência de diversos processos biológicos, como por exemplo: expressão de fatores de virulência – que permite a bactéria infectar (colonizar e se multiplicar) o hospedeiro; formação de biofilme - comunidades bacterianas estruturadas e dinâmicas, resultante da associação das bactérias com seus produtos extracelulares, e esporulação - formação do esporo que confere resistência a condições adversas, apresentando assim, um papel importante na manutenção e dispersão dessas bactérias.

É de grande importância para as células bacterianas de uma mesma linhagem, que coabitam um mesmo ambiente, que haja a coordenação dos mesmos, conforme a situação do meio, a partir das mudanças no ambiente. Desta forma, diversos processos celulares tem uma regulação minuciosa, como a esporulação, que é um processo complexo e dispendioso para a bactéria e, em algum ponto, irreversível e, portanto, iniciado apenas se não houver outras vias de sobrevivência disponíveis (PEREGO, 2013).

Todos esses processos, utilizados pelas bactérias para responder aos estímulos do meio ambiente, são controlados por sistemas de *quorum-sensing*. Esse sistema é um modo de comunicação entre as células que envolve a secreção de moléculas que são reconhecidas pelas bactérias cognatas. Nas bactérias Gram positivas, esta comunicação se dá por meio de sistemas baseados na secreção e reconhecimento de oligopeptídeos, o que permite às células bacterianas coordenar a expressão gênica envolvida em processos biológicos de acordo com a densidade populacional (PERCHAT et al., 2011; PEREGO et al., 1996).

Alguns dos sistemas regulados pelos mecanismos de *quorum sensing* fazem parte dos chamados sistemas de transdução de sinal de dois componentes. Nesses sistemas, um dos dois componentes é uma molécula sensora, geralmente uma histidina-quinase, que reconhece o sinal; e o outro, um regulador de resposta, que executa a ação, frequentemente um fator de transcrição (BRUNSING et al., 2005; PEREGO et al., 1996).

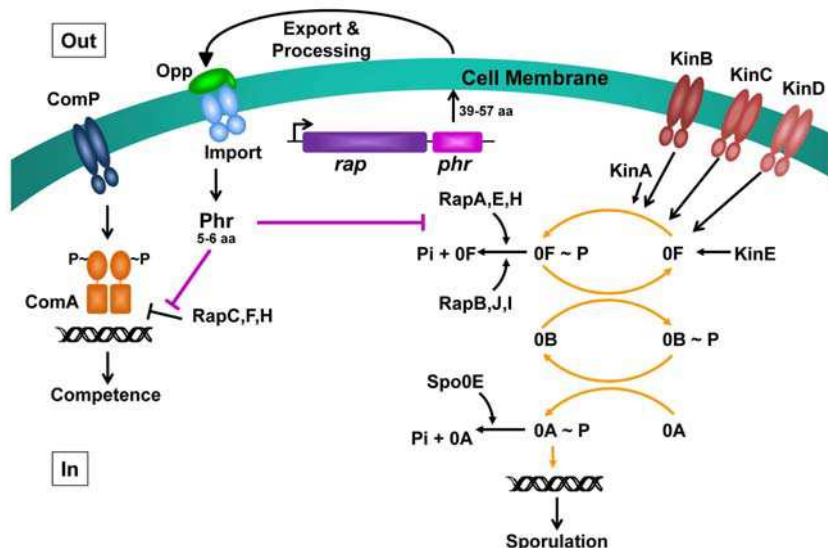
A cadeia de transferência de fosfato, ou *phosphorelay*, que regula a iniciação da esporulação é um exemplo de sistema de transdução de sinal de dois componentes. Essa cadeia apresenta regulação complexa, pois possuem fatores adicionais que atuam como fosfatases, os reguladores de resposta aspartato-fosfatase (Rap), com função de equilibrar os sinais recebidos com a resposta a ser gerada, neutralizando as proteínas quinases (BONGIORNI et al., 2005).

Nesse sistema, o estímulo atua sobre quinases (KinA a KinE) que fosforilam o regulador de resposta Spo0F, utilizado como substrato para a fosfotransferase Spo0B, que transfere o fosfato de Spo0F para Spo0A, responsável pela regulação da transcrição de genes envolvidos no início da esporulação (ISHIKAWA; CORE; PEREGO, 2002). As proteínas Rap atuam sobre esta via desfosforilando os reguladores de resposta Spo0F, que no estado desfosforilado não tem condições de fornecer à Spo0B o grupamento fosfato necessário para fosforilar Spo0A, o regulador de resposta final da via de esporulação. Outro sistema regulado pelas proteínas Rap é o sistema de transdução de sinal de dois componentes ComP-ComA, que controla o desenvolvimento de competência para transformação genética, ou seja, a habilidade de incorporar DNA exógeno, e a síntese de produtos secretados em *Bacillus subtilis*. ComP é um sensor histidina-quinase que ativa o regulador de resposta fator de transcrição ComA (AUCHTUNG; LEE; GROSSMAN, 2006).

A família das proteínas Rap tem sido amplamente estudada em *B. subtilis*, bactéria considerada como espécie modelo para o estudo desse sistema e na qual já foram descritas 11 proteínas Rap (RapA a RapK). Pesquisas com as proteínas Rap demonstram que RapA, RapB, RapE, RapI e RapJ atuam na via de esporulação, por meio da desfosforilação do regulador de resposta Spo0F da cadeia de transferência de fosfato; RapC e RapF regulam a indução de competência ao se ligarem ao domínio de ligação ao DNA do regulador de resposta ComA, impedindo sua atuação como fator de transcrição, e RapH age sobre ambas as vias. RapI também atua na regulação da mobilidade do transposon ICEBs1. RapG inibe a atividade de ligação ao DNA do regulador de resposta DegU que controla vários processos celulares, inclusive a produção de proteases extracelulares. Enquanto que RapD e RapK ainda não tiveram suas funções reveladas (Figura 1) (AUCHTUNG; LEE; GROSSMAN, 2006; BONGIORNI et al., 2005; PEREGO, 2013).

A atividade das proteínas Rap é regulada negativamente por peptídeos denominados Phr. A maioria dos genes *rap* tem um gene *phr* localizado imediatamente à jusante, formando um operon. Os dois genes estão geneticamente ligados e ligeiramente sobrepostos. A expressão gênica de *phr* é regulada pelo promotor localizado à montante de *rap* e também, na maioria dos casos, por um promotor próprio situado no interior dos genes *rap*, reconhecido pela RNA polimerase que contém o fator σ^H , o qual aumenta a quantidade de Phr transcrito durante a fase estacionária (Figura 2). A transcrição regulada pelo promotor adicional pode assegurar que a quantidade do oligopeptídeo Phr durante esta fase seja suficiente para inibir a proteína Rap (McQUADE; COMELLA; GROSSMAN, 2001; POTTATHIL; LAZAZZERA, 2003).

Figura 1 – Representação esquemática das vias de sinalização reguladas por módulos Rap-Phr. Do lado esquerdo está representado a via de transdução de sinal de dois componentes (ComP-ComA) que regula a competência. Enquanto que do lado esquerdo está a via de transferência de fosfato (phosphorelay) responsável pela regulação da iniciação da esporulação.



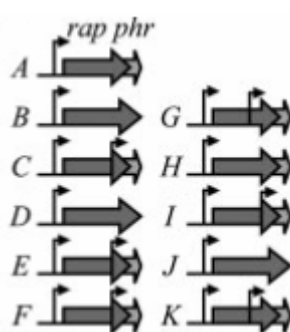
Fonte: PEREGO, 2013

Oito dos onze genes *rap* descritos para *B. subtilis* possuem o gene *phr* correlato (Figura 2). O produto do gene *phr* é um pequeno peptídeo de 39 a 57 aminoácidos, o qual é secretado, processado para a forma de oligopeptídeo (normalmente pentapeptídeo) e re-importado para dentro da célula por meio de transportadores Oligopeptídeo permease (Opp), atuando como sensor de quórum.

Em geral, a atividade de um determinado peptídeo Phr é específica para sua proteína Rap correlata, assim PhrA inibe especificamente a proteína RapA. Com exceção do peptídeo PhrC, também conhecido como CSF (de fator estimulante de competência), que também inibe a atividade das proteínas RapB e RapJ não correlatas (AUCHTUNG; LEE; GROSSMAN, 2006; BONGIORNI et al., 2005; PARASHAR; JEFFREY; NEIDITCH, 2013; PEREGO, 1996, 2013).

Os peptídeos pré-Phr apresentam características estruturais típicas de proteínas exportadas, pois possuem sequência sinal para exportação, com sítio de clivagem. Provavelmente são exportados pelo sistema dependente de Sec (complexo translocase) e processados por peptidases sinal tipo I. Acredita-se que um segundo evento de processamento ocorra sob ação de uma das três proteases redundantes subtilisina, Epr (protease serina extracelular) ou Vpr (protease extracelular menor). A sequência correspondente ao peptídeo Phr maduro normalmente se localiza na porção C-terminal, mas as diferentes classes do peptídeo não são processadas de maneira idêntica (LANINGAN-GERDES et al., 2008; PEREGO; BRANNIGAN, 2001; POTTATHIL; LAZAZZERA, 2003).

Figura 2- Representação esquemática dos operons *rap-phr*. As letras que antecedem cada operon estão identificando as classes dos genes *rap*, as setas largas e escuras representam os genes *rap*, enquanto que as largas e claras retratam os genes *phr*. As setas finas correspondem aos promotores.



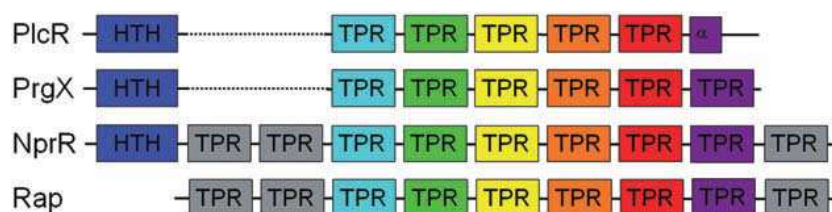
Fonte: Bischof, 2009

Uma vez que a competência e a esporulação são antagônicas, as bactérias desenvolveram mecanismos de controle cruzado para garantir a sucessão correta dos eventos. No início do crescimento exponencial, quando a densidade celular ainda é baixa, RapC (aqui generalizando os outros genes *rap* que atuam na

esporulação) inibe o desenvolvimento de competência ao se ligar no domínio de ligação ao DNA de ComA. O acúmulo de peptídeos PhrC, já no final da fase exponencial, inativa RapC, estimulando a competência. Durante a transição entre as fases de crescimento exponencial e estacionária (pós-exponencial), ComA estimula a expressão de RapA (generalizando os genes *rap* que atuam na esporulação), que ao desfosforilar Spo0F, reprime a esporulação. Durante a fase estacionária, quando normalmente a disponibilidade de nutrientes se torna limitada e ocorre a indução da transcrição dos peptídeos Phr a partir do fator σ^H da RNA polimerase, RapA é inibida por seu peptídeo Phr correlato, permitindo o início da esporulação (KOETJE et al., 2003).

As proteínas Rap fazem parte de uma família de proteínas denominada RNPP, cujo nome é denominado da fusão dos nomes das proteínas Rap, NprR - fator de transcrição necrotrófico, PlcR - regulador pleiotrópico de transcrição de fatores de virulência e PrgX - repressor transcricional. Essas proteínas estão presentes no gênero *Bacillus* e, apresentam em comum a modulação por pequenos peptídeos, conforme a densidade populacional, caracterizando-as como sistemas de *quorum sensing*. Os membros da família RNPP compartilham a mesma organização estrutural contendo repetições tetratricopeptídicas, discutida mais detalhadamente abaixo. A proteína Rap se diferencia das outras proteínas do grupo RNPP pela ausência do domínio HTH (hélice-giro-hélice) de ligação ao DNA, normalmente presente em fatores de transcrição (Figura 3) (PERCHAT et al., 2011; SLAMTI et al., 2014).

Figura 3 – Representação esquemática da organização estrutural das proteínas da família RNPP. TPR: repetições tetratricopeptídicas; HTH: domínio de ligação ao DNA hélice-giro-hélice.



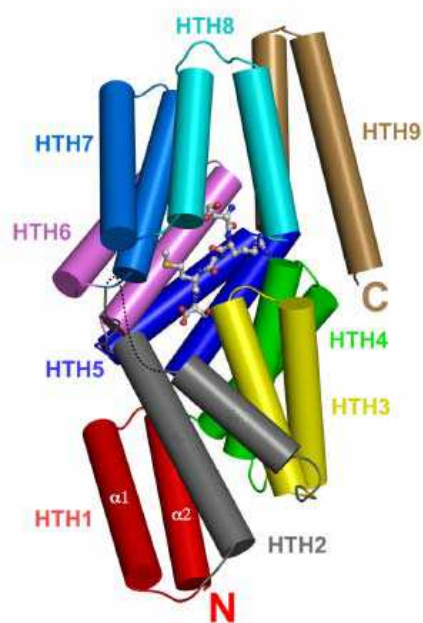
Fonte: Perchat, 2011

A repetição tetratricopeptídica é um motivo estrutural presente em uma grande variedade de proteínas identificadas em diversos organismos, de bactérias a humanos, e envolvidas em vários processos biológicos, como regulação do ciclo

celular, controle transcricional, proteínas de transporte mitocondrial e peroxissomal e dobramento de proteínas. Sua repetição básica é constituída por 34 aminoácidos, que podem estar dispersos na proteína ou em repetições de 3 a 16 vezes em tandem. Sua sequência consenso não é definida por resíduos específicos, mas por uma preferência por uma classe de aminoácidos (com resíduo pequeno, grande ou aromático). Os motivos TPR, em geral, adotam um arranjo estrutural hélice-giro-hélice (HTH, que não se liga ao DNA, e que é distinto do domínio HTH anteriormente citado) com um ângulo de empacotamento entre as duas hélices de $\sim 24^\circ$, importante para sua atuação em interações proteína-proteína ou na formação de complexos multiproteicos (D'ANDREA; REGAN, 2003).

A maioria dos estudos definem o motivo TPR presente nas proteínas Rap como sendo composto por seis domínios TPR (Figura 4). Precedendo esse motivo TPR estão três hélices no domínio N-terminal da proteína. O peptídeo Phr se liga a proteína Rap correlata em uma conformação estendida ao longo da superfície côncava formada pelo domínio TPR. Assim, o sítio de ligação para os peptídeos Phr nas proteínas Rap estão em suas porções C-terminal. A ligação do peptídeo induz uma constrição que se propaga para o agrupamento de três hélices N-terminal, resultando em uma alteração estrutural caracterizada por um domínio composto por nove repetições semelhantes à TPR (D'ANDREA; REGAN, 2003; DIAZ et al., 2012; PEREGO, 2013).

Figura 4 - Estrutura espacial da proteína RapJ ligada ao peptídeo PhrC de *B. subtilis*. Os domínios TPR de RapJ estão retratados em diferentes cores e são chamados aqui de HTH. As α -hélices estão representadas como cilindros (como destacada nas hélices de HTH1). N: terminal amino; C: terminal carboxil. A molécula de PhrC está retratada pelo modelo de esferas e bastões ('ball and stick').



RapJ-PhrC

Fonte: Parashar, 2013.

Os genes *rap-phr* também estão presentes nos genomas de *Bacillus cereus lato sensu* (*Bacillus cereus sensu stricto*, *Bacillus thuringiensis*, *Bacillus anthracis*), contudo, ainda há poucos estudos envolvendo este sistema neste grupo. Entre eles, um estudo que utilizou a sequência de aminoácidos da proteína RapA de *B. subtilis* 168 para uma pesquisa de sequências homólogas em *B. anthracis* A2012, detectando cinco prováveis genes *rap*. Desses, apenas o produto de dois dos genes *rap* afetaram a esporulação em testes funcionais (BONGIORNI et al., 2006).

Slamti et al. (2014) identificaram possíveis genes *rap* no genoma de *B. thuringiensis* 407 utilizando análises de BLAST com a sequência de aminoácidos de uma das proteínas Rap identificadas em *B. anthracis* por Bongiorno et al. (2006). Por análises comparativas os autores identificaram que pelo menos três sistemas Rap-Phr desta linhagem podem estar envolvidos no controle da esporulação da bactéria.

Devido à pouca produção científica no grupo *B. cereus* envolvendo esses genes, à importância dos mesmos e supondo-se que as proteínas Rap e os peptídeos Phr de *B. cereus lato sensu* tenham funções similares às descritas em *B. subtilis*, torna-se interessante pesquisar a presença desses genes nas espécies do grupo. Inclusive, investigar a possível utilização desses genes em estudos envolvendo as questões taxonômicas desse grupo, as quais continuam sendo tema de debates, desde a década de 1990.

1.2 BACILLUS CEREUS LATO SENSU

O grupo do *Bacillus cereus* abrange bactérias Gram positivas formadoras de esporos com células em forma de bastonete. Essas bactérias são encontradas em diversos habitats, como solo, folhas ou rizosfera de plantas e capazes de colonizar diversos hospedeiros. Possuem importância econômica e/ou médica e são estreitamente relacionadas geneticamente (JENSEN et al., 2003; TOURASSE et al., 2011; VILAS-BÔAS; PERUCA; ARANTES, 2007). Atualmente, sete espécies fazem parte deste grupo: *Bacillus cereus (sensu stricto)*, *Bacillus thuringiensis*, *Bacillus anthracis*, *Bacillus weihenstephanensis*, *Bacillus mycoides*, *Bacillus pseudo-mycoides*, e a mais recentemente reconhecida como pertencente ao grupo, *Bacillus cytotoxicus* (GUINEBRETIERE et al., 2013; VILAS-BÔAS; PERUCA; ARANTES, 2007).

A espécie tipo do grupo, *B. cereus* é uma bactéria do solo, provavelmente ubíqua, e um patógeno oportunista, causa comum de intoxicação alimentar e contaminação hospitalar. *B. thuringiensis* produz, durante a esporulação, cristais proteicos tóxicos para uma grande variedade de insetos, sendo a base dos pesticidas biológicos mais utilizados mundialmente. *B. anthracis* é o agente patogênico da doença anthrax que atinge mamíferos, inclusive humanos e, devido a sua elevada toxicidade, já foi utilizada como arma biológica. *B. weihenstephanensis* é uma bactéria tolerante ao frio, podendo ser encontrada como contaminante de alimentos congelados. *B. mycoides* podem melhorar o crescimento de plantas e *B. pseudo-mycoides* já foi identificada em infecções em humanos como endoftalmite. *B. cytotoxicus* é uma bactéria termotolerante associada à intoxicação alimentar (GUINEBRETIERE et al., 2013; HELGASON et al., 2000; JENSEN et al., 2003; TOURASSE et al., 2011).

Os membros do grupo *B. cereus* exibem genomas complexos, pois a maioria das linhagens carregam plasmídeos em diferentes números e tamanhos e possuem um cromossomo com sequências repetitivas e numerosos componentes genéticos móveis, como sequências de inserção, transposons e prófagos, indicando um genoma dinâmico (GILLIS; MAHILLON, 2014; KOLSTØ; TOURASSE; ØKSTAD, 2009).

Bacillus cereus, *B. thuringiensis* e *B. anthracis* apresentam genomas altamente similares, o que as torna taxonomicamente muito próximas, apesar de demonstrarem patótipos muito diferentes. A distinção entre estas espécies é excepcionalmente complexa, visto que os determinantes genéticos específicos para caracterização de linhagens como pertencentes à espécie *B. thuringiensis* ou à *B. anthracis* estão localizados em plasmídeos. Os genes *cry* (codificantes das proteínas inseticidas) de *B. thuringiensis* e os genes formadores de toxinas e de cápsula de *B. anthracis*, principais determinantes da patogenicidade dessas espécies, estão localizados em grandes plasmídeos, frequentemente conjugativos, os quais podem ser facilmente transferidos para outras espécies do grupo *B. cereus* através de mecanismos de transferência gênica horizontal (conjugação) (HELGASON et al., 2000; KOLSTØ; TOURASSE; ØKSTAD, 2009; VILAS-BÔAS; PERUCA; ARANTES, 2007).

A complexidade das relações filogenéticas entre as três espécies é tal que não há um consenso se devem ser consideradas espécies diferentes ou linhagens especializadas da mesma espécie, podendo ser denominadas de *B. cereus lato sensu*. Atualmente, são consideradas espécies diferentes (Bergey's Manual of Systematic Bacteriology – VOS et al., 2009), embora alguns autores propõem que *B. cereus stricto sensu* seja considerado como a espécie-tipo, e *B. thuringiensis* e *B. anthracis* como subespécies ou linhagens, as quais codificam suas principais características fenotípicas a partir de DNA extracromossomal. Por exemplo, caso os plasmídeos contendo os genes *cry* de *B. thuringiensis* sejam perdidos, a bactéria torna-se indistinguível de *B. cereus*, pois a única característica que distingue as duas espécies é dada pelos genes *cry*, os quais estão localizados em plasmídeos, e frequentemente associados a transposons. Situação similar também é encontrada para *B. anthracis* (HELGASON et al., 2000; MAHILLON et al., 1994; VILAS-BÔAS; PERUCA; ARANTES, 2007).

Assim, na busca da resolução dessa questão taxonômica, diversos autores utilizando métodos variados tentaram esclarecer a questão. A análise das sequências nucleotídicas do rDNA 16S das três espécies apresenta apenas 1% de divergência, abaixo dos 3% de divergência que tradicionalmente caracterizam espécies como diferentes. Experimentos utilizando eletroforese de enzimas em multi-locos (MLEE) (HELGASON et al., 1998), sequência de genes distintos codificadores de proteínas (HELGASON et al., 2000), amplificação de fragmentos de tamanhos polimórficos (AFLP) (TICKNOR et al., 2001) e tipificação de sequências de múltiplos locos (MLST) sugerem que *B. thuringiensis* e *B. anthracis* devem ser classificadas como subespécies da espécie *B. cereus*. Embora a sintenia cromossômica entre os genomas de *B. cereus lato sensu* seja significativa, estudos que demonstram diferenças na ordem dos genes, rearranjos cromossômicos, variações nucleotídicas e fagos remanescentes não permitem verificar a ocorrência de diferenças espécie específicas (ANDERSON et al., 2005).

Tourasse et al. (2011) reuniram as metodologias de análises genotípicas no banco de dados HyperCAT, combinando os dados de MLST, AFLP e MLEE, resultando em uma análise filogenética em que os isolados são classificados em sete grandes grupos. A informação gerada corrobora com os estudos de tipagem anteriores ao refletir a heterogeneidade do grupo *B. cereus*, formando um grupo filogenético no qual seus membros estão frequentemente misturados.

Análises comparativas sugerem que as principais diferenças entre os membros de *B. cereus lato sensu* se devem mais a discretas alterações na expressão gênica do que no nível de divergência de sequência ou conteúdo gênico (RASKO et al., 2005). Os fatores sigma fazem parte do aparato da RNA polimerase que controla o reconhecimento do promotor. A divergência na sequência do respectivo gene, portanto, pode ter um papel importante para explicar a diversidade das características do grupo. Além disso, as sequências regulatórias e de redes de transcrição são as porções do genoma bacteriano mais flexíveis evolutivamente. A análise filogenética das sequências dos genes de fator sigma revelou que as espécies formam subgrupos relacionados, sugerindo que sejam membros da mesma espécie (SCHMIDT et al., 2011).

Por outro lado, alguns trabalhos identificaram divergência genética suficiente para distinguir *B. anthracis* de *B. cereus* e *B. thuringiensis* devido à baixa variabilidade genética de *B. anthracis*, espécie que apresenta ainda propriedades

fenotípicas e marcadores genéticos que resultam num diferencial em relação às duas outras espécies. As propriedades fenotípicas, ausência de atividade hemolítica, de motilidade, e sensibilidade ao fago γ , são consequências de uma mutação pontual no regulador pleiotrópico PlcR. Em acréscimo, os cromossomos de *B. anthracis* possuem quatro regiões cromossômicas que parecem ser específicas da espécie, correspondentes a profagos (Lambda 01-04). No entanto, além dessas regiões não há nenhuma sequência cromossômica específica que possa ser utilizada para distinguir a espécie dos outros membros de *B. cereus lato sensu* (KOLSTØ; TOURASSE; ØKSTAD, 2009).

A diferenciação entre *B. cereus* e *B. thuringiensis* foi identificada quando as metodologias levaram em consideração o nicho ecológico das espécies. Pois, espera-se que diferentes espécies bacterianas possuam propriedades genéticas, ecológicas, evolucionárias e filogenéticas específicas (COHAN, 2002).

Em concordância com este pensamento, Vilas-Bôas et al. (2002) realizaram um estudo com amostras de solo de *B. cereus* e *B. thuringiensis* provenientes de dois locais diferentes, e observaram por meio de MEE que as populações de cada uma das espécies são geneticamente mais semelhantes entre si do que em relação às populações simpátricas da outra espécie. Verificaram também que a taxa de fluxo gênico é maior entre os isolados da mesma espécie, do que entre as espécies. Contudo, a troca de material genético entre as duas espécies é possível.

Sabendo da possibilidade da troca de material genético entre *B. cereus* e *B. thuringiensis*, Santos et al. (2010) analisaram a transferência de um plasmídeo, por conjugação, entre linhagens das duas espécies e a multiplicação de células vegetativas em larvas do inseto *Bombyx mori*. Os resultados demonstraram que embora a linhagem doadora do plasmídeo não tenha preferência para intercâmbio genético entre as linhagens de *B. cereus* ou *B. thuringiensis* receptoras, a presença de células vegetativas é necessária para a conjugação, ou seja, as duas bactérias (doadora e receptora) precisam compartilhar o mesmo nicho ecológico. Entretanto, as linhagens de *B. cereus* germinam e se multiplicam menos eficientemente em *B. mori* do que as linhagens de *B. thuringiensis*, o que demonstrou diferenças ecológicas importantes entre linhagens das duas espécies.

A análise do gene *csaB* (necessário para ancoragem de proteínas na superfície celular) foi realizada considerando que as análises filogenéticas baseadas em genes conservados apresentaram filogenia incongruente, buscou-se, assim,

obter uma maior discriminação entre as espécies do grupo por meio de genes com maior diversidade genética. Entretanto, essa análise não foi capaz de distinguir as 122 linhagens de *B. cereus lato sensu* quanto à espécie de origem, mas reuniu-as em dois grupos com relação ao hospedeiro de onde foram isoladas, mamíferos ou insetos; refletindo a pressão seletiva dos hospedeiros. Linhagens isoladas do solo se distribuíram igualmente entre os dois grupos, ou seja, o solo parece não exercer uma pressão seletiva sobre as linhagens, parecendo ser reservatório das linhagens de ambos hospedeiros (ZHENG et al., 2013).

Embora diversos métodos tenham sido adotados na investigação da questão taxonômica envolvendo essas espécies, os resultados são controversos quanto à taxonomia do grupo variando conforme as técnicas e linhagens utilizadas. Desta forma, novos estudos para auxiliar na solução dessas questões taxonômicas, devem ser conduzidos, principalmente por meio de metodologias que permitam uma análise baseada no nicho ecológico das bactérias, ou seja, que levem em consideração a origem dos isolados empregados, ou o comportamento dos mesmos nos diferentes ambientes. Assim, a partir dos resultados já obtidos, genes que estejam de alguma forma envolvidos na interação da bactéria com seu hospedeiro, ou com o ambiente como um todo, são fortes candidatos a desempenhar essa função.

1.2.1 Plasmídeos em *B. cereus lato sensu*

Como já apresentado na seção anterior, as principais características fenotípicas das espécies de *B. cereus lato sensu*, inclusive as fundamentais para patogenicidade e resultantes da adaptação ao meio, são codificadas nos plasmídeos. Como a proteína cereulide em *B. cereus stricto sensu* e as já descritas proteínas Cry de *B. thuringiensis* e toxinas e cápsula de *B. anthracis* (TOURASSE et al., 2011).

A maioria das linhagens do grupo *B. cereus* contém vários plasmídeos, com uma variação considerável em número (de zero a mais de 10) e no tamanho (de 2 a 600 kb) desses replicons (KOLSTØ; TOURASSE; ØKSTAD, 2009; VILAS-BÔAS; LEMOS, 2004; ZHENG et al., 2013). Análises do perfil plasmidial de linhagens de *B. thuringiensis* revelaram uma grande variedade e complexidade do conteúdo dos plasmídeos e de sua importância para esta espécie (VILAS-BÔAS; LEMOS, 2004); enquanto que as análises genéticas de *B. anthracis* revelam a presença de apenas

dois plasmídeos, pXO1 e pXO2, característicos da espécie (KOLSTØ; TOURASSE; ØKSTAD, 2009).

Devido a estas características, o grupo do *B. cereus* vem sendo usado como modelo para a realização de estudos que questionam a visão corrente de que os plasmídeos são parte integrante do genoma de uma espécie bacteriana, como apenas elementos genéticos acessórios que fornecem vantagens seletivas às células que os abrigam. Uma nova linha de pensamento coloca os plasmídeos como elementos genéticos autônomos, como bacteriófagos ou vírus, que podem permanecer em células bacterianas de diversas espécies que coexistem em um determinado ambiente (BAHL; HANSEN; SØRENSEN, 2009; RAYMOND et al., 2010).

Raymond et al. (2012) realizaram estudos com a linhagem *B. thuringiensis kurstaki* HD1, comparando populações bacterianas com proporções variadas de células da mesma linhagem, produtoras e não produtoras da toxina Cry. A produção da toxina Cry e a manutenção do plasmídeo que contém os genes *cry* elevam os custos energéticos para a célula bacteriana, assim as bactérias não produtoras da toxina se multiplicam mais rapidamente que as produtoras. Entretanto, a produção da toxina é benéfica tanto para as células produtoras quanto para as não produtoras. Assim, ambos os tipos celulares (produtoras e não produtoras) são favorecidos pela ação das toxinas Cry, pois esta é necessária para virulência da bactéria, permitindo a invasão da hemolinfa do hospedeiro, tanto pelas células produtoras de cristal, como pelas não produtoras. Essas últimas, no entanto, exatamente por não ter que suportar os custos de manutenção do plasmídeo que contém o gene *cry*, multiplicam-se mais rapidamente, o que permite à população bacteriana, a tomada mais rápida e eficaz do cadáver do inseto, em relação a uma situação em que houvessem apenas células portadoras do plasmídeo.

Embora Raymond et al. (2012) tenham trabalhado apenas com uma única linhagem bacteriana com duas sub-populações, uma produtora e outra não produtora de toxina Cry, a união dos resultados desse estudo com os obtidos por Santos et al. (2010), demonstra a importância do processo de comunicação celular ou *quorum sensing*, em um determinado ambiente. Assim, um plasmídeo pode ser mantido simultaneamente em diferentes linhagens bacterianas através do intercâmbio genético entre as mesmas num mesmo ambiente. No entanto, ele não precisa ser mantido em todas as células de uma determinada linhagem bacteriana.

Zheng et al. (2015) em uma análise genética dos plasmídeos do grupo *B. cereus* revelou 598 famílias gênicas compartilhadas por plasmídeos e cromossomos. Dessas famílias gênicas, as envolvidas na replicação, recombinação e reparo representavam a maior proporção dos genes plasmidiais, que continham também, em grande quantidade, famílias envolvidas na transcrição, modificações pós-transcricionais e dobramento de proteínas. Menos frequentemente estavam presentes famílias gênicas envolvidas no metabolismo basal e na tradução. Estes autores observaram também que quando os genes estavam presentes no plasmídeo e no cromossomo na mesma célula, geralmente apresentavam sequências promotoras e terminadoras diferentes, sugerindo que são transcricionalmente regulados por diferentes elementos.

Em geral, estudos sobre genes contidos em plasmídeos salientam os genes que conferem vantagem às bactérias, como resistência a antibióticos ou fatores de virulência. Atualmente, no entanto, relatos de plasmídeos contendo genes essenciais ao metabolismo da célula têm sido mais comumente encontrados. Como é o caso da bactéria fixadora de nitrogênio *Rhizobium etli*, na qual alguns genes e vias metabólicas completas estão localizadas em seus plasmídeos (GONZÁLES et al., 2006). Ou ainda, genes que codificam o regulador de resposta aspartato-fosfatase (*rap*) em espécies do gênero *Bacillus* (BONGIORNI et al., 2006; KOETJE et al., 2003; PARASHAR et al., 2013). A presença de genes essenciais em plasmídeos de outras espécies além do grupo *B. cereus*, demonstram que os resultados obtidos dentro desse grupo poderão ser transpostos para outros grupos bacterianos, o que poderá permitir uma mudança quanto a forma de pensar no papel dos plasmídeos na evolução do genoma bacteriano e no comportamento da bactéria no ambiente.

A disponibilidade de sequências do genoma completo, inclusive dos plasmídeos, de diversas linhagens de *B. cereus lato sensu* possibilita novas pesquisas para tentar compreender melhor as relações filogenéticas entre as três espécies e como os plasmídeos devem ser considerados. Neste contexto, os genes *rap*, cujos produtos regulam processos biológicos essenciais para as espécies do gênero *Bacillus*, como esporulação, competência genética e formação de biofilme e que ainda não foram explorados nas espécies do grupo *B. cereus*, foram utilizados como possível ferramenta para acrescentar dados a essas discussões, envolvendo tanto a questão taxonômica, quanto o papel dos plasmídeos no grupo.

2 OBJETIVOS

2.1 OBJETIVO GERAL

Identificar e caracterizar os genes *rap-phr* por meio de ferramentas computacionais nos cromossomos e plasmídeos de *B. cereus lato sensu*, visando avançar no conhecimento quanto à ocorrência e variabilidade desses genes, bem como, auxiliar na compreensão do papel dos plasmídeos nesse grupo bacteriano.

2.2 OBJETIVOS ESPECÍFICOS

- Identificar as linhagens de *B. cereus*, *B. thuringiensis* e *B. anthracis* com sequência do genoma completo disponível em banco de dados público;
- Obter as sequências nucleotídicas e proteicas dos sistemas *rap-phr* dessas linhagens;
- Realizar a caracterização molecular das sequências, com relação a tamanho e características específicas dessas proteínas;
- Identificar a localização cromossômica e plasmidial dos genes *rap-phr*, bem como os genes localizados à montante e à jusante dos mesmos;
- Comparar as sequências identificadas como *rap-phr* com os sistemas de *B. subtilis*;
- Analisar a distribuição e similaridade dos sistemas *rap-phr* dentro de uma mesma linhagem e entre diferentes linhagens de *B. cereus lato sensu*;
- Analisar filogeneticamente as proteínas selecionadas observando as relações taxonômicas entre as três espécies, agrupando-as em classes.

REFERÊNCIAS

- ANDERSON, I. et al. Comparative genome analysis of *Bacillus cereus* group genomes with *Bacillus subtilis*. **FEMS Microbiology Letters**, v.250, p. 175–184, 2005.
- AUCHTUNG, J. M.; LEE, C. A.; GROSSMAN, A. D. Modulation of the comA-dependent quorum response in *Bacillus subtilis* by multiple Rap proteins and Phr peptides. **Journal of Bacteriology**, v.188, n.14, p. 5273–5285, jul. 2006.
- BAHL, M. I.; HANSEN, L. H.; SØRENSEN, S. J. Persistence mechanisms of conjugative plasmids. In: **Horizontal Gene Transfer: Genomes in Flux**. Humana Press, 2009. p. 73-102.
- BISCHOF, I. B., et al. Complexity in bacterial cell– cell communication: Quorum signal integration and subpopulation signaling in the *Bacillus subtilis* phosphorelay. **PNAS**, v.106, n.16, p. 6459–6464, abr. 2009.
- BONGIORNI, C. et al. Synergistic regulation of competence development in *Bacillus subtilis* by two Rap-Phr systems. **Journal of Bacteriology**, v.187, n.13, p. 4353–4361, jul. 2005.
- BONGIORNI, C. et al. Rap phosphatase of virulence plasmid pXO1 inhibits *Bacillus anthracis* sporulation. **Journal of Bacteriology**, v.188, n.2, p. 487–498, jan. 2006.
- BRUNSING, R. L. et al. Characterization of sporulation histidine kinases of *Bacillus anthracis*. **Journal of Bacteriology**, v.187, n.20, p. 6972–6981, out. 2005.
- COHAN, F.M. what are bacterial species? **Annual Review of Microbiology**, v.56, p.457–487, 2002.
- D'ANDREA, L. D.; REGAN, L. TPR proteins: the versatile helix. **TRENDS in Biochemical Sciences**, v.28, n.12, p.655-662, dez. 2003.
- DIAZ, A. R. *Bacillus subtilis* RapA phosphatase domain interaction with its substrate, phosphorylated Spo0F, and its inhibitor, the PhrA peptide. **Journal of Bacteriology**, v.194, n.6, p. 1378–1388, mar.2012.
- GILLIS, A.; MAHILLON, J. Phages preying on *Bacillus anthracis*, *Bacillus cereus*, and *Bacillus thuringiensis*: past, present and future. **Viruses**, v.6, p. 2623-2672, 2014.
- GUINEBRETIERE, M.-H., et al. *Bacillus cytotoxicus* sp. nov. is a novel thermotolerant species of the *Bacillus cereus* Group occasionally associated with food poisoning. **International Journal of Systematic and Evolutionary Microbiology**, v.63, p. 31–40, 2013.
- GONZÁLEZ, V. The partitioned *Rhizobium etli* genome: Genetic and metabolic redundancy in seven interacting replicons. **PNAS**, v.103, n.10, p. 3834–3839, mar. 2006.

HELGASON, E. et al. Genetic diversity of *Bacillus cereus*/*Bacillus thuringiensis* isolates from natural sources. **Current Microbiology**, v.37, p. 80–87, 1998.

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, jun. 2000.

ISHIKAWA, S.; CORE, L.; PEREGO, M. Biochemical characterization of aspartyl phosphate phosphatase interaction with a phosphorylated response regulator and its inhibition by a pentapeptide. **The Journal of Biological Chemistry**, v.277, n. 23, p. 20483–20489, jun. 2002.

JENSEN, G. B. et al. The hidden lifestyles of *Bacillus cereus* and relatives. **Environmental Microbiology**, v.5, n.8, p. 631–640, 2003.

KOETJE, E. J. et al. A plasmid-borne Rap–Phr system of *Bacillus subtilis* can mediate cell-density controlled production of extracellular proteases. **Microbiology**, v.149, p. 19–28, set. 2003.

KOLSTØ, A.-B.; TOURASSE, N. J.; ØKSTAD, O. A. What sets *Bacillus anthracis* apart from other *Bacillus* species? **The Annual Review of Microbiology**, v.63, p. 451–76, 2009.

LANIGAN-GERDES, S., et al. Identification of residues important for cleavage of the extracellular signaling peptide CSF of *Bacillus subtilis* from its precursor protein. **Journal of Bacteriology**, v.190, n.20, p. 6668–6675, out. 2008.

McQUADE, R. S.; COMELLA, N.; GROSSMAN, A. D. Control of a family of phosphatase regulatory genes (*phr*) by the alternate sigma factor sigma-H of *Bacillus subtilis*. **Journal of Bacteriology**, v.183, n.16, p. 4905–4909, ago. 2001.

PARASHAR, V. et al. A plasmid-encoded phosphatase regulates *Bacillus subtilis* biofilm architecture, sporulation, and genetic competence. **Journal of Bacteriology**, v.195, n.10, p. 2437–2448, maio 2013.

PARASHAR, V.; JEFFREY, P. D.; NEIDITCH, M. B. Conformational change-induced repeat domain expansion regulates Rap phosphatase Quorum-Sensing signal receptors. **PLOS Biology**: v.11, n.3, mar. 2013.

PERCHAT, S. et al. A cell–cell communication system regulates protease production during sporulation in bacteria of the *Bacillus cereus* group. **Molecular Microbiology**, v.82, n.3, p. 619–633, set. 2011.

PEREGO, M. et al. Aspartyl-phosphate phosphatases deactivate the response regulator components of the sporulation signal transduction system in *Bacillus subtilis*. **Molecular Microbiology**, v.19, n.6, p. 1151-1157, 1996.

PEREGO, M.; BRANNIGAN, J. A. Pentapeptide regulation of aspartyl-phosphate phosphatases. **Peptides**, v.22, p. 1541–1547, 2001.

PEREGO, M. Forty years in the making: understanding the molecular mechanism of peptide regulation in bacterial development. **PLOS Biology**, v.11, n. 3, mar. 2013.

RASKO, D. A., et al. Genomics of the *Bacillus cereus* group of organisms. **FEMS Microbiology Reviews**, v.29, p. 303–329, 2005.

RAYMOND, B. et al. Environmental factors determining the epidemiology and population genetic structure of the *Bacillus cereus* group in the field. **PLoS Pathogens**, v.6, n.5, p. 1-13, maio 2010.

RAYMOND, B. et al. The dynamics of cooperative bacterial virulence in the field. **Science**, v.337, n.85, p. 85-88, jul. 2012.

SANTOS, C. A. et al. Conjugal transfer between *Bacillus thuringiensis* and *Bacillus cereus* strains is not directly correlated with growth of recipient strains. **Journal of Invertebrate Pathology**, v.105, p. 171–175, jul. 2010.

SCHMIDT, T. R.; SCOTT, E. J.; DYER, D. W. Whole-genome phylogenies of the family *Bacillaceae* and expansion of the sigma factor gene family in the *Bacillus cereus* species-group. **BMC Genomics**, v.12, n.430, p. 1-16, 2011.

SLAMTI, L et al. Quorum sensing in *Bacillus thuringiensis* is required for completion of a full infectious cycle in the insect. **Toxins**, v.6, p. 2239-2255, jul. 2014.

TICKNOR, L. O. et al. Fluorescent amplified fragment length polymorphism analysis of Norwegian *Bacillus cereus* and *Bacillus thuringiensis* soil isolates. **Applied and Environmental Microbiology**, v.67, p. 4863–4873, out. 2001.

TOURASSE, N. J. et al., Extended and global phylogenetic view of the *Bacillus cereus* group population by combination of MLST, AFLP, and MLEE genotyping data. **Food Microbiology**, v.28, p. 236-244, 2011.

VILAS-BOAS, G. et al. Genetic differentiation between sympatric populations of *Bacillus cereus* and *Bacillus thuringiensis*. **Applied and environmental microbiology**, v.68, n.3, p. 1414–1424, mar. 2002.

VILAS-BÔAS, G. T.; LEMOS, M. V. F. Diversity of *cry* genes and genetic characterization of *Bacillus thuringiensis* isolated from Brazil. **Canadian Journal of Microbiology**, v.50, p. 605-613, set. 2004.

VILAS-BÔAS, G.; PERUCA, A.; ARANTES, O. Biology and taxonomy of *Bacillus cereus*, *Bacillus anthracis*, and *Bacillus thuringiensis*. **Canadian Journal of Microbiology**, v.53, n. 6, p. 673-87, jun. 2007.

VOS, P., et al. **Bergey's Manual of Systematic Bacteriology**. 2. ed. New York: Springer, 2009. v. 3.

ZHENG, J., et al. Differentiation of *Bacillus anthracis*, *B. cereus*, and *B. thuringiensis* on the Basis of the *csaB* Gene Reflects Host Source. **Applied and Environmental Microbiology**, v.79, n.12, p. 3860–3863, jun. 2013.

ZHENG, J., et al. Plasmids are vectors for redundant chromosomal genes in the *Bacillus cereus* group. **BMC Genomics**, v.16, n.1, jan. 2015.

3 ARTIGO:

**ANÁLISE *IN SILICO* DE GENES *rap-phr* EM
CROMOSSOMOS E PLASMÍDEOS DE LINHAGENS DE
*Bacillus cereus lato sensu***

3.1 RESUMO

Espécies do gênero *Bacillus* produzem proteínas Rap (regulador de resposta aspartato fosfatase) que regulam vias essenciais para a bactéria, como esporulação, competência e formação do biofilme. Rap é inibida pelo peptídeo Phr, que é secretado, processado e internalizado para cumprir sua função, atuando assim como sensor de *quorum*. Essas proteínas têm sido investigadas em *B. subtilis* e com poucos estudos em *Bacillus cereus lato sensu*. Um grupo bacteriano ao qual pertencem *B. cereus stricto sensu*, *B. thuringiensis* e *B. anthracis*, bactérias fenotipicamente diferentes, mas geneticamente muito similares. Atualmente não há consenso se estas bactérias devem ser consideradas espécies distintas ou variantes especializadas de uma única espécie, inclusive porque suas principais características fenotípicas são codificadas em plasmídeos. O objetivo deste trabalho foi identificar e caracterizar os sistemas *rap-phr* nos genomas destas espécies. Foram identificados 189 genes *rap* e 147 genes *phr* a partir do genoma de 34 linhagens do grupo com sequências genômicas completas. Dos genes obtidos, 23% são plasmidiais, uma frequência alta considerando um gene essencial para o metabolismo da célula. As proteínas Rap e Phr foram comparadas com proteínas ortólogas de *B. subtilis* e entre si. Foi possível notar uma grande variabilidade das sequências e em sua distribuição e na sintenia entre os genes de *B. anthracis*, enquanto que os genes de *B. cereus* e *B. thuringiensis* estão mais distribuídos e intercalados. Não foi possível identificar marcadores quanto a espécie nem em relação a sua localização no genoma. Entretanto, a grande quantidade dos genes plasmidiais nas linhagens de *B. thuringiensis*, em conjunto com a maior plasticidade dos genomas dessa espécie, pode permitir responder melhor às modificações ambientais.

Palavras-chave: esporulação, plasmídeo, *B. cereus stricto sensu*, *B. thuringiensis*, *B. anthracis*

3.2 INTRODUÇÃO

Muitos processos bacterianos são controlados por um sistema de comunicação entre as células denominado *quorum-sensing*. Esta comunicação se dá por meio da secreção e reconhecimento de moléculas por bactérias cognatas, coordenando a expressão gênica de cada célula de acordo com a densidade populacional (PERCHAT et al., 2011; PEREGO et al., 1996). Um exemplo é o sistema Rap-Phr, no qual o regulador de resposta Rap é inibido pelo peptídeo Phr, que necessita ser exportado e processado no meio extracelular para assumir sua conformação ativa. Ao ser reimportado para o ambiente intracelular, o peptídeo Phr é capaz de inibir a proteína Rap correlata. Contudo, esse transporte é dependente da concentração desse peptídeo no meio, que por sua vez está condicionada à densidade populacional (AUCHTUNG; LEE; GROSSMAN, 2006; PEREGO, 1996, 2013).

As proteínas Rap fazem parte de uma família de proteínas denominada RNPP, cujo nome é denominado da fusão dos nomes das proteínas Rap, NprR, PlcR e PrgX. Essas proteínas estão presentes nas espécies do gênero *Bacillus* e apresentam em comum a modulação por pequenos peptídeos, conforme a densidade populacional, como sistemas de *quorum sensing*. Os membros da família RNPP compartilham, também, a mesma organização estrutural contendo repetições tetratricopeptídicas (TPR) (PERCHAT et al., 2011; SLAMTI et al., 2014).

TPR é motivo estrutural que está normalmente envolvido em funções mediadas por interações proteína-proteína e consiste em repetições em tandem de 34 resíduos de aminoácidos. No entanto, sua sequência consenso é definida por um padrão preferencial de aminoácidos com resíduos pequenos ou grandes e não pela presença de um aminoácido invariável em determinada posição, permitindo, assim, uma grande versatilidade nas sequências proteicas (D'ANDREA; REGAN, 2003).

Essas proteínas têm sido amplamente estudadas em *Bacillus subtilis*, espécie na qual já foram descritas 11 proteínas Rap. Pesquisas com as proteínas Rap demonstram que RapA, RapB, RapE, RapH, RapI e RapJ atuam na via de esporulação; RapC, RapF e RapH regulam a indução de competência para transformação. RapG inibe o regulador de resposta DegU que controla vários processos celulares, inclusive a produção de proteases extracelulares. Enquanto

que RapD e RapK ainda não tiveram suas funções reveladas. Oito dos 11 genes *rap*, possuem à jusante o gene *phr* correlato (AUCHTUNG; LEE; GROSSMAN, 2006; BONGIORNI et al., 2005; PEREGO, 2013).

Alguns trabalhos relatam a presença de genes *rap-phr* plasmidiais análogos aos genes cromossomais conhecidos. Análises funcionais desses genes revelam envolvimento com a regulação da produção de enzimas proteolíticas (KOETJE et al., 2003) ou com a esporulação, competência para transformação gênica e arquitetura do biofilme (PARASHAR et al., 2013). Ou ainda, regulando a conjugação (SINGH et al., 2013)

Os genes *rap-phr* também já foram identificados em algumas linhagens de *B. thuringiensis* (SLAMTI et al., 2014) e *B. anthracis* (BONGIORNI et al., 2006), contudo ainda não foram caracterizados molecular e funcionalmente em *Bacillus cereus lato sensu*. Este agrupamento bacteriano é composto pelas espécies *Bacillus cereus sensu stricto*, *Bacillus thuringiensis*, *Bacillus anthracis*, que são espécies estreitamente relacionadas geneticamente mas apresentam patótipos diferentes.

As principais características fenotípicas que distinguem *B. thuringiensis* e *B. anthracis* de *B. cereus stricto sensu*, relacionadas ao patótipo ou nicho ecológico, são codificadas por genes presentes em grandes plasmídeos conjugativos, os genes *cry* e os genes para formação de toxina e cápsula, respectivamente (TOURASSE et al., 2011).

Desta forma, essas características tornam o grupo um modelo biológico interessante para estudos que permitam um melhor entendimento do papel dos plasmídeos para as bactérias. Questiona-se se os plasmídeos estão preferencialmente presentes em uma das espécies do grupo ou se numa população de *B. cereus lato sensu* os plasmídeos podem ser compartilhados livremente entre células das três espécies (BAHL; HANSEN; SØRENSEN, 2009; RAYMOND et al., 2010). Em uma análise de populações simpátricas de *B. cereus* e *B. thuringiensis*, Vilas Boas et al (2002) observou a taxa de fluxo gênico é maior entre os isolados da mesma espécie, do que entre as espécies. Contudo, a troca de material genético entre as duas espécies é possível.

Santos et al (2010), além de confirmar a ocorrência de transferência genética por conjugação entre as duas espécies, notou a necessidade da presença de células vegetativas para que essa transferência ocorra.

Assim, este trabalho tem como objetivo identificar e caracterizar molecularmente, por meio de ferramentas de bioinformática, os genes *rap-phr* presentes nas linhagens de *B. cereus lato sensu*. Visando compreender melhor a distribuição, a variabilidade numérica e de sequência desses genes nas linhagens dessas espécies, focando principalmente na comparação entre genes comossômicos e plasmidiais, com intuito de contribuir para um melhor entendimento tanto das questões taxonômicas envolvendo essas espécies, quanto do papel desses genes e do papel dos plasmídeos nestas espécies.

3.3 METODOLOGIA

3.3.1 Linhagens bacterianas

Foram realizadas consultas na seção Genoma do NCBI (<http://www.ncbi.nlm.nih.gov/genome/>), visando identificar linhagens do grupo *B. cereus*, pertencentes às espécies *B. cereus*, *B. thuringiensis* e *B. anthracis* com sequência completa do genoma disponível até abril de 2014.

Foram registradas as principais características disponíveis sobre cada uma das linhagens: origem, tamanho e conteúdo GC dos genomas e cromossomos, número de genes e proteínas preditos e o número de plasmídeos.

Para possibilitar uma visão da relação filogenética entre as linhagens foi utilizado o ‘*Bacillus cereus* group Multilocus and MultiData Typing website’ da Universidade de Oslo (<http://mlstoslo.uio.no/>). As sequências provenientes deste banco de dados foram alinhadas no programa Mega 6.1 (TAKAMURA et al., 2013) pelo algoritmo ClustalW e uma árvore filogenética foi construída utilizando o algoritmo Neighbor-Joining.

3.3.2 Levantamento dos genes

Todas as sequências genômicas identificadas como pertencentes a essas três espécies foram selecionadas para a busca de genes *rap* e *phr*, utilizando o *GenBank* (BENSON et al., 2015) como base.

3.3.2.1 Genes *rap*

A busca pelas sequências foi realizada em duas etapas, inicialmente empregando-se palavras-chaves: '*rap*' e '*response regulator aspartate phosphatase*'. Em seguida, os genes identificados foram utilizados para buscas por similaridade de sequência por meio do algoritmo BLASTn. As sequências nucleotídicas e proteicas de cada gene *rap* identificado foram selecionadas para a realização das etapas posteriores desse estudo. Adicionalmente, o primeiro critério acima descrito foi empregado para a busca de genes *rap* em todas as linhagens de *B. subtilis* com a sequência completa do genoma disponível.

3.3.2.1 Genes *phr*

A procura pelos genes *phr* se deu a partir de sua localização em relação ao gene *rap*. Foram considerados genes *phr* aqueles que se encontravam ligados e/ou sobrepostos aos genes *rap* identificados pelos critérios adotados nesse estudo. Também foi realizada a busca empregando-se as palavras-chaves: '*phr*', '*response regulator aspartate phosphatase inhibitor*' e '*rap inhibitor*'. As sequências nucleotídicas e proteicas de cada gene *phr* putativo foram selecionadas para a realização das etapas posteriores desse estudo. Adicionalmente, as sequências nucleotídicas e proteicas dos genes *phr* da linhagem de *B. subtilis* subsp. *subtilis* 168 foram salvas.

3.3.3 Caracterização dos genes/proteínas

3.3.3.1 Genes *rap*

A detecção da repetição tetratricopeptídica, correspondente ao domínio TPR, foi utilizada como um dos critérios para validar as sequências obtidas como pertencentes à família de proteínas Rap. Para tanto, as sequências nucleotídicas obtidas a partir da busca inicial foram analisadas em dois bancos de dados para domínios e famílias proteicas. O CDD (Conserved Domain Database) (MARCHLER-BAUER et al., 2015), no NCBI, um recurso para anotação de unidades funcionais em proteínas (<http://www.ncbi.nlm.nih.gov/Structure/cdd/wrpsb.cgi>) e o Pfam (<http://pfam.xfam.org/search>) (FINN et al., 2014), uma grande coleção de famílias

proteicas. Também foram analisadas as sequências de aminoácidos utilizando o banco de dados de proteínas InterPro Scan (<http://www.ebi.ac.uk/interpro/sequence-search>) (JONES et al., 2014), que fornece uma análise funcional das proteínas ao classificá-las em famílias e predizer domínios e sítios importantes.

3.3.3.2 Genes *phr*

A validação das sequências classificadas como *phr*, foi realizada utilizando o programa SignalP 4.1 Server (<http://www.cbs.dtu.dk/services/SignalP/>), que prediz a presença e localização dos sítios de clivagem de peptídeo sinal em sequências de aminoácidos (PETERSEN et al., 2011), como realizado por Pottathil e Lazazzera (2003).

3.3.4 Localização dos genes *rap-phr*

O panorama da localização dos operons *rap-phr* de cada linhagem foi gerado por meio da plataforma Artemis (RUTHERFORD et al., 2000), que permite a visualização das posições das sequências em relação ao cromossomo ou plasmídeo como um todo. Nos plasmídeos de *B. thuringiensis* que apresentam genes *rap-phr* também foram observadas a presença e localização dos genes *cry*.

Em acréscimo, foram registrados os genes dispostos mais proximamente à montante e à jusante dos genes *rap-phr*, na relação completa do GenBank de cada linhagem. No caso de genes anotados como proteína hipotética foi realizada a comparação por BLASTp para tentar inferir a família proteica dessas sequências.

3.3.5 Alinhamento cromossômico

Após identificada a localização dos genes *rap-phr* nos genomas, sequências genômicas foram alinhadas duas a duas pelo programa Genomes Match da plataforma Softberry (<http://www.softberry.com/cgi-bin/programs/scanh/gmatch.pl>), nos casos em que se identificou a possibilidade de modificações quanto a localização dos genes em uma dada linhagem, em relação à localização do mesmo gene nas demais linhagens, devido a possível ocorrência de inversão no genoma daquela linhagem.

3.3.6 Análises de similaridade entre as sequências

3.3.6.1 Similaridade com proteínas de *B. subtilis*

Outro critério para validar as sequências obtidas como pertencentes à família de proteínas Rap e dos peptídeos Phr foi a análise de identidade entre sequências dos genes *rap-phr* de *B. cereus*, *B. thuringiensis* e *B. anthracis* e sequências de proteínas de *B. subtilis* 168. Desta forma, a sequência dos genes *rap* identificadas foram comparadas com as sequências proteicas depositadas no banco de dados curado Swiss-Prot/UniProt (<http://www.uniprot.org/blast/>) (THE UNIPROT CONSORTIUM, 2015). Equivalentemente as sequências nucleotídicas dos genes *phr* preditos foram comparadas com a sequência peptídica dos Phr de *B. subtilis*. Essa comparação das sequências dos Phr também foi realizada por meio de BLASTp.

3.3.6.2 Análises de similaridade entre proteínas Rap ou Phr - De uma mesma linhagem

A similaridade entre todas as proteínas identificadas como Rap em uma mesma linhagem foi investigada por meio do BLASTp para as linhagens *B. subtilis* 168, *B. cereus* ATCC 14789 e *B. thuringiensis kurstaki* HD-73. Do mesmo modo foi realizada análise de similaridade entre os peptídeos Phr dessas mesmas linhagens.

3.3.6.3 Análises de similaridade entre proteínas Rap – Entre linhagens diferentes

Com o objetivo de verificar semelhanças entre as sequências anotadas como pertencentes a uma mesma classe de proteínas Rap, bem como verificar a existência de inconsistências nas anotações, sequências de proteínas anotadas como RapA e Rap D nos genomas das linhagens de *B. cereus* e de *B. thuringiensis*, além da proteína RapD de *B. subtilis* 168, foram submetidas a alinhamentos de sequências pela ferramenta Polydot, pertencente ao pacote de programas de análise Emboss (<http://emboss.bioinformatics.nl/cgi-bin/emboss/polydot>) (RICE; LONGDEN; BLEASBY, 2000).

3.3.6.4 Análises de similaridade entre proteínas Rap – Agrupamento

Para realizar o agrupamento em classes de todas as proteínas Rap identificada nas linhagens de *B. cereus lato sensu*, foi verificada a similaridade entre elas utilizando o algoritmo BLASTp. A linha de corte adotada para a formação dos grupos foi o valor de *e-value* igual a zero, representando identidade maior ou igual a 68%. Nos grupos em que foram verificadas identidade menor do que 75%, os constituintes foram rearranjados em subgrupos.

3.3.7 Análises filogenéticas

Análises filogenéticas entre as sequências das proteínas Rap, bem como entre as sequências das proteínas Phr foram realizadas através de alinhamento com o uso do algoritmo *ClustalW*. Para tanto foi construída uma árvore filogenética pelo método *Neighbor-Joining* com o uso do programa MEGA 6.1 (TAKAMURA et al., 2013). Valores de *bootstrap* foram adicionados para demonstração da sustentação dos diversos agrupamentos da árvore.

3.4 RESULTADOS E DISCUSSÃO

3.4.1 Linhagens bacterianas

O grupo *B. cereus* é o grupo bacteriano com maior número de linhagens com sequenciamento do genoma completo ou em andamento. Até abril de 2014, 34 linhagens de *B. cereus*, *B. thuringiensis* e *B. anthracis* tiveram a sequência do seu genoma completo (cromossomo e plasmídeos) disponibilizadas no NCBI. As sequências completas do genoma das 34 linhagens do grupo do *B. cereus*, incluindo 13 linhagens de *B. cereus sensu stricto*, 12 linhagens de *B. thuringiensis* e nove linhagens de *B. anthracis*, disponíveis na seção ‘Genoma’ do NCBI foram utilizadas como base para a realização desse estudo. Na Tabela 1 (Apêndice, pág. 77), podem ser visualizados os dados dos genomas dessas linhagens bacterianas, bem como o número de acesso de cada uma das sequências genômicas.

Por serem filogeneticamente muito próximas (HELGASON et al., 2000; VILAS-BÔAS; PERUCA; ARANTES, 2007), linhagens das três espécies apresentam algumas características muito similares, como por exemplo, a porcentagem do conteúdo de GC, de cerca de 35% para todas as linhagens. O tamanho do

cromossomo é outra característica conservada, variando de 5,21 a 5,23 Mb entre as linhagens de *B. anthracis*, de 5,08 a 5,42 Mb em *B. cereus*, e de 5,24 a 6 Mb para *B. thuringiensis*, o que demonstra que, principalmente nessa última espécie, ocorreram alterações que permitiram o aumento do tamanho do cromossomo (Tabela 1).

Essa variação no tamanho do cromossomo pode ser resultado de diversos eventos genéticos, alguns deles consequentes da ação de elementos genéticos móveis, como sequências de inserção (IS), transposons, íntrons grupo I e II e IStrons (íntrons associados a ISs) que contribuem para a dinâmica do genoma e que também podem ter impacto fenotípico (GILLIS; MAHILLON, 2014; KOLSTØ; TOURASSE; ØKSTAD, 2009). Elementos móveis, como transposons e IS são frequentemente encontrados em plasmídeos, inclusive em plasmídeos conjugativos, podendo assim ser transferidos para outras bactérias juntamente com esses. Além disso, ao se deslocar, um transposon pode também carregar uma região adjacente do genoma, e a frequente ocorrência de IS facilita a ocorrência de recombinações entre regiões do genoma, incluindo cromossomo e plasmídeos. Essas características resultam em duas consequências: i) aumento do número de linhagens que contém elementos genéticos móveis e ii) ocorrência de modificações na localização de genes, o que contribui para aumento na plasticidade do genoma da espécie (GILLIS; MAHILLON, 2014; MAHILLON et al., 1994).

Quanto às variações genômicas, estudos demonstram que linhagens de *B. thuringiensis* apresentam um grande número de plasmídeos (VILAS-BÔAS et al., 2004, ZHENG et al., 2015), o que pode ser verificado também nas linhagens utilizadas nesse trabalho, onde o conteúdo plasmidial variou de um plasmídeo ao máximo de 14 plasmídeos diferentes numa mesma linhagem (Tabela 1).

Assim, a presença de plasmídeos é uma característica importante que confere à bactéria, um aumento no genoma, que chegou até a 23%, no caso das linhagens *B. thuringiensis* MC28 e *B. thuringiensis* serovar *thuringiensis* IS5056. Mas a quantidade de DNA plasmidial pode ser superior à quantidade de DNA cromossômico em uma única célula de *B. thuringiensis* se o número de cópias de cada plasmídeo for considerado (ZHONG et al., 2011).

Em *B. cereus*, entre as linhagens empregadas nesse estudo, o número de plasmídeos diferentes variou de zero até cinco, o que resultou em um aumento máximo de 10% no tamanho do genoma, como no caso da linhagem *B. cereus* E33L, enquanto que em *B. anthracis*, a maioria das linhagens apresentou a

presença de dois plasmídeos diferentes, pXO1 e pXO2, que carregam os genes envolvidos com a virulência e patogenicidade dessa bactéria. No entanto, algumas linhagens de *B. anthracis* apresentaram a perda de um ou de ambos os plasmídeos, no momento em que foram sequenciadas. Desta forma, analisando os dados da Tabela 01, nessa espécie, a presença dos plasmídeos permitiu um acréscimo de material genético de no máximo 5%. No entanto, vale lembrar que a presença dos genes codificantes para toxinas e cápsula, localizados nos plasmídeos pXO1 e pXO2, respectivamente, é uma condição obrigatória para a classificação de uma bactéria como *B. anthracis*.

A observação do número total de genes identificados por anotação nos genomas das linhagens das três espécies permitiu verificar uma maior variação no número de genes entre as linhagens de *B. thuringiensis*, as quais apresentaram um mínimo de 4945 genes até o máximo de 7060, uma diferença de 2115 genes. Essa diferença foi menor entre as linhagens de *B. anthracis* estudadas, onde o genoma com maior número de genes apresentou um total de 6099 genes, enquanto que o menor número de genes foi 5265 genes, gerando uma diferença de 834 genes. No entanto, a menor variação foi encontrada entre as linhagens de *B. cereus*, com a diferença de 520 genes entre o genoma com maior número de genes identificados (6014 genes) e o genoma com menor número de genes identificados (5494 genes). É ainda importante destacar que, considerando as linhagens analisadas dessas três espécies, tanto o maior número de genes (7060), quanto o menor número (4945), foram identificados em linhagens de *B. thuringiensis*, o que demonstra uma maior variabilidade quantitativa no genoma dessa espécie, em relação às outras duas espécies.

Essas diferenças entre os números de genes identificados podem ser explicadas por diversas razões, uma delas devido às diferentes ferramentas de bioinformática empregadas para a anotação dos genomas. Entre as 34 linhagens estudadas, houveram linhagens que tiveram o genoma sequenciado e anotado no ano 2002 (*B. anthracis* Ames e *B. cereus* ATCC 10987) enquanto outras tiveram seu genoma anotado somente no ano 2014 (*B. anthracis* SVA11). Dessa forma, considerando-se que o sequenciamento e a anotação dos genomas foram feitos por diversos grupos distintos e as metodologias de anotação evoluíram juntamente com o desenvolvimento das tecnologias de sequenciamento, pode haver uma disparidade entre a anotação do genoma das diferentes linhagens. Acredita-se que

as sequências genômicas disponibilizadas mais recentemente e com as anotações revisadas manualmente estejam mais completamente e corretamente anotadas, já que técnicas de anotação automatizada utilizam as sequências e anotações disponíveis de linhagens estreitamente relacionadas para treinamento dos modelos de predição.

No entanto, mesmo considerando as diferenças entre as ferramentas de bioinformática empregadas para a anotação dos genomas, as diferenças entre a quantidade de genes identificados entre as três espécies, podem refletir que linhagens de *B. thuringiensis* apresentam uma maior plasticidade genômica quantitativa que as linhagens das outras espécies, como pode ser visto tanto pelo critério que trata da quantidade de plasmídeos diferentes quanto pela quantidade de genes identificados no cromossomo das mesmas.

Quando os dados da Tabela 1 são analisados visando a identificação do número de genes codificantes para proteínas nos genomas das três espécies avaliadas, os resultados encontrados permitem verificar a ocorrência da mesma situação descrita acima para a quantidade de genes. Desta forma, foi possível a identificação de uma menor variação no número de genes codificantes para proteínas entre as linhagens de *B. cereus* (626 genes), seguido pelas linhagens de *B. anthracis* (1018 genes), e com a maior variação tendo sido observada entre as linhagens de *B. thuringiensis* (2124 genes). A diferença entre o número de genes e o número de proteínas preditos para cada linhagem, se dá pela presença de pseudogenes, sequências repetitivas, elementos móveis, regiões não traduzidas (RNAs).

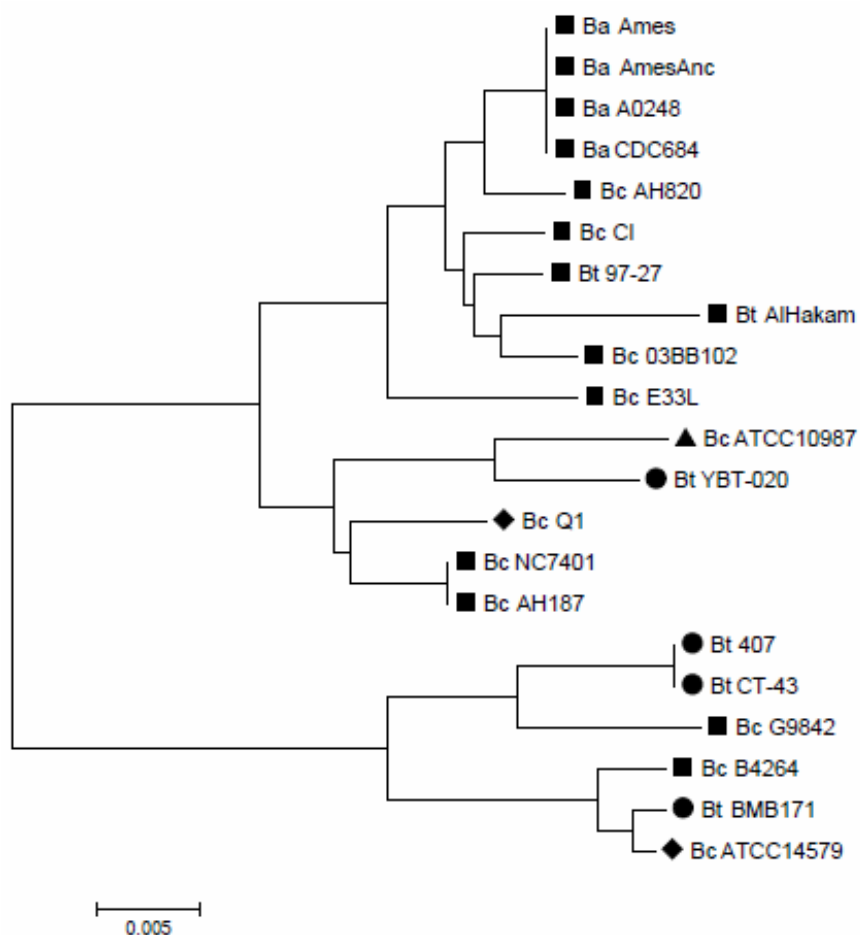
Para analisar as relações filogenéticas entre as linhagens selecionadas, foram utilizadas as sequências de MLST já disponíveis no 'Bacillus cereus group Multilocus and MultiData Typing website', que reúne os cinco esquemas de MLST já publicados para o grupo *B. cereus*. Todas as sequências selecionadas nos cinco esquemas são regiões conservadas de genes essenciais para o metabolismo da bactéria (housekeeping). Como a última atualização do banco de dados ocorreu em 2012, as sequências das linhagens cujo resultado do sequenciamento foi liberado mais recentemente não estavam disponíveis. Das 24 linhagens presentes no site, apenas 21 foram utilizadas para a construção da árvore, pois as outras três (*B. thuringiensis* HD73, *B. anthracis* Sterne e *B. cereus* F837/76) não foram avaliadas

em todos os cinco esquemas de MLST e a ausência de alguns genes poderia alterar a estrutura da árvore e diminuir a confiabilidade dos dados.

A árvore gerada (Figura 1) a partir destas sequências revelou a formação de dois grandes grupos. No primeiro é possível visualizar dois subgrupos, em um deles há o agrupamento das linhagens de *B. anthracis* (todas reunidas no mesmo ramo) e as linhagens de *B. cereus* e *B. thuringiensis* que apresentam elevada patogenicidade aos mamíferos. No outro subgrupo há uma mistura de linhagens de *B. cereus* patogênica aos mamíferos, uma linhagem obtida a partir de amostra de solo, uma de campo de petróleo e outra a partir de produtos lácteos e uma linhagem de *B. thuringiensis* inseticida.

No grupo mais distante das linhagens de *B. anthracis*, estão as linhagens inseticidas de *B. thuringiensis* e linhagens de *B. cereus* isoladas do solo, além de algumas patogênicas aos mamíferos. No entanto, a maioria (10 de 13) das linhagens de *B. cereus* com sequência completa do genoma disponível, é composta por linhagens patogênicas aos mamíferos, dificultando assim a compreensão do posicionamento das linhagens não-patogênicas na árvore e não refletindo toda a extensa variabilidade genética apresentada principalmente por linhagens de *B. cereus* e *B. thuringiensis*.

Figura 1 – Árvore filogenética de seqüências dos cinco esquemas de MLST das espécies de *B. cereus lato sensu* (Bc – *B. cereus*, Bt – *B. thuringiensis* e Ba – *B. anthracis*). Para a construção da árvore foi utilizado o algoritmo *Neighbor-Joining*.



Origem ou atividade das linhagens: ■- patogênicos aos mamíferos; ◆- isolado de solo ou campo de petróleo; ●- atividade inseticida; ▲- produtos lácteos.

Os agrupamentos filogenéticos observados na Figura 1 não permitem a diferenciação das linhagens entre as três espécies estudadas, mas os ramos representam a afinidade dos grupos quanto ao potencial patogênico e/ou ao ambiente de origem delas. Esse resultado está de acordo com análises moleculares mais amplas realizadas por meio de MEE, AFLP ou MLST que demonstraram a ocorrência de linhagens de *B. cereus* e *B. thuringiensis* patogênicas aos mamíferos mais intimamente relacionadas às linhagens de *B. anthracis* do que às outras linhagens de *B. cereus* e *B. thuringiensis* obtidas a partir de amostras de solo. Adicionalmente, quando consideradas apenas as linhagens de *B. cereus* e *B. thuringiensis*, pode-se observar que os membros das duas espécies se encontram

intercalados por todo o agrupamento filogenético (HELGASON et al., 2000; TICKNOR et al., 2001; TOURASSE, 2006 e 2011).

A partir das análises filogenéticas é possível constatar que, apesar do alto nível de similaridade e sintonia gênica entre as linhagens, o grupo como um todo apresenta um elevado grau de heterogeneidade genética e adaptações genômicas relacionadas ao nicho ecológico no qual estas linhagens estão incluídas. Além disso, o conteúdo plasmidial nem sempre coincide com sua filogenia (TOURASSE et al., 2006). Entretanto, a determinação da espécie da qual um isolado ou linhagem faz parte é dependente de características codificadas nos plasmídeos, enquanto os estudos filogenéticos são baseados em marcadores cromossômicos.

É possível notar que as linhagens com sequências do genoma completo disponíveis não representam a amplitude da variabilidade genética que o grupo tão diverso representa, pois, a escolha das linhagens é dada por interesses econômicos e/ou médicos e não ao acaso. Assim, o agrupamento dessas linhagens ou as características apresentadas pelo grupo de sequências disponíveis não necessariamente refletem a realidade das espécies que elas representam.

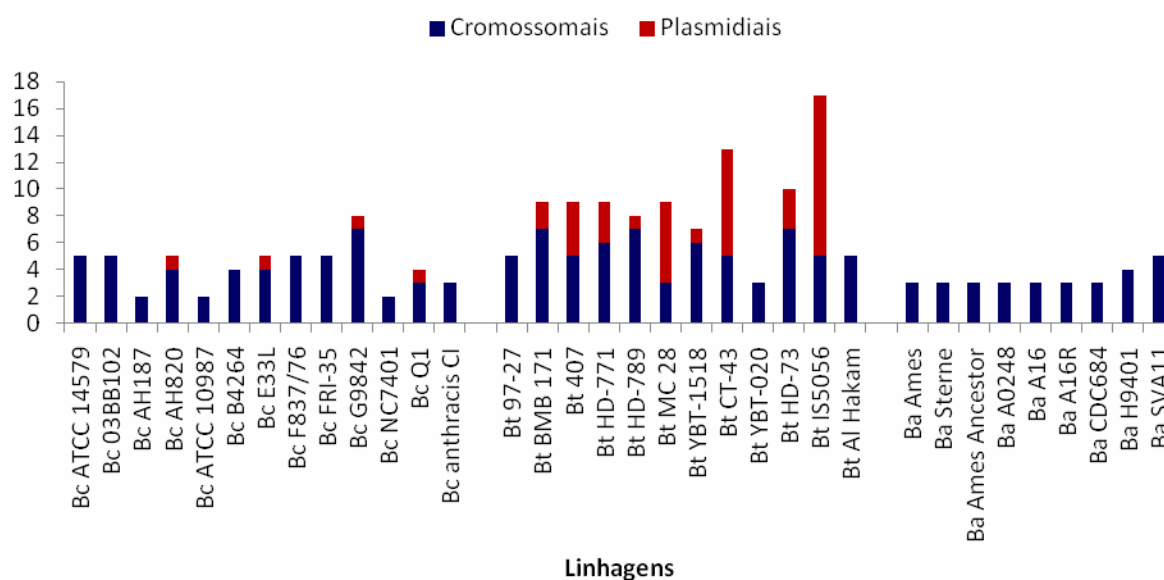
3.4.2 Levantamento e localização dos genes *rap-phr*

3.4.2.1 Levantamento dos genes *rap-phr*

Entre os genes codificantes para proteínas, aqueles que codificam as proteínas Rap e os peptídeos Phr foram o foco desse estudo. A pesquisa pelos genes *rap* permitiu a identificação de um total de 189 genes nos genomas das 34 linhagens estudadas (Tabela 2 – Apêndice, pág 80). A identificação desses genes foi iniciada, a princípio, através da busca pelo nome do gene, usando “*rap*” como palavra-chave, o que permitiu a identificação de 26 genes, todos categorizados dentro de classes, conforme descrito para *B. subtilis* (*rapA*, *rapB*, etc). Em seguida, empregou-se o nome completo do gene para a identificação de novos genes nos genomas das linhagens analisadas, ou seja, “*response regulator aspartate phosphatase*”, o que permitiu a identificação de outros 91 genes. Por último, as sequências dos 117 genes foram empregadas para busca por similaridade, o que possibilitou a identificação de outros 72 genes que não se encontravam anotados como *rap*.

Do total de genes *rap* identificados, 55% dos genes foram localizados no genoma de linhagens de *B. thuringiensis*, enquanto que 29% dos genes foram localizados nas linhagens de *B. cereus* e outros 16% em linhagens de *B. anthracis*. Em todas as linhagens pesquisadas foram identificados no mínimo dois genes *rap*, como é o caso das linhagens *B. cereus* AH187, ATCC10987 e NC7401. A linhagem *B. thuringiensis* IS5056 foi a que apresentou o número máximo de genes identificados, num total de 17 genes, seguida pela linhagem *B. thuringiensis* CT-43 com 13 genes *rap* identificados (Figura 2).

Figura 2 – Número de genes *rap* por linhagem



Bt: *B. thuringiensis*, Bc: *B. cereus*, Ba: *B. anthracis*.

Quanto aos genes *phr* (Tabela 2), foram reconhecidos 144 genes à jusante dos 189 genes *rap*, ou seja, prevalência de 76,2%. Similar à prevalência observada em *B. subtilis* 168, considerada padrão da espécie, que foi de 72,7% (8 dos 11), como já relatado na literatura (PEREGO, 2013). Os prováveis genes *phr* de *B. cereus lato sensu* estão sobrepostos aos genes *rap* por um ou três nucleotídeos, em sua maioria. Enquanto que os genes *phr* de *B. subtilis* 168 apresentam um a 43 nucleotídeos sobrepostos ao gene *rap*. Tanto em *B. cereus lato sensu* quanto em *B. subtilis* 168, os genes *phr* estão no mesmo sentido da fita de DNA que os genes *rap*, mas são transcritos em quadros de leitura diferentes.

Quando a busca de genes *phr* foi realizada por meio de palavra-chave, foram encontrados outros três genes anotados como '*phr*' que não estavam ligados à um

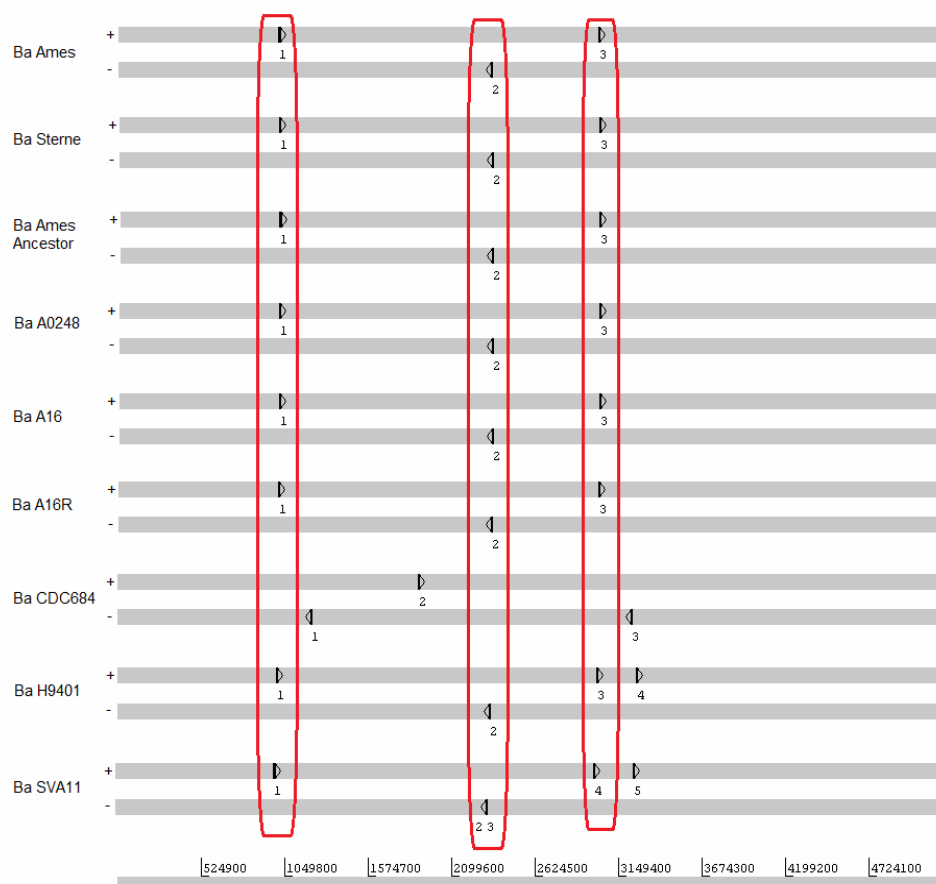
gene *rap*. Desses, os genes anotados como “response regulator aspartate phosphatase inhibitor” nas linhagens *B. thuringiensis* YBT-1518 e *B. cereus* NC7401, estão localizados próximos a genes que não apresentam nenhuma relação com os genes *rap* e não estão sobrepostos a esses genes, nem à montante, nem à jusante. Mas apresentaram similaridade a alguns dos genes *phr* identificados anteriormente quando tiveram suas sequências comparadas por BLASTn.

O gene anotado como “Phr-like protein” de *B. thuringiensis* HD-73 apresentou sobreposição ao gene à montante, anotado como “Helix-turn-helix domain protein”, cujo produto quando comparado ao banco de dados do NCBI pelo BLASTp resultou em proteínas similares a NprR (fator de transcrição necrotrófico). Além disso a comparação por BLASTn não revelou similaridade com outros *phr* de *B. cereus lato sensu*. Deste modo, pode-se supor que este gene na verdade se trata do gene codificante para o peptídeo NprX, que regula a proteína NprR, também pertencente à família RNPP. Tanto Phr quanto NprX são oligopeptídeos em sua forma ativa e regulam as proteínas Rap e NprR, respectivamente. Seus genes estão localizados à jusante dos genes Rap e NprR, com os quais formam uma unidade transcricional, apesar de em alguns casos possuírem promotor próprio. Entretanto, o produto do gene identificado apresenta 84 aminoácidos de comprimento, enquanto que a literatura descreve NprX como um peptídeo de 43 aminoácidos (SLAMTI et al., 2014).

3.4.2.2 Localização dos genes *rap-phr*

Para observar a posição que os genes *rap-phr* ocupam nos cromossomos ou plasmídeos que os contém, foi gerado um panorama de localização (Figuras 3, 5 a 8). Na Figura 3, pode-se evidenciar a conservação da localização dos genes *rap* no genoma das linhagens de *B. anthracis* estudadas. O resultado obtido aponta para o elevado grau de clonalidade que ocorre entre as linhagens de *B. anthracis* já destacado por outros autores (HELGASON et al., 2000; KOLSTØ; TOURASSE; ØKSTAD, 2009). A maioria das linhagens apresentou três sistemas *rap-phr* cuja localização cromossômica foi conservada. Apenas a linhagem *B. anthracis* CDC684, apresentou alteração na posição desses genes, no entanto, essa alteração na posição foi resultante da ocorrência de um evento de inversão da região cromossômica contendo os mesmos.

Figura 3 - Representação gráfica da localização dos genes *rap* no cromossomo das linhagens de *B. anthracis* (Ba)*.



Numeração do *rap* é correspondente à identificada na Tabela 2, * em destaque estão as áreas com localização conservada dos genes.

A ocorrência da inversão no genoma da linhagem CDC684 foi comprovada pela construção de gráficos de matrizes de pontos gerados pelo 'Genomes Match – Softberry' (Figura 4). Os pontos do gráfico representam a similaridade de blocos gênicos entre as linhagens comparadas. No alinhamento entre as linhagens de *B. anthracis* Ames e SVA11, a linha reta diagonal corresponde a manutenção da localização e posicionamento dos blocos gênicos. A variação e o diferente ângulo de parte da reta nos alinhamentos entre CDC 684 X Ames e CDC 684 X SVA11 representam uma modificação na posição e ordenamento destes blocos gênicos.

Esse evento genético também já foi descrito por Okinaka et al (2011). A inversão compreende uma região de 3,3 Mb e parece ter sido causada por um evento de recombinação interna entre as regiões homólogas dentro de dois profagos lisogênicos semelhantes a Lambda, que são encontrados em todos os genomas de *B. anthracis*. A linhagem CDC 684 foi originalmente classificada como *Bacillus*

megaterium e apesar de conter todas as características determinantes de *B. anthracis*, apresenta toxicidade severamente atenuada e cinética de crescimento alterada.

Para as linhagens de *B. cereus* e *B. thuringiensis* pôde-se observar um maior número de genes *rap-phr* que em linhagens de *B. anthracis*, além da ocorrência de uma maior variação na quantidade de genes localizados nos cromossomos (Figuras 5 e 6). Em ambas as espécies, dividindo-se o cromossomo em três partes de cerca de 2.000.000 de pb, a partir da origem de replicação, pode-se perceber que a maioria dos genes *rap-phr* identificados, está localizado principalmente nos dois primeiros terços do cromossomo, ou seja, até o nucleotídeo 4.000.000, sendo que somente três linhagens de *B. cereus* e três de *B. thuringiensis* apresentaram genes *rap-phr* na região que compreende o último terço do cromossomo.

Outro aspecto que pode ser observado nas Figura 5 e 6, para as linhagens de *B. cereus* e *B. thuringiensis*, é a grande concentração de genes posicionados na região de aproximadamente 1,05 Mb nos cromossomos de nove das 13 linhagens de *B. cereus*, e que também pode ser observada em seis das 12 linhagens de *B. thuringiensis*. Esse mesmo evento de concentração de grande quantidade de genes ocorre também em uma outra região cromossômica, entre 3 e 3,7Mb, em 12 linhagens de *B. cereus* e em 10 de *B. thuringiensis*. Estas duas concentrações estão em posições próximas às localizações conservadas de *B. anthracis*. A aglomeração de genes *rap-phr* em *B. anthracis* na posição de aproximadamente 2,3Mb também é possível observar em *B. cereus*, o qual apresenta 6 linhagens com genes nesta localização.

Figura 4 – Representação gráfica dos alinhamentos entre blocos gênicos de linhagens de *B. anthracis* (Ba) gerado pelo Genomes Match

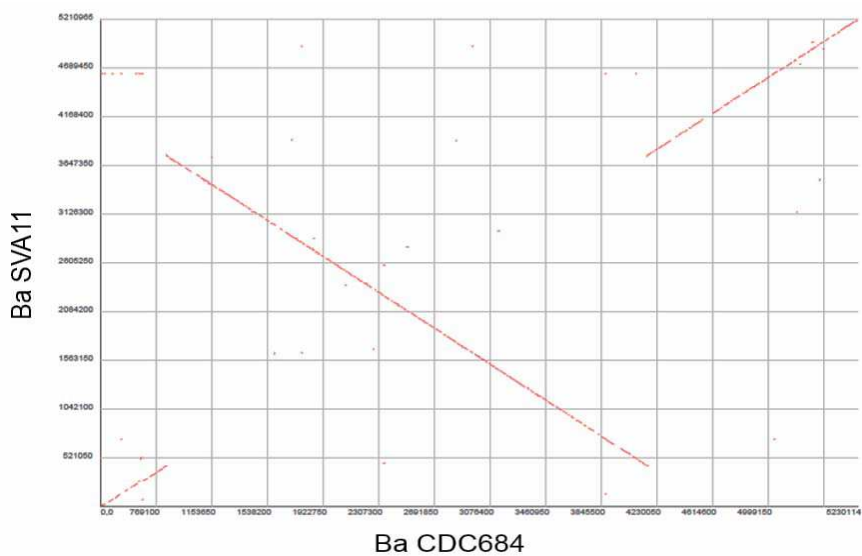
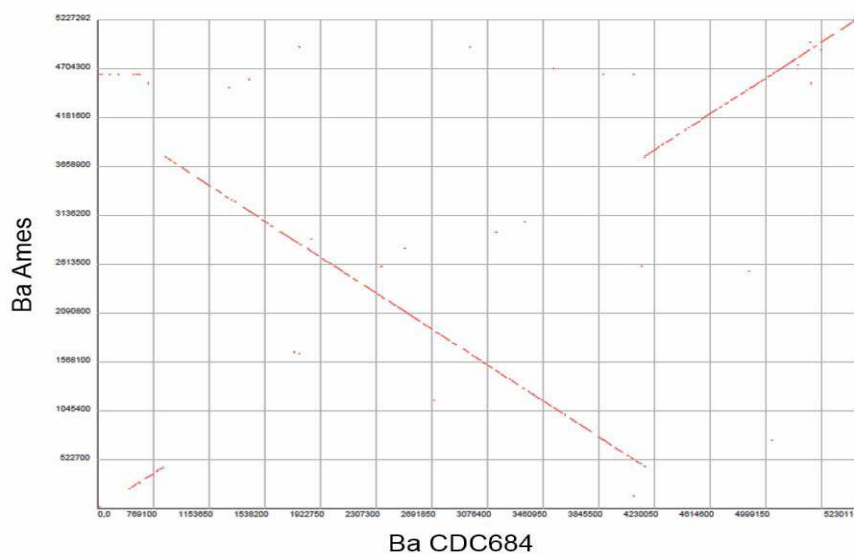
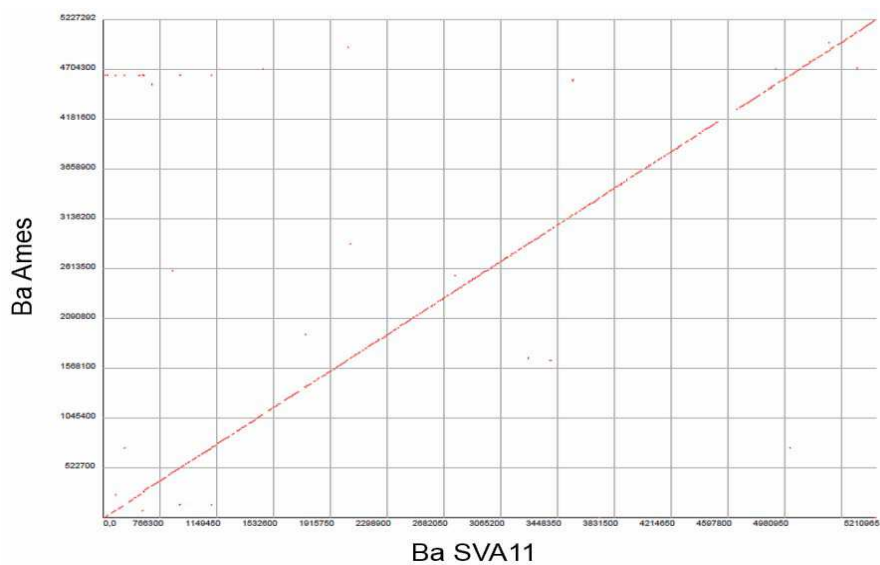
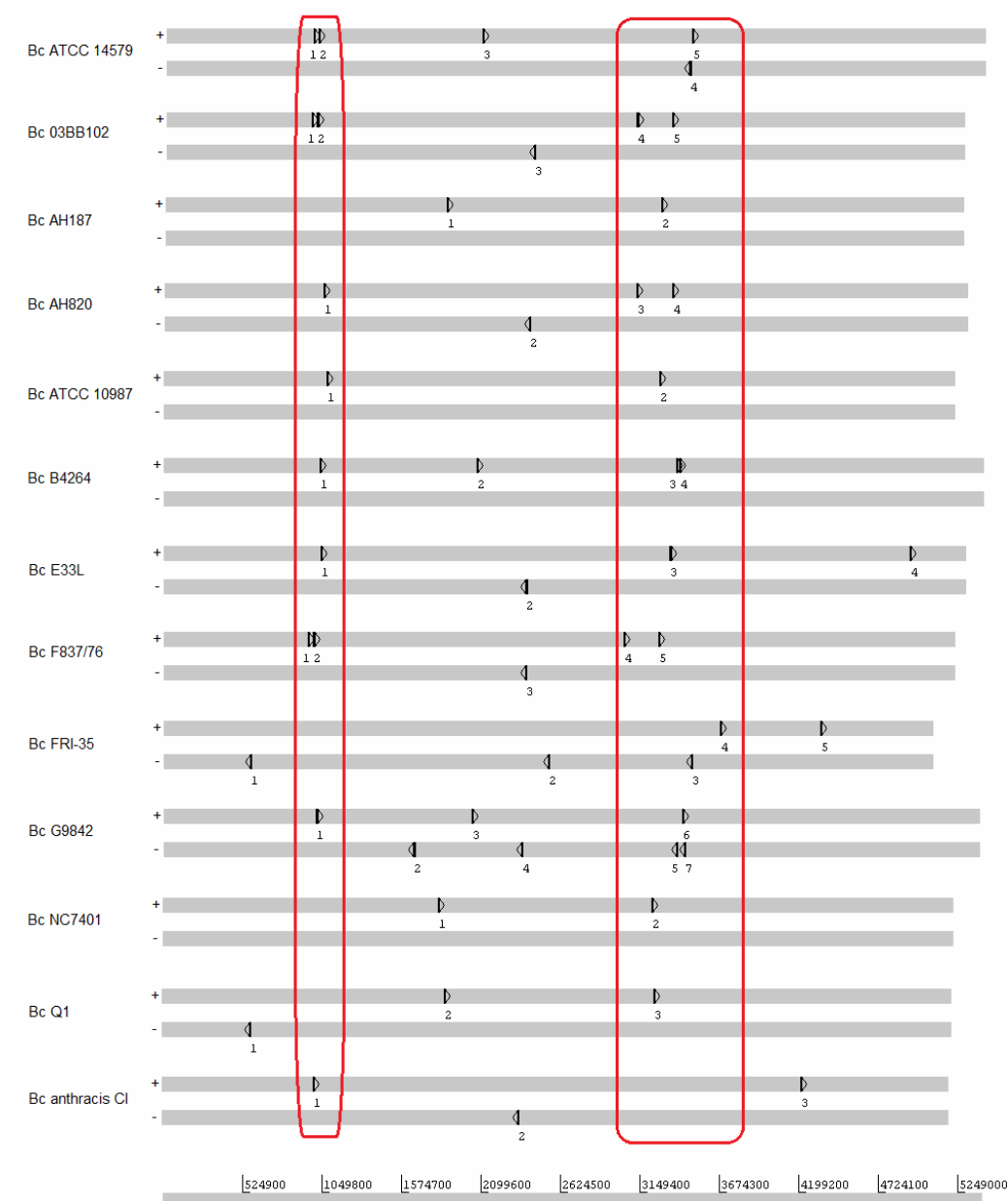


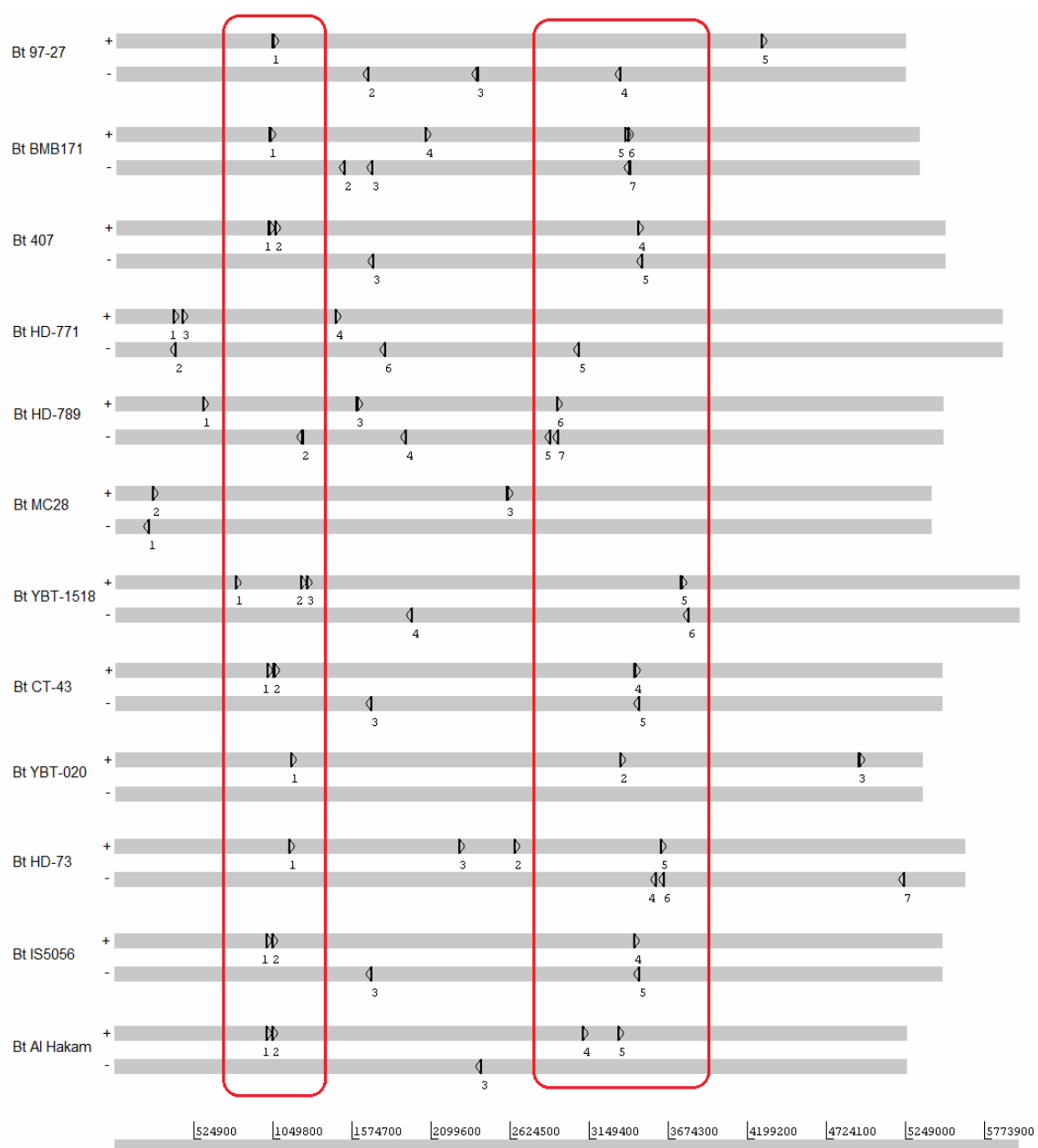
Figura 5 – Representação gráfica da localização dos genes *rap* no cromossomo das linhagens de *B. cereus* (Bc)*.



Numeração do *rap* é correspondente à identificada na Tabela 2, * em destaque estão as áreas com localização conservada dos genes

Também é possível observar que a maioria dos sistemas identificados está localizada nos arredores do terminal da replicação (~2,7 Mb), região conhecida por sua plasticidade genética (RASKO et al., 2005). O que poderia justificar a variação de número e localização dos genes *rap-phr* nesta área.

Figura 6 – Representação gráfica da localização dos genes *rap* no cromossomo das linhagens de *B. thuringiensis* (Bt)



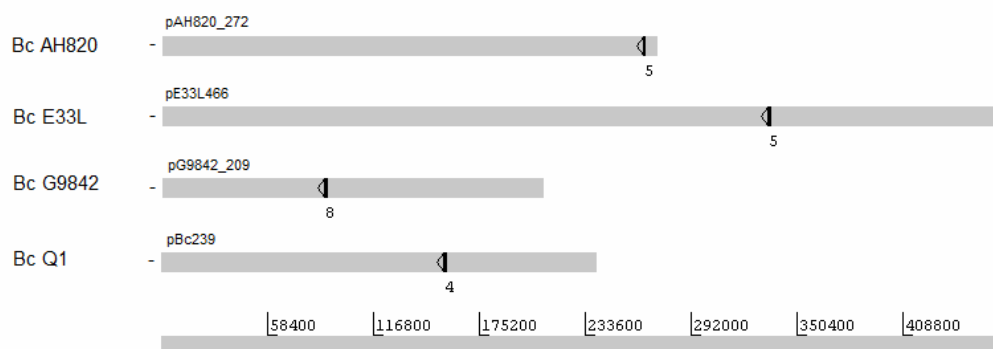
3.4.2.3 Genes plasmidiais

Foram identificados 44 genes *rap* localizados nos plasmídeos das espécies de *B. cereus lato sensu* (Tabela 3 – Apêndice, pág 86), dos quais a maioria estava presente nas linhagens de *B. thuringiensis* (91%). Não foram encontrados genes *rap* plasmidiais em linhagens de *B. anthracis*. Das 34 linhagens selecionadas para este

estudo, em 38% delas (ou seja, 13 linhagens) foram identificados genes localizados em plasmídeos, além dos genes cromossômicos.

Os genes *rap* foram encontrados em plasmídeos com uma grande variedade de tamanho, de 6.880 pb a 502 kb, com maior ocorrência em grandes plasmídeos, maiores que 70 kb. Adicionalmente, em linhagens de *B. thuringiensis*, em plasmídeos maiores que 280 kb foram encontrados até três genes *rap* diferentes, como é o caso do plasmídeo pMC429, de 430 kb, na linhagem *B. thuringiensis* MC 28.

Figura 7 – Representação gráfica da localização dos genes *rap* nos plasmídeos das linhagens de *B. cereus* (Bc). Numeração do *rap* é correspondente à identificada na Tabela 2.



Das linhagens de *B. subtilis* que apresentam a sequência completa do genoma disponível em bancos de dados, a maioria não apresenta conjuntamente a sequência de nenhum de seus plasmídeos. Foi identificada somente a sequência do plasmídeo pBEST195S da linhagem *B. subtilis* subsp. *natto* BEST195, no qual não foram encontrados genes *rap* anotados. No entanto, desde a década de 1990, alguns autores já relataram a presença de genes *rap-phr* nos plasmídeos dessa espécie, sistemas análogos aos genes cromossômicos conhecidos (MEIJER et al., 1995 e 1998).

A partir da identificação desses cassetes nos plasmídeos, alguns estudos se voltaram para a análise funcional desses genes, como Koetje et al. (2003) que detectaram a presença do sistema Rap60-Phr60 no plasmídeo pTA1060 de *B. subtilis* 168, em que a proteína Rap60 mostrou-se efetiva na repressão da produção de enzimas proteolíticas. Adicionalmente, Parashar et al. (2013) publicaram a descoberta do cassete *rapP-phrP* localizado em um plasmídeo de 80 kb que está

presente em isolados selvagens de *B. subtilis* mas não em linhagens consideradas como padrão dentro da espécie. Esse autor descreveu a atuação da proteína Rap na cadeia de transferência de fosfato, regulando a esporulação, competência para transformação gênica e a arquitetura do biofilme. Porém, diferentemente dos outros sistemas Rap-Phr relatados, o produto do gene *phrP* não afetou significativamente a função da proteína RapP.

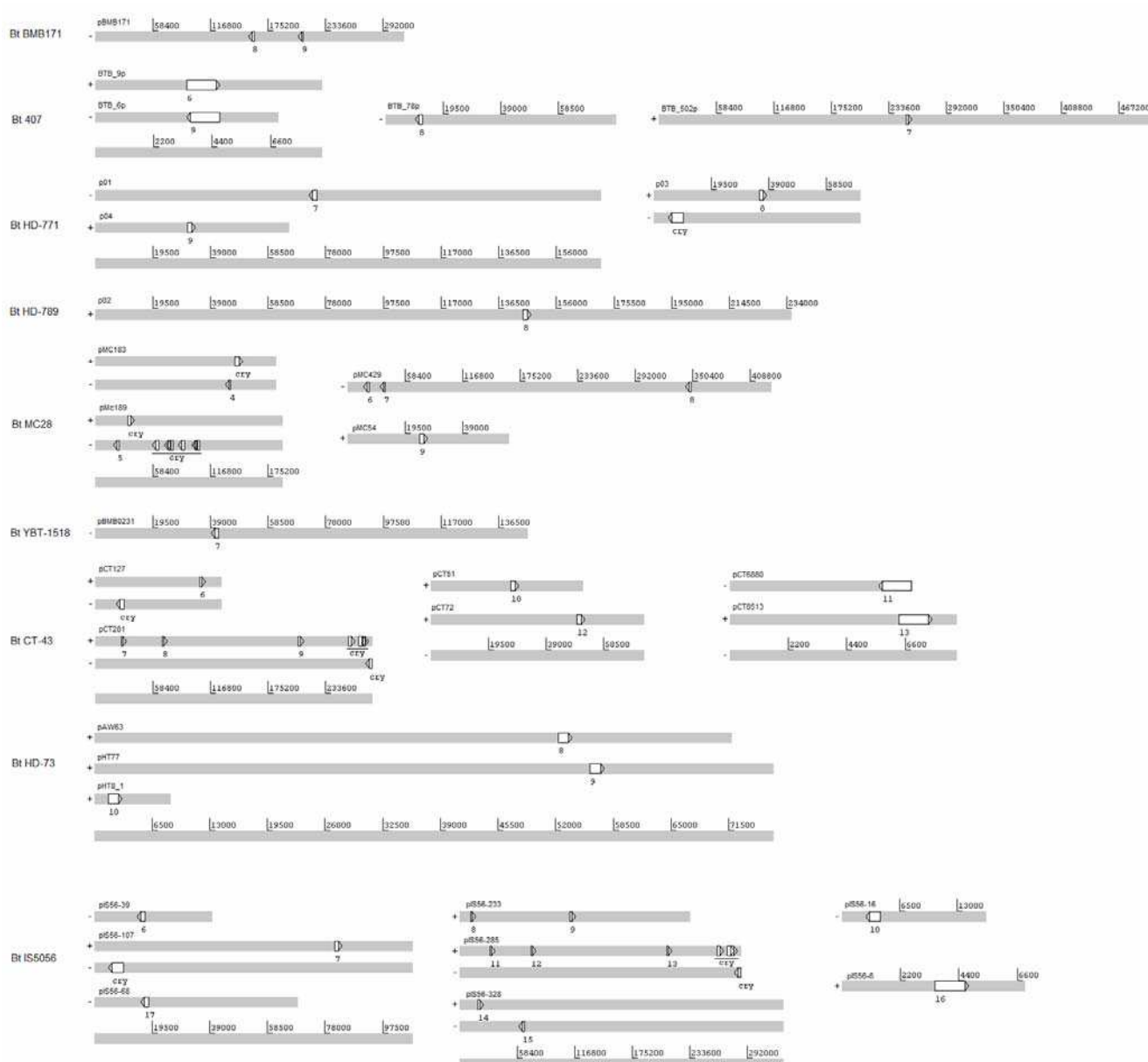
Além disso, os resultados de Singh et al. (2013) revelaram que o sistema *rap-phr* do plasmídeo conjugativo pLS20 regula sua mobilidade. No qual a proteína Rap_{LS20} atua como anti-repressor da conjugação, enquanto é regulada negativamente pelo peptídeo Phr_{LS20}, sinalizando que este peptídeo determina a temporização da conjugação. Quando todas ou a maioria das células de determinada população produzirem o peptídeo Phr_{LS20}, sua concentração no meio será elevada e a conjugação inibida. No meio em que somente algumas células produzirem o peptídeo, sua concentração não será suficiente para inibir a proteína Rap e, conseqüentemente, a conjugação do plasmídeo. Assim, os genes de conjugação só serão ativados quando a célula doadora estiver próxima a células que não possuem este plasmídeo, ou seja, células potencialmente receptoras.

Ainda que não tenham sido encontrados genes *rap* em plasmídeos de *B. anthracis* pelos métodos adotados neste trabalho, Bongiorno et al. (2006) relataram a presença de um sistema Rap-Phr no plasmídeo de virulência pXO1 de *B. anthracis* A2012, linhagem com sequenciamento genômico ainda parcialmente disponível e, por isso, não selecionada para a realização desse estudo. Os genes encontrados por esses autores apresentam 100% de identidade com seqüências presentes nos plasmídeos pXO1 de outras linhagens de *B. anthracis* cujas seqüências genômicas estão disponíveis. Em acréscimo, a análise da seqüência de aminoácidos apresentou 90% de identidade com uma proteína codificada no plasmídeo pBc10987 de *B. cereus* ATCC10987, que também não foi identificada pelos métodos aqui utilizados. Segundo esses autores, o sistema encontrado afeta a iniciação da esporulação, sugerindo que o sistema Rap-Phr plasmidial pode proporcionar uma vantagem seletiva para a virulência de *B. anthracis*.

Segundo Brunsing et al. (2005) a ocorrência da esporulação em *B. anthracis* pode estar em oposição à patogênese eficiente. Dessa forma, as proteínas Rap podem ser responsáveis por uma regulação mais eficaz da via de esporulação, adiando seu início dentro do hospedeiro, pois macrófagos são capazes de captar e

destruir esporos, mas não células vegetativas encapsuladas, contudo mantendo a eficiência de esporular em ambientes menos favoráveis à bactéria, assim regulando de forma conveniente as vias de patogenicidade. Observação que presumivelmente pode ser estendida para *B. cereus* e *B. thuringiensis*.

Figura 8 – Representação gráfica da localização dos genes *rap* nos plasmídeos das linhagens de *B. thuringiensis*; os genes *cry*, quando presentes também estão identificados. Numeração do *rap* é correspondente à identificada na Tabela 2.



Adicionalmente, foi analisada a presença concomitante de genes *rap-phr* e de genes *cry* em plasmídeos de *B. thuringiensis*, não tendo sido encontrada qualquer

regra. De tal modo que foram identificados plasmídeos portadores de ambos os genes (como por exemplo: pMC183), portadores de somente genes *cry* (pHT73) e portadores de somente genes *rap*, com (BTB_502p) ou sem *phr* (pCT6880). Assim, não foi possível inferir nenhuma relação específica significativa, quanto a presença ou localização mútuas, entre os genes *rap-phr* plasmidiais e os genes *cry* anotados nos plasmídeos das linhagens de *B. thuringiensis*, gene codificante da principal característica fenotípica diferencial da espécie (Figura 8).

Segundo Koetje et al. (2003), a presença de sistemas Rap-Phr plasmidiais, em *B. subtilis*, pode proporcionar às bactérias que os comportam mecanismos adicionais para regular a expressão de genes que são suscetíveis ao *quorum sensing*. Assim, a expressiva quantidade de genes *rap-phr* plasmidiais em *B. thuringiensis* identificadas neste estudo pode conferir vantagens comportamentais às linhagens desta espécie, quando comparadas às linhagens de *B. cereus* e *B. anthracis*. Experimentos realizados em lagartas de *Bombyx mori* sugerem que linhagens de *B. thuringiensis* tem maior habilidade de germinação de esporos e multiplicação no inseto que linhagens de *B. cereus* (SANTOS et al., 2010), e a presença dessa extensa rede de plasmídeos e de genes *rap* adicionais, pode ser uma das razões que justificam essa melhor capacidade adaptativa da espécie.

Por se tratar de um gene essencial ao metabolismo da bactéria, atuando em vias fundamentais como esporulação, competência e formação de biofilme, a presença desses genes num dado plasmídeo pode tornar essencial a presença desse plasmídeo nas linhagens bacterianas. Um dado relevante do ponto de vista do papel do plasmídeo dentro do ecossistema.

Assim, pode-se presumir que a ocorrência de genes *rap-phr* plasmidiais em cerca de 40% das linhagens estudadas pode dar suporte à ideia de um equilíbrio entre os três elementos de um dado ecossistema: bactéria, plasmídeo e ambiente. Para as bactérias, a presença de plasmídeos com genes essenciais que interferem em funções vitais, as permite responder melhor às modificações ambientais, o que pode torná-las melhor representadas no ecossistema dinâmico em que estão inseridas, uma vez que estas apresentam vantagens adaptativas em relação às outras bactérias que não contêm esses elementos. Considerando-se a existência de genes *rap* ou *rap-phr* em plasmídeos conjugativos, e a ocorrência de trocas genéticas entre diferentes bactérias de um mesmo ambiente, a presença desses genes em plasmídeos pode favorecer a estabilidade dos mesmos em diferentes

bactérias, o que aumenta a ocorrência do mesmo num dado ambiente. Desta forma, plasmídeos contendo genes *rap* serão mais amplamente mantidos neste ecossistema devido a transferência genética por conjugação entre bactérias de uma mesma ou de diferentes espécies, geneticamente compatíveis, e estavelmente mantidos nessas bactérias desde que carreguem genes essenciais ao metabolismo das mesmas.

3.4.2.3 Genes localizados à montante e à jusante ao cassete *rap-phr*

Foi realizado também o levantamento dos genes localizados mais proximamente (à montante e à jusante) ao sistema *rap-phr* na ordem da relação dos genes do GenBank. Esses genes, seu tamanho e localização estão dispostos na Tabela 2.

Como a maioria dos genes destacados estavam anotados como proteína hipotética, foi realizado uma busca por similaridade por BLASTp e os resultados quando disponíveis estão descritos entre parênteses na Tabela 2 (após 'hypothetical protein'). Ainda assim, mais de 80 desses genes permaneceram identificados apenas como proteína hipotética.

Os genes que estavam anotados ou tiveram sua família identificada por similaridade fazem parte, em sua maioria, de famílias gênicas relacionadas ao metabolismo de macromoléculas (como glioxalase; lipase / alpha/beta hidrolase; 3-oxoacil-ACP sintase; CDP-glucose 4,6 hidratase; 5-nucleotidase, entre outros) ou reguladores da transcrição (como de esporulação e síntese de protease PAI2; de ligação ao açúcar LacI; entre outros). Também é possível observar genes relacionados a elementos genéticos móveis, como transposase, proteína Hfq (essencial para replicação de alguns bacteriófagos) e proteínas de elementos móveis. Além de pseudogenes próximos a genes *rap-phr* em linhagens de *B. anthracis*.

Em geral, os genes vizinhos são conservados em genes que apresentam maior similaridade. O gene mais próximo ao gene *rap* está a 45 pb de distância, verificado na linhagem *B. cereus anthracis* CI, enquanto que 2865 pb separam o gene mais distante do gene *rap*, observado na linhagem *B. cereus* G9842.

3.4.3 Caracterização dos genes *rap-phr*

3.4.3.1 Comparação com genes/proteínas de *B. subtilis*

Em *B. subtilis*, espécie na qual os genes *rap* têm sido amplamente estudados, há 11 classes de genes *rap* (PEREGO, 2013). Num levantamento dos genes *rap* nas linhagens dessa espécie, com genoma completamente sequenciado disponíveis no NCBI, foi possível notar que, em geral, cada uma das linhagens apresenta uma cópia de cada umas das classes de genes *rap* (Tabela 4 e 5). Em acréscimo, os dados da Tabela 5 (Apêndice, pág 89) permitem verificar que o gene *rapB* foi anotado quatro vezes em *B. subtilis* TU-B-10 e três vezes em *B. subtilis* RO-NN-1; e o gene *rapH* apareceu duas vezes em *B. subtilis* BSP1.

Linhagem	<i>rap</i> Anotados em classes		'reponse regulator aspartate phosphatase'		Total
	Normais	Curtos	Normais	Curtos	
Bs subtilis 168	11	-	-	-	11
Bs spizizenii W23	8	-	-	1	9
Bs BSn5	-	-	10	-	10
Bs spizizenii TU- B-10	11	5	-	-	16
Bs subtilis RO- NN-1	9	2	-	-	11
Bs QB928	9	-	-	-	9
Bs subtilis BSP1	9	-	-	-	9
Bs BEST7613	9	-	-	-	9
Bs XF-1	6	3	-	-	9
Bs subtilis 6051- HGW	11	-	-	-	11
Bs subtilis BAB- 1	2	2	7	-	11
Bs BEST7003	10	-	-	-	10
Bs PY79	9	-	-	-	9
Bs subtilis JH642 s. AG174	11	-	-	-	11
Bs subtilis AG1839	11	-	-	-	11
Bs subtilis OH131.1	-	-	-	-	-

Tabela 4 – Distribuição dos genes *rap* nas linhagens de *B. subtilis* (Bs)

Anderson et al. (2005) em uma análise comparativa entre os genomas do grupo do *B. cereus* e de *B. subtilis*, relataram semelhanças entre os genomas dessas bactérias, exceto para as categorias de processamento de informação, transdução de sinal, virulência e subsistemas de transporte. Em relação ao sistema

de transdução de sinal, os genomas de *B. subtilis* apresentam 11 genes codificantes para fosfatases da família Rap, enquanto que nos genomas dos membros do grupo do *B. cereus*, há uma maior variação na quantidade desses genes, 8 em *B. thuringiensis* sub spp. *israelensis*, 6 em *B. anthracis* A 2012 e 5 em *B. cereus* ATCC 14579, por exemplo.

Foi realizada uma análise comparativa entre as proteínas Rap entre *B. cereus lato sensu* e *B. subtilis* por meio do banco de dados Swissprot (BLASTx), a partir do qual foi observada uma identidade entre 20 e 35% entre os genes *rap* do grupo de linhagens analisado com as proteínas Rap de *B. subtilis* 168.

Como o banco de dados Swissprot não reconheceu similaridade entre os peptídeos Phr de *B. cereus lato sensu* e *B. subtilis* 168, a análise comparativa foi realizada por meio de alinhamentos no BLASTp (NCBI) entre os 147 peptídeos Phr aqui identificados com os oito peptídeos da linhagem de *B. subtilis* 168 (Phr A, C, E, F, G, H, I e K). Quarenta peptídeos Phr de *B. cereus lato sensu* não apresentaram similaridade alguma com os Phr de *B. subtilis*. Entretanto, apenas dois alinhamentos resultaram em similaridade significativa (com e-value menor que e^{-3}), o Phr7 de *B. thuringiensis* Bt407 com PhrE (59% de cobertura e 29% de identidade) e o Phr1 de *B. cereus* B4264 com PhrA (29% de cobertura e 28% de identidade).

A cadeia de transferência de fosfato, responsável pela iniciação e regulação das vias de esporulação e competência, tem sido amplamente estudada em *B. subtilis*, enquanto que os estudos nas espécies de *B. cereus lato sensu* se restringe a poucos trabalhos em *B. anthracis* (BRUSING et al., 2005; BONGIORNI et al., 2006). Justificando assim, a investigação pela similaridade dos genes *rap-phr* do grupo com à espécie tipo.

O regulador de resposta final Spo0A, fator de transcrição da via, apresenta sequência de aminoácidos altamente conservada em todas espécies do gênero *Bacillus* e *Clostridium*. Da mesma maneira, a comparação de outros componentes proteicos da via entre *B. subtilis* e *B. anthracis* revelou uma alta homologia entre os ortólogos Spo0F e Spo0B. Por outro lado, os domínios sensores das histidinas-quinases são fracamente conservados entre as espécies, dificultando a identificação dos ortólogos (BRUSING et al., 2005). Da mesma forma, os ortólogos de *rap-phr* de *B. subtilis* e *B. cereus lato sensu* parecem não apresentar uma identidade forte entre eles. Diferença que pode ser justificada também pela necessidade do sistema possuir uma sequência espécie/linhagem-específica para o

phr atuar efetivamente como sensor de quórum sobre sua proteína *rap* correlata específica.

3.4.3.2 Tamanho dos genes *rap* e *phr*

O tamanho médio dos genes *rap* encontrados nos genomas das três espécies estudadas foi de 1033 pb, variando de 108 a 1233 pb. Quando os genes curtos (menores que 1000 pb, mais detalhados no tópico abaixo) foram desconsiderados, a média do tamanho apresentado pelos genes *rap* foi elevada para 1102 pb.

Em *B. subtilis*, o tamanho dos genes *rap* varia entre 1065 e 1176 pb, sendo constante em cada tipo, por exemplo, os genes *rapA* de todas as linhagens apresentam 1137 pb de comprimento e *rapC* 1149 pb. Além do mesmo tamanho, foi possível observar que genes codificantes para o mesmo tipo de *rap* são localizados no mesmo 'frame' em todas as linhagens (Tabela 5).

No grupo do *B. cereus* não existe uma classificação dos genes *rap* baseada em tamanho da sequência gênica. Houve uma tentativa de agrupar os genes *rap* aqui estudados conforme o tamanho, baseado no exemplo de *B. subtilis*, mas não foi possível, já que 33% dos genes encontrados apresentam um mesmo tamanho de 1095pb, aparecendo 62 vezes, e grande variação no tamanho dos genes restantes. Há outros tamanhos que apareceram repetidas vezes, mas em frequências bem menores. Não foi possível observar um tipo característico que estivesse presente em todas as linhagens analisadas.

Assim, entre as três espécies estudadas, duas apresentam variações quanto à quantidade e localização dos genes *rap*. Dessa forma, análises de similaridade foram iniciadas visando verificar o grau de variabilidade de sequências desses genes e de suas proteínas tanto dentro do grupo, quanto em relação às sequências de *B. subtilis*.

O tamanho dos genes identificados como *phr* variou entre 123 a 330 pb com média de 170,47 pb, mais variáveis e maiores dos que os genes *phr* de *B. subtilis* 168, cujos tamanhos variam de 120 a 174 pb, com média de 130, 87 pb.

3.4.3.3 Genes *rap* curtos

Um total de 18 genes (9,52% do total de genes identificados) apresentou comprimento abaixo de 1000 pb (Tabela 6). Esses genes curtos podem ter sido gerados devido a erros no sequenciamento ou à ocorrência de mutações que resultaram em deslocamentos do quadro de leitura que levaram ao aparecimento de términos abruptos, que, conseqüentemente, tornaram não funcionais as proteínas codificadas por eles. Como demonstrado anteriormente, esses genes curtos também podem ser observados nos genomas de *B. subtilis* (Tabela 4).

Tabela 6 – Características principais dos genes *rap* curtos de *B. cereus lato sensu*

Linhagem	Localização	Gene situado mais proximamente	Rap*	Tam.	Presença de Phr	Gene situado mais proximamente
<i>Bc</i> AH820	pAH820_272	Hypothetical protein (transposase)	rapB 5	177	+	{tetratricopeptide repeat protein-1035pb} / Transcriptional regulator, ArsR family
<i>Bc</i> E33L	pE33L466	Transposase	rapD 5	915	-	Hypothetical protein
<i>Bc</i> G9842	Cromossomo	Hypothetical protein	rap 7	231	-	Sensory box/GGDEF family protein
<i>Bt</i> BMB171	Cromossomo	Cellulase	rap 7	474	-	Sensory box/GGDEF family protein
	pBMB171	Hypothetical protein	rap 9	108	+	Transposase
<i>Bt</i> 407	Cromossomo	Cellulase	rap 5	207	-	Sensory box/GGDEF family protein
<i>Bt</i> HD-771	Cromossomo	Sensory box/GGDEF family protein	rap 1	219	-	Glycosidase, family 5
<i>Bt</i> HD-789	Cromossomo	Glycosidase, family 5	rap 7	246	-	Sensory box/GGDEF family protein
	Cromossomo	Putative endoglucanase	rap 6	219	-	Sensory box/GGDEF family protein
<i>Bt</i> YBT-1518	Cromossomo	Hypothetical protein	rap 7	894	-	{tetratricopeptide domain-containing protein-120pb}/ Hypothetical protein
	pBMB0231	Hypothetical protein	rap 7	894	-	{tetratricopeptide domain-containing protein-120pb}/ Hypothetical protein
<i>Bt chinensis</i> CT-43	Cromossomo	Cellulose	rap 5	219	-	Sensory box/GGDEF family protein
<i>Bt kurstaki</i> HD-73	Cromossomo	Hypothetical protein (SAM-dependent methyltransferase)	rap 2	300	-	ATP-binding protein IstB
	Cromossomo	Hypothetical protein (glycosyl hydrolase family 5)	rap 6	474	-	Sensory box/GGDEF
<i>Bt thuringiensis</i> IS 5056	Cromossomo	Putative endoglucanase	rap 5	207	-	Sensory box/GGDEF family protein
<i>Ba</i> H9401	Cromossomo	Hypothetical protein (histidine kinase)	rap 4	345	+	Transcriptional repressor of sporulation and protease synthase
<i>Ba</i> SVA 11	Cromossomo	2',3'-cyclic	rap 2	141	+	DNA-binding protein

	nucleotide 2'-phosphodiesterase	rap 3	987		
Cromossomo	Hypothetical protein (histidine kinase)	rap 5	345	+	Transcriptional repressor of sporulation and protease synthase

Numeração do *rap* é correspondente à identificada na Tabela 2; * genes grifados apresentam similaridade ao gene *rap1* de *B. cereus* ATCC14579

Alguns desses genes estão aparentemente divididos, como o caso dos genes *rap2* e *rap3* da linhagem *B. anthracis* SVA11 (Figura 3), na qual os dois genes curtos que estão anotados como “*response regulator aspartate phosphatase*”, apresentam uma sobreposição, e se fossem preditos como um só, apresentariam um tamanho condizente com um gene *rap* normal. Nos outros dois eventos observados, apenas uma das partes está anotada como *rap*, mas a outra ao ser comparada por BLASTp também apresenta similaridade com *rap*, como ocorre nos genes *rap5* de *B. cereus* AH820 e *rap7* *B. thuringiensis* YBT-1518.

Também foi identificada por similaridade sequências correspondentes aos genes curtos *rap4* de *B. anthracis* H9401 e *rap5* de *B. anthracis* SVA11, que apresentam localização cromossômica similar e mesmo tamanho, nas outras linhagens dessa espécie. Mas como nesta etapa do processo o tamanho médio dos genes já era conhecido, essas sequências foram descartadas.

Cinco genes *rap* curtos possuem o gene *phr* correlato. Quatro genes *rap* curtos estão localizados em plasmídeos, incluindo dois genes que fazem parte dos casos apresentados acima (genes separados em duas partes) e os outros dois estão associados a transposases.

Dos genes curtos restantes, nove apresentam similaridade com o gene *rap1* de *B. cereus* ATCC14579 (genes *rap* sublinhados na Tabela 6) e são cromossômicos. Os genes localizados ao lado desses genes curtos são conservados, de um lado celulose/glicosidase e de outro uma proteína da família GGDEF (molécula sinalizadora intracelular), divergindo assim do gene na linhagem ATCC14579, que é ladeado por proteínas hipotéticas. Estes genes curtos não estão estritamente relacionados a elementos geneticamente móveis, como plasmídeos ou transposons, e oito deles estão localizados na região entre 2,9 a 3,8 Mb dos respectivos cromossomos.

3.4.3.4 Caracterização das sequências

A anotação gênica nos genomas, por ser uma tarefa trabalhosa, somada ao avanço das tecnologias de sequenciamento, que tem elevado exponencialmente a quantidade de genomas sequenciados, tem sido realizada com o emprego de métodos automatizados. Entretanto, a acurácia dessa abordagem tem sido questionada, pois os modelos estão baseados, em sua maioria, em identidades de sequências relativamente fracas e são dependentes de treinamento. Genes com anotação funcional errada ou genes não preditos podem afetar diversos tipos de análise (POPTSOVA; GOGARTEN, 2010).

Por ser um processo de várias etapas, é difícil definir se os erros aconteceram durante o sequenciamento, na predição dos genes ou na atribuição de suas funções (POPTSOVA; GOGARTEN, 2010).

Os genes *rap* identificados por similaridade estavam anotados no GenBank como proteína hipotética (*hypothetical protein*), proteína contendo domínio tetratricopeptíco (*tetratricopeptide domain-containing protein*), putativo regulador de resposta (*response regulator, putative*), proteína com domínio TPR (*TPR domain-containing protein*), regulador de resposta (*response regulator*) e proteína com repetição TPR (*TPR repeat protein*).

Para verificar a anotação das sequências obtidas e confirmar sua classificação como regulador de resposta aspartato-fosfatase, todas as sequências foram comparadas com três bancos de dados para domínio e famílias proteicas diferentes: CDD (NCBI), Pfam e InterPro, visando a identificação do domínio TPR, característica obrigatoriamente constante na estrutura da proteína (PEREGO, 2013) e presente na família proteica Rap em *B. subtilis*, composta por 11 membros, todos com relativa similaridade de sequência e com organização estrutural em comum, caracterizada por repetições de tetratricopeptídeos (TPR). A detecção desse domínio nas proteínas analisadas das linhagens de *B. cereus lato sensu* auxilia na validação dessas proteínas como pertencentes a família de proteínas Rap.

A relação completa sobre as proteínas Rap e os dados obtidos podem ser observadas na Tabela 7 (Apêndice, pág 92). Todas as proteínas elencadas, com exceção das sequências curtas (<115 aminoácidos), apresentaram o domínio TPR identificados por pelo menos dois dos bancos de dados explorados. Apenas uma

proteína Rap de *B. thuringiensis kurstaki* HD-73, com 99 aminoácidos, apresentou o domínio TPR em apenas um banco de dados, o CDD.

Proteínas Rap menores que 300 aminoácidos foram desconsideradas das análises posteriores, pois a ausência de todo domínio TPR, ou parte dele, poderiam alterar os resultados de similaridade.

Da mesma forma, todas as sequências identificadas nesse estudo como gene *phr* foram investigadas quanto a presença de peptídeos sinais, critério esse adotado por Pottathil e Lazazzera (2003). O programa Signal P identificou a sequência sinal em 116 dos peptídeos (Tabela 2), ou seja, em apenas 78,9 % dos precursores de Phr de *B. cereus lato sensu*. Mas análises funcionais realizadas por Stepherson et al (2003) indicaram que o peptídeo sinal predito do sítio de clivagem não é necessário para o processamento do peptídeo Phr.

Também avaliamos os peptídeos Phr de *B. subtilis* 168 quanto a presença do peptídeo sinal, o qual não foi identificado em PhrE, PhrG e PhrI. Pottathil e Lazazzera (2003) previram que estes mesmos Phr não possuem alta correspondência com sítios de clivagem peptídeo sinal e que nem todos os pré-peptídeos parecem ser processados da mesma maneira. Mesmo os peptídeos nos quais não foram identificados peptídeos sinais, é possível observar no gráfico gerado pelo programa, que a porção amino-terminal é carregada positivamente, como descrito na literatura, característica estrutural típica de proteínas exportadas (PEREGO; BRANNIGAN, 2001; POTTHATIL; LAZAZZERA, 2003).

3.4.4 Análise de similaridade de proteínas Rap

Bongiorni et al.(2006), utilizaram a sequência de aminoácidos da proteína RapA de *B. subtilis* para pesquisar pelo BLAST a presença dessas fosfatases na sequência do genoma de *B. anthracis* A2012. Identificaram seis genes, cujas proteínas compartilham o mesmo peso molecular e 34% de resíduos idênticos com RapA. Dessas, apenas duas afetaram a esporulação *in vivo* e uma outra pareceu ter tido um deslocamento de quadro de leitura, gerando uma proteína truncada.

Por meio das análises por BLASTp entre as proteínas Rap da linhagem de *B. subtilis* 168 foi possível verificar a similaridade entre as proteínas Rap dos 11 tipos diferentes, com mínimo de 23% e máximo de 57% de identidade e cobertura variando de 64 a 100% (Quadro 1, Apêndice, pág 95). Perego (2013) retratou o alto

grau de homologia compartilhada entre as sequências das proteínas Rap de *B. subtilis*, que considerou ser de aproximadamente 45%, sugerindo que a conformação estrutural é semelhante, mesmo havendo uma diversidade de alvos com os quais as proteínas Rap interagem e afetam a função.

Da mesma forma, foi realizada a avaliação de similaridade entre as proteínas Rap de linhagens específicas de *B. cereus* (ATCC 14579) (Quadro 2, Apêndice, pág 96) e *B. thuringiensis* (HD-73) (Quadro 3, Apêndice, pág 97), as quais foram escolhidas por serem consideradas linhagens referência e muito estudadas nessas espécies. *B. cereus* ATCC 14579 apresenta cinco genes *rap*, todos cromossômicos e *B. thuringiensis* HD-73 apresenta 10 genes *rap*, sendo três genes localizados em plasmídeos diferentes. Foi possível constatar que a similaridade das proteínas Rap nessas espécies é ligeiramente mais elevada do que em *B. subtilis*, entre 39 e 53% de identidade na linhagem de *B. cereus*, e 32 e 58% de identidade na linhagem de *B. thuringiensis*. Quanto à cobertura, os valores são ainda mais significativos para *B. cereus* ATCC 14579 e *B. thuringiensis* HD-73, a quais apresentaram cobertura mínima de 98%, com relação à variação na cobertura entre 64 e 100% entre as proteínas de *B. subtilis* 168.

Do mesmo modo, a análise de similaridade entre os peptídeos Phr de cada uma das três linhagens foi realizada, resultados demonstrados nos Quadros 4 (*B. subtilis* 168) e 5 (*B. thuringiensis* HD-73) (Apêndice, págs 98 e 99). É possível observar nestes quadros que muitos dos alinhamentos resultaram em ausência de similaridade, motivo pelo qual também não foi construído um quadro para os peptídeos de *B. cereus* ATCC 14579. Apenas dois dos cinco genes *rap* desta espécie possuem o gene *phr* correlato (Phr3 e Phr5), e as sequências dos peptídeos codificados por eles não apresentam nenhuma similaridade significativa.

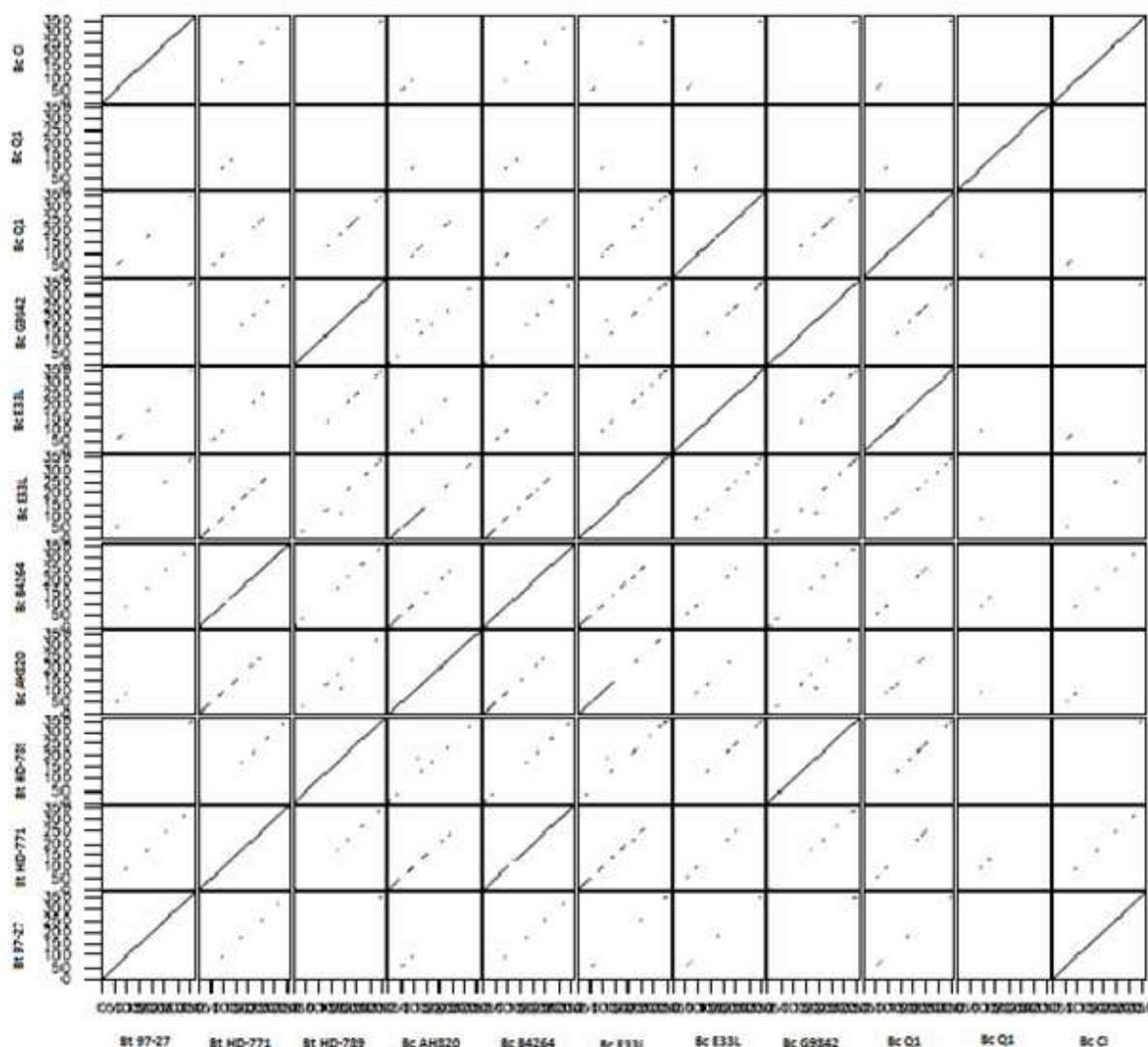
Esta ausência de similaridade entre os peptídeos identificados, como também observado na comparação entre as sequências de *B. subtilis* e *B. cereus lato sensu* por análise no SwissProt, não quer dizer que estes não possam ser identificados como Phr, pois a sequência estrita dos aminoácidos não é necessária para a funcionalidade do peptídeo, sendo consideradas as características bioquímicas dos resíduos desses aminoácidos, as quais determinam a sequência do peptídeo sinal, sua exportação, clivagem e sua forma oligopeptídica ativa.

Para conferir a similaridade entre as sequências anotadas como de mesma classe, como o agrupamento em *B. subtilis*, nas linhagens de *B. cereus* e *B.*

thuringiensis foram desenhados gráficos de pontos que representam regiões de similaridade entre múltiplas sequências alinhadas duas a duas por meio da ferramenta Polydot para comparação entre as sequências anotadas como RapA (Figura 9) e como RapD (Figura 10).

Não há uma similaridade significativa entre todas as proteínas anotadas como RapA. Nenhuma das proteínas RapA demonstram identidade completa (quadros com linha diagonal integral) com mais de 1 proteína RapA de outra linhagem. Em geral, observa-se uma similaridade pequena ou nula entre essas proteínas.

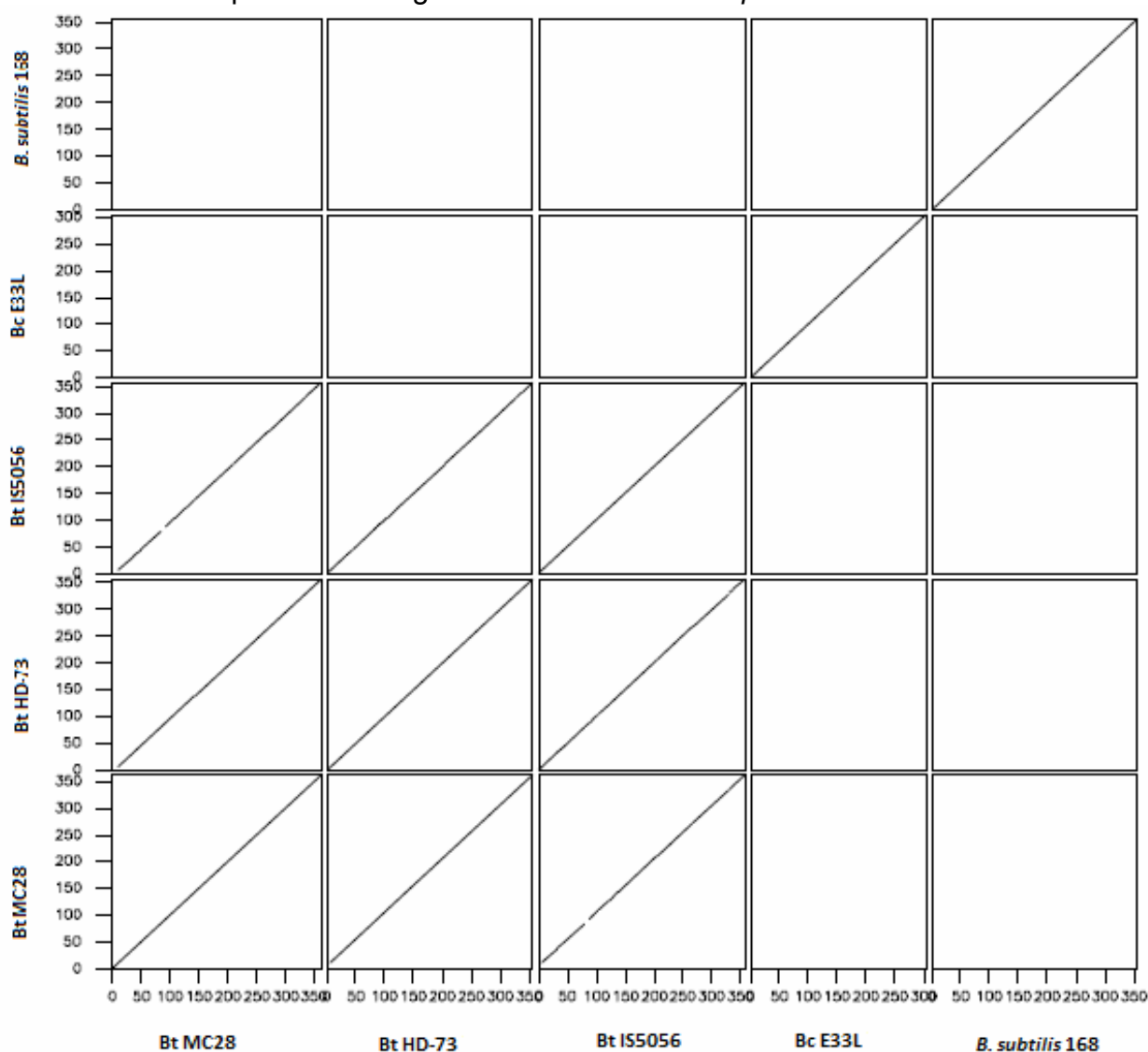
Figura 9 – Representação gráfica do alinhamento entre as sequências proteicas dos produtos dos genes anotados como *rapA*



Bt: *B. thuringiensis*, Bc: *B. cereus*; o número abaixo da linhagem corresponde à numeração encontrada na Tabela 2.

Também não é possível observar uma similaridade significativa entre todas as proteínas anotadas como RapD (Figura 10). Nas três sequências anotadas como RapD de linhagens de *B. thuringiensis*, os genes codificantes dessas proteínas estão localizados em plasmídeos, mas não há uma coincidência entre suas posições ou tamanho dos plasmídeos que os comportam (Figura 8). Os resultados dos alinhamentos das proteínas RapA e RapD demonstram que não há um critério definido para a anotação dos genes *rap* de *B. cereus lato senso* em classes como observado nos genomas de *B. subtilis*.

Figura 10 – Representação gráfica do alinhamento entre as sequências proteicas dos produtos dos genes anotados como *rapD*



Bt: *B. thuringiensis*, Bc: *B. cereus*; o número abaixo da linhagem corresponde à numeração encontrada na Tabela 2.

A identidade entre as proteínas RapD de *B. cereus* e *B. thuringiensis* em relação às proteínas Rap de *B. subtilis* foi observada no Swissprot mas não no Polydot. Esse resultado pode ser justificado pela possível localização dispersa dos resíduos idênticos ou similares na sequência proteica que não permite que sejam visualizadas na representação gráfica, visto que o Swissprot não apresenta numericamente a cobertura entre as sequências alinhadas. Para confirmar a similaridade observada no Swissprot, foi realizada uma comparação dessas sequências pelo BLASTp, revelando uma identidade de 24 a 25% entre as proteínas Rap presentes nas linhagens de *B. cereus lato sensu* e em *B. subtilis* 168, com cobertura variando de 64 a 85%. Ou seja, o nível de detecção de similaridade do Polydot é menor do que do BLASTp.

A análise de similaridade pelo agrupamento foi realizada na tentativa de reunir as proteínas Rap do grupo *B. cereus* em classes como em *B. subtilis*, observar a presença de proteínas idênticas na mesma linhagem e fazer uma relação entre a similaridade de sequência e a localização dessas proteínas no genoma (Figuras 3, 5 a 8). Essa análise resultou em 25 grupos (Quadro 6, Apêndice, pág 100), variando de dois a 25 componentes e seis proteínas Rap que não apresentaram similaridade com nenhuma outra entre as identificadas em *B. cereus lato sensu*. Nesta etapa do estudo as proteínas curtas não foram consideradas.

A linhagem IS5056 de *B. thuringiensis* foi a única que apresentou proteínas idênticas no mesmo genoma. No entanto, os genes codificantes dessas proteínas estão presente em diferentes plasmídeos. Nesta linhagem, Rap7 (pIS56-107) com Rap8 (pIS56-233) são altamente similares, assim como Rap10 (pIS56-16) com Rap16 (pIS56-285) e Rap13 (pIS56-285) com Rap14 (pIS56-328) são similares entre si.

Nenhuma das linhagens de *B. cereus* ou *B. thuringiensis* apresenta conteúdo de proteínas Rap igual. Nos grupos onde há apenas proteínas de *B. thuringiensis* (10, 14, 18, 20 e 24) há uma prevalência de proteínas plasmidiais. Apenas dois grupos têm apenas sequências de *B. cereus*: 21 e 25 e nos dois há sequências da linhagem *B. cereus* Q1.

Quando observados as proteínas plasmidiais mais especificamente, é possível observar um evidente agrupamento de proteínas plasmidiais, apesar de, com exceção dos Grupos 14, 18, 22 e 24, os grupos com proteínas plasmidiais

possuírem também pelo menos uma proteína cromossômica. E também três proteínas plasmidiais não se agruparam com nenhuma outra Rap.

Relacionando as Figuras 3 a 7 com o Quadro 6, foi possível observar que, principalmente para os grupos maiores, as proteínas similares são codificadas por genes com localização semelhante. Como por exemplo os genes da região conservada intermediária/mediana (2,3 Mb) de *B. anthracis* correspondem às proteínas do Grupo 1, com membros também nas linhagens de *B. cereus* e *B. thuringiensis*. Os genes cromossômicos que não estão na posição correspondente codificados na fita contrária, estão associados à genes codificantes de proteínas de ligação ao DNA (DNA-binding protein), mas não há uma relação definida quanto a presença ou não do gene *phr*.

Relação que não foi observada na posição de aproximadamente 1,05 Mb. Apesar de detectar a similaridade de sequência desses genes de *B. anthracis*, a similaridade observada no agrupamento não é verdadeira para os genes concentrados nesta região de *B. cereus* e *B. anthracis*. Outra observação peculiar é da proximidade de cinco dos seis elementos do Grupo 8 com elementos do Grupo 2.

Por meio dessas relações, foi possível observar que os plasmídeos pCT281 (de *B. thuringiensis* CT-43) e pIS56-285 (de *B. thuringiensis* IS5056) são muito semelhantes nos critérios analisados nesse estudo, pois contém os mesmos três grupos Rap proteicos com localização correspondente, além da presença de genes *cry* de mesma classe (2Ab, 1Aa, 1Ia e 2Aa) localizados em regiões correspondentes. O alinhamento da sequência completa dos dois plasmídeos por BLASTn confirmou a probabilidade de serem praticamente o mesmo elemento genético ao resultar em cobertura e identidade de 99%.

Também é possível notar a presença da mesma classe de genes nos plasmídeos com 6,9kb de *B. thuringiensis*: BTB_6p (Bt 407), pCT6880 (CT-43) e pIS56-6 (IS5056). Apesar de os genes *rap* estarem em diferentes localizações dentro dos plasmídeos (Figura 8) o alinhamento desses plasmídeos por BLASTn revelou 100% de identidade e 100% de cobertura. Conseqüentemente, caracterizando-os como mesmo plasmídeo. A presença do mesmo plasmídeo não conjugativo em três linhagens diferentes, pode ser justificada pela ocorrência de mobilização por plasmídeos conjugativos (VILAS-BOAS; SANTOS, 2012).

Não foi possível construir o quadro de agrupamento por similaridade para os peptídeos Phr, pois a baixa identidade auxiliada pelo curto comprimento da proteína, não permitiram a formação de grupos definidos.

Para melhor avaliar a similaridade entre as proteínas Rap obtidas, as sequências de aminoácidos foram alinhadas por meio do programa MEGA 6 e uma árvore filogenética foi construída pelo método de Neighbor-Joining (Figura 11). Foram excluídas do alinhamento as proteínas menores de 300 aminoácidos, resultando em 173 sequências.

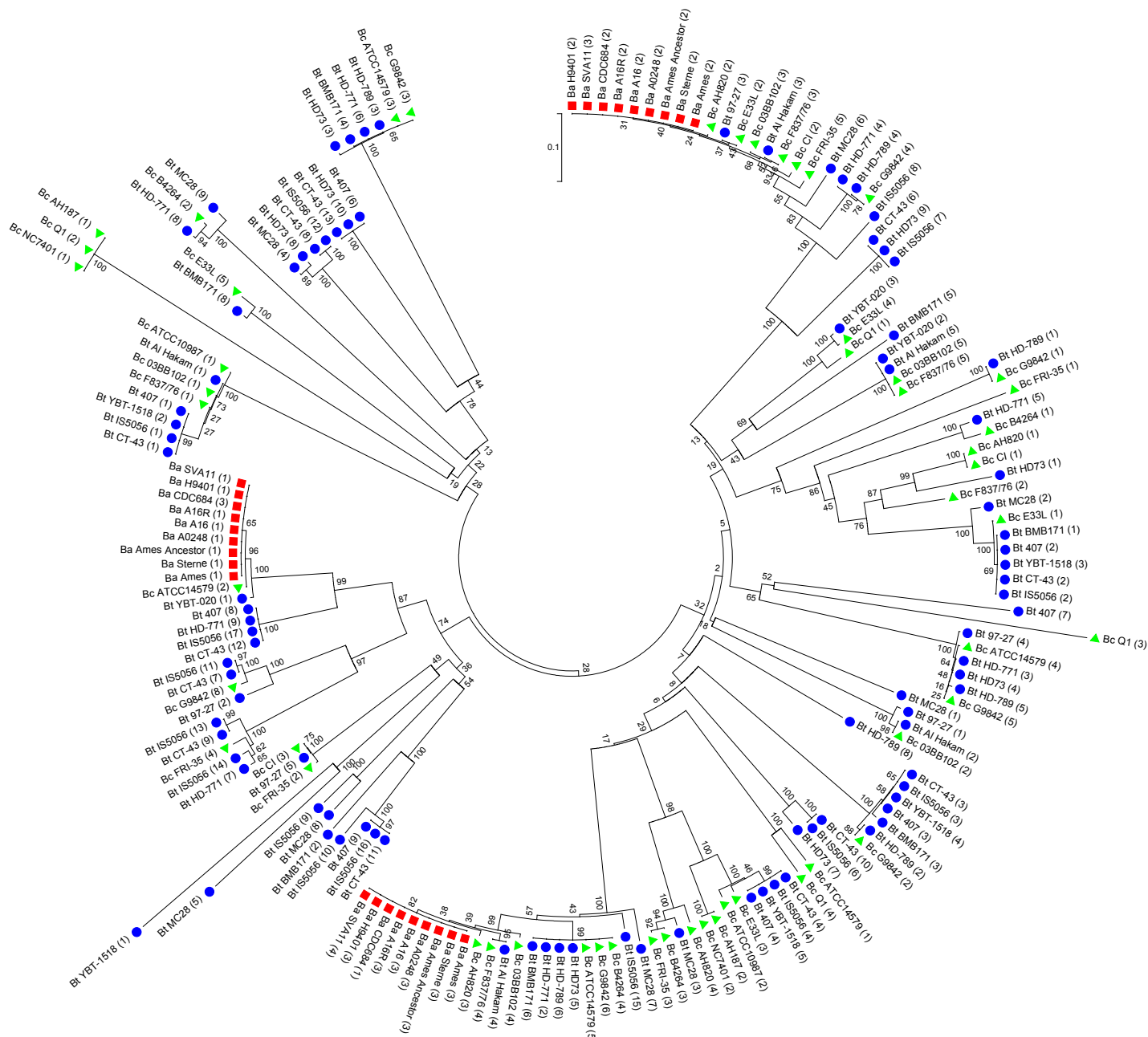
Na Figura 11 é possível notar claramente o agrupamento das sequências proteicas observadas no Quadro 6, como esperado pois a proximidade dos ramos da árvore está relacionada com a similaridade de sequência.

Na árvore também é possível observar a concentração das proteínas de *B. anthracis*, em contraste com às de linhagens de *B. cereus* e *B. thuringiensis*, cujas proteínas Rap aparecem misturadas na árvore. Este agrupamento das proteínas de *B. anthracis* confirma a sintonia observada na Figura 3, pois as sequências das proteínas nesta espécie apresentam extensa homologia e mesmo ordenamento genético, ou seja, os genes com localização conservada codificam proteínas com sequências similares, e até mesmo idênticas.

As mesmas características observadas na árvore filogenética resultante dos genes *rap*, podem ser observadas na árvore gerada para os genes *phr* (Figura 12). No alinhamento (resultado não demonstrado) do qual esta árvore foi gerada, é possível observar vários e grandes *gaps*, o que demonstra que, mesmo nas sequências mais longas, o padrão necessário para a funcionalidade é mantido, de forma que as sequências correspondentes aos *gaps* nos peptídeos mais curtos, provavelmente, levam à formação de alças.

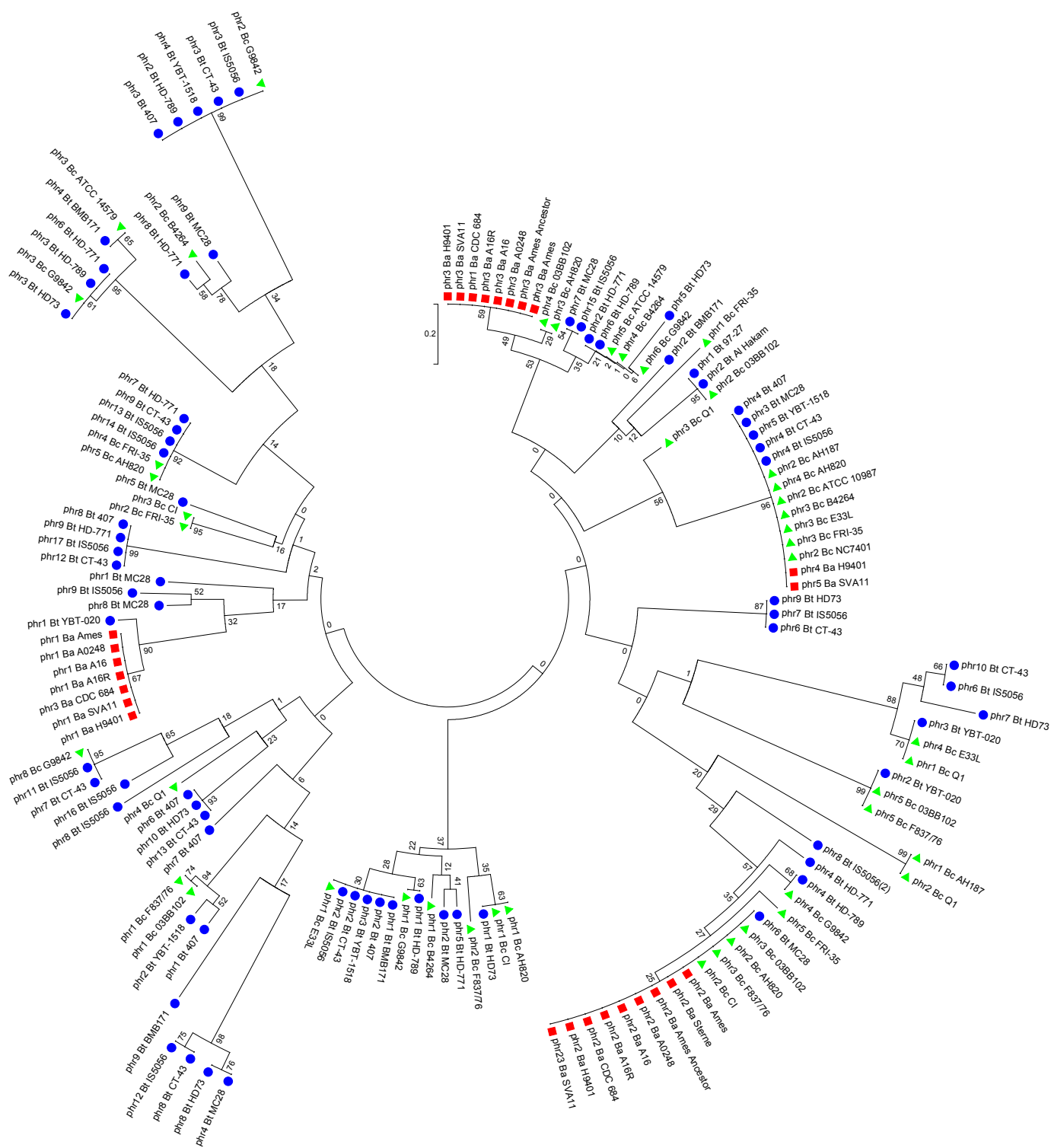
A relevância da análise filogenética se dá por proporcionar a visualização dos agrupamentos das proteínas mais estreitamente relacionadas nas árvores. Esta análise é confiável mesmo com alguns valores de *bootstrap* baixos, pois nos pequenos agrupamentos a formação dos *clusters* é estatisticamente significativa. Os valores mais baixos do *bootstrap* estão presentes quando os grandes *clusters* são relacionados, a ausência de uma similaridade significativa entre eles pode ser justificada pela ausência de necessidade da conservação da sequência estrita dos aminoácidos das proteínas Rap e peptídeos Phr.

Figura 11 – Árvore filogenética das proteínas Rap de *B. cereus lato sensu* construída pelo algoritmo *Neighbour-Joining*. Numeração correspondente à apresentada na Tabela 2.



▲ : *B. cereus*, ● : *B. thuringiensis*, ■ : *B. anthracis*

Figura 12 – Árvore filogenética dos peptídeos Phr de *B. cereus lato sensu* construída pelo algoritmo *Neighbour-Joining*. Numeração correspondente à apresentada na Tabela 2.



Neste estudo, os sistemas Rap-Phr das linhagens de *B. cereus lato sensu* com sequência do genoma completo disponível foram analisados quanto à sua localização, proximidade a elementos transponíveis, tamanho, similaridade aos sistemas de *B. subtilis*, identificação de características determinantes (domínio TPR ou sequência sinal), conservação de sequências entre e dentro das linhagens e quanto ao seu agrupamento em classes por meio de ferramentas de bioinformática. Entretanto, a análise *in silico* tem suas limitações, pois não permite inferir a funcionalidade desses sistemas. Desta forma são necessárias análises funcionais para compreender o papel desses sistemas nas espécies do grupo e a comparação entre os genes plasmidiais e cromossômicos auxiliarão na discussão do papel do plasmídeo para estas bactérias. Devido à grande quantidade de sistemas identificados não é possível realizar a análise funcional de todos eles, mas sim o estudo em linhagens representativas do grupo. Como por exemplo, *B. thuringiensis* HD-73, com sete genes *rap* cromossômicos e três plasmidiais e *B. cereus* G9842, com sete genes *rap* cromossômicos e um plasmidial.

3.5 CONCLUSÃO

A busca e caracterização dos genes *rap-phr* nos genomas de linhagens das três espécies de *B. cereus lato sensu*, bem como o levantamento bibliográfico realizado em torno desses genes, mostrou a grande diversidade dos mesmos em vários aspectos: ocorrência, localização, similaridade de sequências e aspectos funcionais. Levando em consideração as questões taxonômicas relacionadas às espécies do grupo, os dados obtidos neste trabalho não permitiram a identificação de padrões de genes *rap-phr* espécie específicos que possam diferenciar linhagens das três espécies. No entanto, a análise dos resultados desse estudo permite constatar a maior plasticidade genômica das linhagens de *B. thuringiensis* em relação às outras duas espécies. Isto, em razão da maior variação do tamanho dos cromossomos e do número de genes preditos nestes, do maior número de plasmídeos por linhagem e da maior quantidade de genes *rap-phr* por linhagem, principalmente genes plasmidiais.

Quando os dados são analisados quanto à localização do gene, cromossômica ou plasmidial, não foi possível detectar marcadores em relação ao posicionamento desses genes no genoma. No entanto, foram encontrados muitos

genes *rap-phr* nos plasmídeos de *B. cereus* e *B. thuringiensis*. Considerando a relevância funcional dos genes plasmidiais apontada por diversos autores, os resultados obtidos reforçam a importância dos plasmídeos para este grupo bacteriano. Assim, plasmídeos contendo sistemas *rap-phr* podem ser mantidos de forma mais estável em um maior número de espécies ou linhagens que co-existem em um determinado ambiente, uma vez que a presença desses plasmídeos pode trazer uma melhor resposta das bactérias frente às modificações ambientais.

A identificação e a caracterização dos genes *rap-phr* irão auxiliar nas análises funcionais desses genes nas espécies de *B. cereus lato sensu* e, em conjunto com essas análises, permitirá verificar a atividade dos produtos desses genes e, principalmente, contribuir na identificação do real papel dos plasmídeos neste conjunto de bactérias.

REFERÊNCIAS

- AUCHTUNG, J. M.; LEE, C. A.; GROSSMAN, A. D. Modulation of the comA-dependent quorum response in *Bacillus subtilis* by multiple Rap proteins and Phr peptides. **Journal of Bacteriology**, v.188, n.14, p. 5273–5285, jul. 2006.
- ANDERSON, I. et al. Comparative genome analysis of *Bacillus cereus* group genomes with *Bacillus subtilis*. **FEMS Microbiology Letters**, v.250, p. 175–184, 2005.
- BAHL, M. I.; HANSEN, L. H.; SØRENSEN, S. J. Persistence mechanisms of conjugative plasmids. In: **Horizontal Gene Transfer: Genomes in Flux**. Humana Press, 2009. p. 73-102.
- BENSON, D. A. et al. GenBank. In: The 2015 Nucleic Acids Research Database Issue and Molecular Biology Database Collection. **Nucleic Acids Research**, v.43, n.D1, p. D30-D35, jan. 2015.
- BONGIORNI, C. et al. Synergistic regulation of competence development in *Bacillus subtilis* by two Rap-Phr systems. **Journal of Bacteriology**, v.187, n.13, p. 4353–4361, jul. 2005.
- BONGIORNI, C. et al. Rap phosphatase of virulence plasmid pXO1 inhibits *Bacillus anthracis* sporulation. **Journal of Bacteriology**, v.188, n.2, p. 487–498, jan. 2006.
- BRUNSING, R. L. et al. Characterization of sporulation histidine kinases of *Bacillus anthracis*. **Journal of Bacteriology**, v.187, n.20, p. 6972–6981, out. 2005.
- D'ANDREA, L. D.; REGAN, L. TPR proteins: the versatile helix. **TRENDS in Biochemical Sciences**, v.28, n.12, p.655-662, dez. 2003.
- FINN, R.D. et al. The Pfam protein families database. **Nucleic Acids Research**, v. 42, p. D222–D230, 2014.
- GILLIS, A.; MAHILLON, J. Phages preying on *Bacillus anthracis*, *Bacillus cereus*, and *Bacillus thuringiensis*: past, present and future. **Viruses**, v.6, p. 2623-2672, 2014.
- 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, jun. 2000.
- JONES, P. et al. InterProScan 5: genome-scale protein function classification. **Bioinformatics**, v.30, n.9, p. 1236–1240, jan. 2014.
- KOETJE, E. J. et al. A plasmid-borne Rap–Phr system of *Bacillus subtilis* can mediate cell-density controlled production of extracellular proteases. **Microbiology**, v.149, p. 19–28, set. 2003.
- KOLSTØ, A.-B.; TOURASSE, N. J.; ØKSTAD, O. A. What sets *Bacillus anthracis* apart from other *Bacillus* species? **The Annual Review of Microbiology**, v.63, p. 451–76, 2009.

MAHILLON, J et al. IS231 and other *Bacillus thuringiensis* transposable elements: a review. **Genetica**, Holanda, v. 93, p. 13-26, maio 1994.

MARCHLER-BAUER, A. et al. CDD: NCBI's conserved domain database. In: The 2015 Nucleic Acids Research Database Issue and Molecular Biology Database Collection. **Nucleic Acids Research**, v.43, p. D222–D226, 2015.

MEIJER, W. J. J. et al. Characterization of the replication region of the *Bacillus subtilis* plasmid pLS20: a novel type of replicon. **Nucleic Acids Research**, v.23, p. 3214-3223, 1995.

MEIJER, W. J. J. et al. Rolling-circle plasmids from *Bacillus subtilis*: complete nucleotide sequences and analyses of genes of pTA1015, pTA1040, pTA1050 and pTA1060, and comparisons with related plasmids from Gram-positive bacteria. **FEMS Microbiology Reviews**, v. 21, p. 337-368, 1998.

OKINAKA, R. T., et al. An attenuated strain of *Bacillus anthracis* (CDC 684) has a large chromosomal inversion and altered growth kinetics. *BMC Genomics*, v.12, n.477, p. 1-13, 2011.

PARASHAR, V. et al. A plasmid-encoded phosphatase regulates *Bacillus subtilis* biofilm architecture, sporulation, and genetic competence. **Journal of Bacteriology**, v.195, n.10, p. 2437–2448, maio 2013.

PERCHAT, S. et al. A cell–cell communication system regulates protease production during sporulation in bacteria of the *Bacillus cereus* group. **Molecular Microbiology**, v.82, n.3, p. 619–633, set. 2011.

PEREGO, M. et al. Aspartyl-phosphate phosphatases deactivate the response regulator components of the sporulation signal transduction system in *Bacillus subtilis*. **Molecular Microbiology**, v.19, n.6, p. 1151-1157, 1996.

PEREGO, M.; BRANNIGAN, J. A. Pentapeptide regulation of aspartyl-phosphate phosphatases. **Peptides**, v.22, p. 1541–1547, 2001.

PEREGO, M. Forty years in the making: understanding the molecular mechanism of peptide regulation in bacterial development. **PLOS Biology**, v.11, n. 3, mar. 2013.

PETERSEN, N. et al. SignalP 4.0: discriminating signal peptides from transmembrane regions. **Nature Methods**, v.8, p. 785-786, 2011.

POPTSOVA, M. S.; GOGARTEN, J. P. Using comparative genome analysis to identify problems in annotated microbial genomes. **Microbiology**, v.156, p.1909–1917, 2010.

POTTATHIL, M.; LAZAZZERA, B. A. The extracellular Phr peptide-Rap phosphatase signaling circuit of *Bacillus subtilis*. **Frontiers in Bioscience**, v.8, p. 32-45, jan. 2003

RASKO, D. A., et al. Genomics of the *Bacillus cereus* group of organisms. **FEMS Microbiology Reviews**, v.29, p. 303–329, 2005.

- RAYMOND, B. et al. Environmental factors determining the epidemiology and population genetic structure of the *Bacillus cereus* group in the field. **PLoS Pathogens**, v.6, n.5, p. 1-13, maio 2010.
- RICE, P.; LONGDEN, I.; BLEASBY, A. EMBOSS: The European Molecular Biology Open Software Suite. **Trends in Genetics**, v.16, n.6, p. 276—277, 2000.
- RUTHERFORD, K. et al. Artemis: sequence visualization and annotation. **Bioinformatics**, Inglaterra, v.16; n.10, p. 944-945, 2000.
- SANTOS, C. A. et al. Conjugal transfer between *Bacillus thuringiensis* and *Bacillus cereus* strains is not directly correlated with growth of recipient strains. **Journal of Invertebrate Pathology**, v.105, p. 171–175, jul. 2010.
- SINGH, P. K. et al. Mobility of the native *Bacillus subtilis* conjugative plasmid pLS20 is regulated by intercellular signaling. **PLOS Genetics**, v. 9, n. 10, p. 1-13, out. 2013.
- SLAMTI, L. et al. Quorum sensing in *Bacillus thuringiensis* is required for completion of a full infectious cycle in the insect. **Toxins**, v.6, p. 2239-2255, jul. 2014.
- STEPHENSON, S., et al. Molecular analysis of Phr peptide processing in *Bacillus subtilis*. **Journal of Bacteriology**, v.185, n.16, p. 4861–4871, ago. 2003.
- TAMURA K., et al. MEGA6: Molecular Evolutionary Genetics Analysis Version 6.0. **Molecular Biology and Evolution**, v.30, p. 2725-2729, 2013.
- THE UNIPROT CONSORTIUM. UniProt: a hub for protein information. In: The 2015 Nucleic Acids Research Database Issue and Molecular Biology Database Collection. **Nucleic Acids Research**, v.43, n.D1, p. D1-D5, 2015.
- TICKNOR, L. O. et al. Fluorescent amplified fragment length polymorphism analysis of Norwegian *Bacillus cereus* and *Bacillus thuringiensis* soil isolates. **Applied and Environmental Microbiology**, v.67, p. 4863–4873, out. 2001.
- TOURASSE, N. J., et al. The *Bacillus cereus* group: novel aspects of population structure and genome dynamics. **Journal of Applied Microbiology**, v.101, p. 579–593, 2006.
- TOURASSE, N. J. et al., Extended and global phylogenetic view of the *Bacillus cereus* group population by combination of MLST, AFLP, and MLEE genotyping data. **Food Microbiology**, v.28, p. 236-244, 2011.
- UNIVERSITY OF OSLO. **Bacillus cereus group MultiLocus and MultiData Typing website**. 2012. Disponível em: <<http://mlstoslo.uio.no>>. Acesso em 18 jan. 2015.
- VILAS-BOAS, G. et al. Genetic differentiation between sympatric populations of *Bacillus cereus* and *Bacillus thuringiensis*. **Applied and environmental microbiology**, v.68, n.3, p. 1414–1424, mar. 2002.
- VILAS-BÔAS, G. T.; LEMOS, M. V. F. Diversity of *cry* genes and genetic characterization of *Bacillus thuringiensis* isolated from Brazil. **Canadian Journal of Microbiology**, v.50, p. 605-613, set. 2004.

VILAS-BÔAS, G.; PERUCA, A.; ARANTES, O. Biology and taxonomy of *Bacillus cereus*, *Bacillus anthracis*, and *Bacillus thuringiensis*. **Canadian Journal of Microbiology**, v.53, n. 6, p. 673-87, jun. 2007.

VILAS-BÔAS, G, T.; SANTOS, C. A. Conjugation in *B. thuringiensis*: Insights into the Plasmids Exchange Process. In: ***B. thuringiensis* Biotechnology**. SANSINENEA, E. (ed.) Springer, p. 159-174, 2012.

ZHENG, J., et al. Plasmids are vectors for redundant chromosomal genes in the *Bacillus cereus* group. **BMC Genomics**, v.16, n.1, jan. 2015.

ZHONG, C., et al. Determination of plasmid copy number reveals the total plasmid DNA amount is greater than the chromosomal DNA amount in *Bacillus thuringiensis* YBT-1520. **PLoS ONE**, v. 6, n.1, jan. 2011.

APÊNDICE

Tabela 1 - Dados gerais dos genomas de *Bacillus cereus* lato sensu empregados neste estudo.

Linhagem	Origem / Atividade	Ano	Nº de acesso	Genoma					Cromossomo	
				Tamanho (Mb)	%GC	Nº de genes	Nº de proteínas	Nº de plasmídeos	Tamanho (Mb)	%GC
<i>Bacillus cereus</i> ATCC 14579	Fazenda / não-patogênico	2003	NC_004722.1	5,43	35,31	5494	5231	1	5,41	35,3
<i>Bacillus cereus</i> 03BB102	Sangue de paciente com pneumonia severa	2009	NC_012472.1	5,45	35,29	5775	5606	1	5,27	35,4
<i>Bacillus cereus</i> AH187	Vômito de surto emético (Londres)	2008	NC_011658.1	5,60	35,52	5961	5783	4	5,27	35,6
<i>Bacillus cereus</i> AH820	Paciente com periodontite	2008	NC_011773.1	5,59	35,31	5949	5810	3	5,3	35,4
<i>Bacillus cereus</i> ATCC 10987	Produtos lácteos (queijo)	2002	NC_003909.8	5,43	35,52	6014	5844	1	5,22	35,6
<i>Bacillus cereus</i> B4264	Sangue e fluído pleural de caso de pneumonia fatal	2008	NC_011725.1	5,42	35,3	5564	5398	0	5,42	35,3
<i>Bacillus cereus</i> E33L	Carcaça de zebra	2004	NC_006274.1	5,84	35,17	5796	5641	5	5,3	35,4
<i>Bacillus cereus</i> F837/76	Ferida da próstata contaminada	2011	NC_016779.1	5,29	35,4	5609	5468	2	5,22	35,4
<i>Bacillus cereus</i> FRI-35		2012	NC_018491.1	5,38	35,45	5580	5435	4	5,08	35,6
<i>Bacillus cereus</i> G9842	Sangue de surto (Nebraska)	2008	NC_011772.1	5,74	35,05	6001	5857	2	5,39	35,3
<i>Bacillus cereus</i> NC7401	Intoxicacao alimentar tipo emética	2011	NC_016771.1	5,55	35,54	5900	5754	5	5,2	35,6
<i>Bacillus cereus</i> Q1	Reservatório de petróleo (China)	2009	NC_011969.1	5,51	35,5	5622	5489	2	5,21	35,6
<i>Bacillus cereus</i> biovar anthracis str. CI	Chipanzé com sintomas de antrax (Costa do Marfim)	2010	NC_014335.1	5,49	35,27	5733	5558	3	5,2	35,4
Média <i>B. cereus</i>				5,52	35,36	5769	5606	2,54	5,27	35,5
<i>Bacillus thuringiensis</i> serovar <i>konkukian</i> str. 97-27	Tecido humano com necrose severa	2004	NC_005957.1	5,31	35,36	5343	5197	1	5,24	35,4
<i>Bacillus thuringiensis</i>	Mutante	2010	NC_014171.1	5,64	35,19	5513	5352	1	5,33	35,3

BMB171	acristalífero derivado de Bt YBT-1463									
<i>Bacillus thuringiensis</i> Bt407	Inseticida – contra lepidoptera	2012	NC_018877.1	6,13	35,02	6590	6402	9	5,5	35,4
<i>Bacillus thuringiensis</i> HD-771		2012	NC_018500.1	6,44	35,04	6703	6568	8	5,89	35,2
<i>Bacillus thuringiensis</i> HD-789 (<i>israelensis</i>)	Inseticida (USDA Agricultural Research Service)	2012	NC_018508.1	6,33	35,18	6626	6462	6	5,5	35,3
<i>Bacillus thuringiensis</i> MC28	Floresta (China) / Inseticida	2012	CP003687.1	6,69	34,92	6843	6722	7	5,41	35,4
<i>Bacillus thuringiensis</i> YBT-1518	Solo (China) / Tóxica para nematoides	2013	NC_022873.1	6,67	35,29	6877	6738	6	6	35,4
<i>Bacillus thuringiensis</i> serovar <i>chinensis</i> CT- 43	Inseticida – contra lepidoptera e diptera (China)	2011	NC_017208.1	6,15	35,12	6380	6206	10	5,59	35,4
<i>Bacillus thuringiensis</i> serovar <i>finitimus</i> YBT- 020	Solo (China) / Inseticida	2011	NC_017200.1	5,68	35,38	5931	5782	2	5,36	35,5
<i>Bacillus thuringiensis</i> serovar <i>kurstaki</i> str. HD73	Inseticida (Centre OILB / Institute Pasteur)	2013	NC_020238.1	5,91	35,19	6334	6194	7	5,65	35,3
<i>Bacillus thuringiensis</i> serovar <i>thuringiensis</i> str. IS5056	Solo (Polônia)	2013	NC_020376.1	6,77	34,91	7060	6922	14	5,49	35,4
<i>Bacillus thuringiensis</i> str. Al Hakam	Suspeita de arma biológica (Irã)	2006	NC_008600.1	5,31	35,41	4945	4798	1	5,26	35,4
Média <i>B. thuringiensis</i>				6,09	35,17	6262	6112	6,00	5,52	35,4
<i>Bacillus anthracis</i> str. Ames	Vaca morta (Texas)	2002	NC_003997.3	5,23	35,4	5401	5039	0	5,23	35,4
<i>Bacillus anthracis</i> str. Sterne		2004	NC_005945.1	5,23	35,4	5265	4955	0	5,23	35,4
<i>Bacillus anthracis</i> str. 'Ames Ancestor'	Vaca morta	2004	NC_007530.2	5,50	35,26	5735	5305	2	5,23	35,4
<i>Bacillus anthracis</i> str. A0248	Isolado humano (USAMRIID)	2009	NC_012659.1	5,50	35,26	6043	5292	2	5,23	35,4

<i>Bacillus anthracis</i> str. A16		2014	CP001970.1	5,50	35,26	6099	5973	2	5,23	35,4
<i>Bacillus anthracis</i> str. A16R		2014	CP001974.1	5,41	35,3	6046	5920	1	5,23	35,4
<i>Bacillus anthracis</i> str. CDC 684		2009	NC_012581.1	5,51	35,26	6042	5902	2	5,23	35,4
<i>Bacillus anthracis</i> str. H9401	Isolado clínico de coreano com antrax cutâneo	2012	NC_017729.1	5,50	35,26	5917	5791	2	5,22	35,4
<i>Bacillus anthracis</i> str. SVA11	Surto sueco de antrax	2014	CP006742.1	5,49	35,26	5905	5741	2	5,21	35,4
Média de <i>B. anthracis</i>				5,43	35,30	5828	5546	1,44	5,23	35,4
Média grupo do <i>B. cereus</i>				5,70	35,27	5959	5749	3,47	5,35	35,4

Tabela 2- Relação dos genes *rap-phr* encontrados, juntamente com os genes localizados mais proximamente ao lado.

Linhagens	Gene			Dist. (pb)	<i>rap</i>				<i>phr</i>			Dist. (pb)	Gene				
	Produto	Tam. (pb)	Localização		Gene (pb)	Localização			Gene (pb)	Localização	SignalP		Produto	Tam. (pb)	Localização		
Bc ATCC 14579	Hypothetical protein (putative membrane protein)	363	970.010-970.372	915	rap 1	1110	971.287-972.396	+	-	-	-	397	Hypothetical protein	390	972.793-973.182		
	Glyoxalase family protein	978	1.005.923-1.006.900	224	rap 2	1113	1.007.124-1.008.236	+	-	-	-	456	Lipase	777	1.008.692-1009468		
	Hypothetical protein (3-oxoacyl-ACP synthase)	348	2.090.879-2.091.226	191	rap 3	1074	2.091.417-2.092.490	+	cromossomo	phr	126	2.092.487-2.092.612	+	298	Hypothetical protein	828	2.092.910-2.093.737
	Hypothetical protein	132	3.456.965-3.457.096	55	rap 4	1098	3.457.151-3.458.248	-	-	-	-	153	Hypothetical protein	348	3.458.401-3.458.748		
	CDP-glucose 4,6-dehydratase	1089	3.473.041-3.474.129	1146	rap 5	1116	3.475.275-3.476.390	+	-	phr	162	3.476.390-3.476.551	-	315	Methyl-accepting chemotaxis protein	1935	3.476.713-3.478.647
Bc 03BB102	Hypothetical protein (antifungal polypeptide)	351	958.713-959.063	416	rap 1	1110	959.479-960.588	+	-	phr	138	960.585-960.722	+	666	HD-domain protein	1125	961.388-962.512
	Glyoxylase family protein	978	995.273-996.250	258	rap 2	1095	996.508-997.602	+	-	phr	246	997.602-997.847	+	237	Hydrolase, alpha/beta fold family	876	998.084-998.959
	Hypothetical protein (2',3'-cyclic nucleotide 2'phosphodiesterase)	1986	2.425.755-2.427.740	276	rap 3	1101	2.428.174-2.429.274	-	cromossomo	phr	159	2.428.016-2.428.174	+	548	DNA-binding protein	201	2.429.822-2.430.022
	Sugar binding transcriptional regulator, LacI family	990	3.107.928-3.108.917	245	rap 4	1116	3.109.162-3.110.277	+	-	phr	162	3.110.277-3.110.438	+	691	Hypothetical protein	375	3.111.129-3.111.503
	Hypothetical protein (selenoprotein O and cysteine-containing-like protein)	1467	3.342.544-3.344.010	256	rap 5	1095	3.344.266-3.345.360	+	-	phr	138	3.345.360-3.345.497	-	75	Transcriptional repressor of sporulation and protease synthase	600	3.345.572-3.346.171
Bc AH187	Hypothetical protein (selenoprotein O and cysteine-containing like protein)	1467	1.860.569-1.861.648	298	rap 1	1095	1.861.946-1.863.040	+	cromossomo	phr	144	1.863.037-1.863.180	-	438	Transcriptional repressor of sporulation and protease synthase	660	1.863.618-1.863.746
	Enterotoxin C	1080	3.277.512-3.278.978	255	rap 2	1095	3.279.233-3.280.327	+	-	phr	189	3.280.324-3.280.512	+	334	Hypothetical protein	129	3.280.846-3.281.505
Bc AH820	Glyoxylase	978	1.048.925-1.049.902	250	rapA 1	1089	1.050.152-1.051.240	+	-	phr	204	1.051.240-1.051.443	+	107	Transposase	321	1.051.550-1.051.870
	Hypothetical protein (2',3'-cyclic nucleotide 2'phosphodiesterase)	1953	2.408.098-2.410.050	275	rap 2	1101	2.410.483-2.411.583	-	-	phr	159	2.410.325-2.410.483	+	540	DNA-binding protein	201	2.412.123-2.412.323
	Sugar-binding transcriptional regulator, LacI family	990	3.117.374-3.118.363	237	rap 3	1116	3.118.600-3.119.715	+	cromossomo	phr	162	3.119.715-3.119.876	+	1642	Hypothetical protein (fascin domain protein)	549	3.121.518-3.122.066
	Hypothetical protein (selenoprotein O and cysteine-containing homologs)	1467	3.351.527-3.352.993	255	rap 4	1095	3.353.248-3.354.342	+	-	phr	189	3.354.339-3.354.527	+	74	Protease synthase and sporulation negative regulatory protein PAI2	600	3.354.601-3.355.200
	Hypothetical protein (transposase)	135	264.154-264.288	398	rapB 5	177	264.817-264.993	-	pAH820_272	phr	135	264.686-264.820	+	1233	{rap 1035pb} / transcriptional regulator, ArsR family	294	266.226-266.519
Bc ATCC 10987	Hypothetical protein (antifungal polypeptide)	351	1.077.885-1.078.235	417	rap 1	1110	1.078.652-1.079.761	+	-	-	-	61	Hypothetical protein	135	1.079.822-1.079.956		
	Hypothetical protein (selenoprotein O and cysteine-containing like protein)	1467	3.276.635-3.278.101	255	rap 2	1095	3.278.356-3.279.450	+	cromossomo	phr	189	3.279.447-3.279.635	+	73	Transcriptional regulator, putative	303	3.279.708-3.280.010
Bc B4264	Glyoxylase	978	1.029.753-1.030.730	241	rapA 1	1095	1.030.971-1.032.065	+	-	phr	324	1.032.065-1.032.388	+	251	Alpha/beta fold family hydrolase	873	1.032.639-1.033.511
	Hypothetical protein	654	2.068.066-2.068.719	445	rapC 2	1011	2.069.164-2.070.174	+	-	phrF	216	2.070.171-2.070.386	+	423	Group specific protein	783	2.070.809-2.071.591
	Hypothetical protein (selenoprotein O and cysteine-containing homologs)	1467	3.388.585-3.390.051	256	rap 3	1095	3.390.307-3.391.401	+	cromossomo	phr	189	3.391.398-3.391.586	+	74	Protease synthase and sporulation negative regulatory protein PAI2	600	3.391.660-3.392.259
	Hypothetical protein	123	3.408.817-3.408.939	798	rap 4	1116	3.409.737-3.410.852	+	-	phr	162	3.410.852-3.411.013	-	162	Methyl-accepting chemotaxis protein	1983	3.411.175-3.413.157
Bc E33L	Glyoxalase family protein	978	1.039.604-1.040.581	255	rapA 1	1095	1.040.836-1.041.930	+	-	phrA	309	1.041.930-1.042.238	+	248	Alpha-beta hydrolase	876	1.042.486-1.043.361
	2',3'-cyclic-nucleotide 2'-phosphodiesterase	2016	2.391.332-2.393.347	402	rap 2	1113	2.393.749-2.394.861	-	-	-	-	529	DNA-binding protein	201	2.395.390-2.395.590		
	Hypothetical protein (selenoprotein O and cysteine-containing like protein)	1467	3.343.536-3.345.002	255	rap 3	1095	3.345.257-3.346.351	+	cromossomo	phr	189	3.346.348-3.346.536	+	74	Transcriptional repressor of sporulation and protease synthase	600	3.346.610-3.347.209
	Group-specific protein	1008	4.926.891-4.927.898	89	rapA' 4	1095	4.927.987-4.929.081	+	-	phr	306	4.929.081-4.929.386	+	393	t-RNA-Arg-2	79	4.929.779-4.929.857

	Transposase	1119	332.618-333.736	93	rapD 5	915	333.829-334.743	-	pE33L466	-	-	-	674	Hypothetical protein	405	335.417-335.821	
Bc F837/76	Hypothetical protein (antifungal polypeptide)	351	950.976-951.326	417	rap 1	1110	951.743-952.852	+		phr	138	952.849-952.986	+	666	Hypothetical protein (HD family phosphohydrolase)	1125	953.652-954.776
	Glyoxalase	978	988.969-989.946	257	rap 2	1095	990.203-991.297	+		phr	330	991.298-991.627	+	234	Lipase	876	991.861-992.736
	5'-nucleotidase	1986	2.386.623-2.388.608	276	rap 3	1101	2.389.042-2.390.142	-	cromossomo	phr	159	2.388.884-2.389.042	+	548	DNA-binding protein	201	2.390.690-2.390.890
	NTD biosynthesis operon regulator NtdR	990	3.038.858-3.039.847	245	rap 4	1116	3.040.092-3.041.207	+		-	-	-	-	2251	Hypothetical protein (fascin domain protein)	549	3.043.458-3.044.006
	Selenoprotein O and cysteine-containing like protein	600	3.267.926-3.269.392	256	rap 5	1095	3.269.648-3.270.742	+		phr	138	3.270.742-3.270.879	-	75	Protease synthase and sporulation negative regulatory protein PAI2	600	3.270.954-3.271.553
Bc FRI-35	Alpha/beta hydrolase	876	572.861-573.736	239	rap 1	1095	574.271-575.365	-		phr	297	573.975-574.271	+	250	Glyoxylase	978	575.615-576.592
	Site-specific tyrosine recombinase XerD	186	2.542.527-2.542.712	491	rap 2	1095	2.543.322-2.544.416	-		phr	123	2.543.203-2.543.325	-	120	Hypothetical protein (type III secretion protein)	429	2.544.536-2.544.964
	Sporulation and protease synthase transcriptional repressor	174	3.488.225-3.488.398	73	rap 3	1095	3.488.656-3.489.750	-	cromossomo	phr	189	3.488.471-3.488.659	-	255	Hypothetical protein (selenoprotein O and cysteine-containing like protein)	1467	3.490.005-3.491.471
	Beta-lactam antibiotic acyclase	2391	3.672.774-3.675.164	268	rap 4	1092	3.675.432-3.676.523	+		phr	135	3.676.520-3.676.654	+	391	Hypothetical protein	705	3.677.045-3.677.749
	DNA-binding protein	201	4.336.854-4.337.054	547	rap 5	1101	4.337.601-4.338.701	+		phr	159	4.338.701-4.338.859	+	323	2',3'-cyclic-nucleotide 2'-phosphodiesterase	531	4.339.182-4.339.712
Bc G9842	Glyoxylase	978	1.008.459-1.009.436	249	rapA 1	1095	1.009.685-1.010.779	+		phr	201	1.010.779-1.010.979	+	241	Alpha/beta hydrolase	873	1.011.220-1.012.092
	Prolipoprotein diacylglycerol transferase	681	1.651.547-1.652.227	124	rap 2	1095	1.652.488-1.653.582	-		phr	138	1.652.351-1.652.488	+	144	Hypothetical protein	177	1.653.726-1.653.902
	Hypothetical protein (3-oxoacyl-ACP synthase)	348	2.038.048-2.038.395	191	rap 3	1074	2.038.586-2.039.659	+		phr	126	2.039.656-2.039.781	+	297	Group-specific protein	864	2.040.078-2.040.941
	Transcriptional repressor PagR	294	2.363.413-2.363.706	552	rap 4	1101	2.364.416-2.365.516	-	cromossomo	phr	159	2.364.258-2.364.416	+	2865	S-layer protein	1419	2.368.381-2.369.799
	Hypothetical protein	132	3.386.866-3.386.997	55	rap 5	1095	3.387.052-3.388.146	-		-	-	-	-	156	Hypothetical protein	348	3.388.302-3.388.649
	CDP-glucose 4,6 dehydratase	1089	3.427.735-3.428.823	1161	rap 6	1116	3.429.984-3.431.099	+		phr	162	3.431.099-3.431.260	-	149	Methyl-accepting chemotaxis protein	1983	3.431.409-3.433.391
	Hypothetical protein	123	3.441.825-3.441.947	293	rap 7	231	3.442.240-3.442.470	-		-	-	-	-	152	Sensory box/GGDEF family protein	2517	3.442.622-3.445.138
	DNA-binding protein HU	273	88.556-88.828	474	rap 8	1095	89.302-90.396	-	pG9842_209	phr	138	89.168-89.305	+	347	Hypothetical protein (ArsR family transcriptional regulator)	414	90.743-91.156
Bc NC7401	Non-expressed Enterotoxin C	1080	1.818.392-1.820.471	298	rap 1	1095	1.820.769-1.821.863	+		-	-	-	417	Hypothetical protein	174	1.822.280-1.822.453	
	Hypothetical protein (selenoprotein O and cysteine-containing like protein)	1467	3.231.672-3.233.138	255	rap 2	1095	3.233.393-3.234.487	+	cromossomo	phr	189	3.234.484-3.234.672	+	334	Protease synthase and sporulation negative regulatory protein PAI2	606	3.235.006-3.235.611
Bc Q1	Hypothetical protein	246	580.653-580.898	354	rapA 1	1095	581.509-582.603	-		phrA	258	581.252-581.509	+	222	Hypothetical protein (type VII secretion protein)	1173	582.825-583.997
	Enterotoxin C	1080	1.862.023-1.863.102	298	rapA' 2	1095	1.863.400-1.864.494	+	cromossomo	phr	144	1.864.491-1.864.634	-	429	Hypothetical protein	138	1.865.063-1.865.200
	Circumsporozoite protein	480	3.249.022-3.249.501	204	rapE 3	1098	3.249.705-3.250.802	+		phr	189	3.250.805-3.250.993	+	73	Protease synthase and sporulation negative regulatory protein pai2	600	3.251.066-3.251.665
	Hypothetical protein	606	154.706-155.311	59	rap 4	1119	155.498-156.616	-	pBc239	phr	132	155.370-155.501	-	286	Hypothetical protein	153	156.902-157.054
Bc anthracis CI	Glyoxalase	978	997.536-998.513	234	rap 1	1089	998.747-999.835	+		phrA	204	999.835-1.000.038	+	237	Lipase, alpha/beta hydrolase fold family	876	1.000.275-1.001.150
	2',3'-cyclic-nucleotide 2'-phosphodiesterase	1953	2.346.779-2.348.731	275	rapA 2	1101	2.349.164-2.350.264	-	cromossomo	phr	159	2.349.006-2.349.164	+	538	DNA-binding protein	201	2.350.802-2.351.002
	Hypothetical protein (type VII secretion protein)	429	4.218.429-4.218.857	45	rap 3	1170	4.218.902-4.220.071	+		phr	123	4.220.068-4.220.190	-	85	Hypothetical protein (membrane protein)	477	4.220.275-4.220.751
Bt konkukian 97-27	Glyoxalase	978	1.034.348-1.035.325	256	rap 1	1095	1.035.581-1.036.675	+	cromossomo	phr	303	1.036.675-1.036.977	+	228	Alpha/beta hydrolase	876	1.037.205-1.038.080
	Hypothetical protein	633	1.668.976-1.669.608	522	rapB 2	1095	1.670.130-1.671.224	-		-	-	-	-	487	Hypothetical protein	612	1.671.711-1.672.322

	2',3'-cyclic-nucleotide 2'-phosphodiesterase	1953	2.391.766-2.393.718	432	rapA 3	1113	2.394.150-2.395.262	-		-	-	519	DNA-binding protein	210	2.395.781-2.395.990		
	Hypothetical protein	1467	3.340.846-3.342.312	293	rapE 4	1098	3.342.605-3.343.702	-		-	-	153	Hypothetical protein	348	3.343.855-3.344.202		
	Hypothetical protein (typeVII secretion protein)	330	4.277.232-4.277.561	274	rap 5	1095	4.277.835-4.278.929	+		-	-	191	Hypothetical protein	489	4.279.120-4.279.608		
Bt BMB171	glyoxalase	978	1.015.550-1.016.527	234	rap 1	1095	1.016.761-1.017.855	+		phr	246	1.017.855-1.018.100	+	238	Lipase (Alpha/beta hydrolase)	876	1.018.338-1.019.213
	transposase	120	1.513.341-1.513.460	789	rap 2	1095	1.514.380-1.515.474	-		phr	135	1.514.249-1.514.383	-	231	Hypothetical protein	558	1.515.705-1.516.262
	Hypothetical protein	138	1.695.106-1.695.243	-	rap 3	1095	1.695.243-1.696.337	-		-	-	-	360	MerR family transcriptional regulator	546	1.696.697-1.697.242	
	Hypothetical protein	201	2.050.717-2.050.917	1077	rap 4	1074	2.051.994-2.053.067	+	cromossomo	phr	126	2.053.064-2.053.189	+	297	Hypothetical protein	828	2.053.486-2.054.313
	Hypothetical protein (selenoproteinO and cysteine-containing like protein)	1467	3.373.531-3.374.997	256	rap 5	1095	3.375.253-3.376.347	+		-	-	-	283	sporulation/protease synthase transcriptional repressor	600	3.376.630-3.377.229	
	CDP-glucose 4,6-dehydratase	1089	3.392.546-3.393.634	1135	rap 6	1116	3.394.769-3.395.884	+		-	-	-	117	Hypothetical protein	114	3.396.001-3.396.114	
	cellulase	1311	3.402.934-3.404.244	471	rap 7	474	3.404.715-3.405.188	-		-	-	-	152	Sensory box/GGDEF family protein	2730	3.405.340-3.408.069	
	Hypothetical protein	168	159.151-159.318	-	rap 8	1095	159.318-160.412	-	pBMB171	-	-	-	811	Hypothetical protein	390	161.223-161.612	
	Hypothetical protein	534	208.948-209.481	216	rap 9	108	209.870-209.977	-		phr	174	209.697-209.870	+	2039	transposase	276	212.016-212.291
Bt 407	Transposase for insertion sequence element IS232	1296	1.007.722-1.009.017	935	rapF1 1	1110	1.009.952-1.011.061	+		phr	138	1.011.058-1.011.195	-	307	S-layer protein/ peptidoglycan endo-beta-N-acetyl glucosaminidase	1734	1.011.502-1.013.235
	Putative ring-cleaving dioxygenase MhqA	978	1.049.356-1.050.333	256	rap1 2	1095	1.050.589-1.051.683	+		phr	291	1.051.683-1.051.973	+	241	Alpha/beta hydrolase family protein	873	1.052.214-1.053.086
	Prolipoprotein diacylglyceryl transferase	681	1.702.130-1.702.810	118	rap 3	1095	1.703.065-1.704.159	-	cromossomo	phr	138	1.702.928-1.703.065	+	360	Transcriptional regulator, MerR family	546	1.704.519-1.705.064
	Hypothetical protein	1467	3.456.408-3.457.874	256	rapF2 4	1095	3.458.130-3.459.224	+		phr	189	3.459.221-3.459.409	+	74	Protease synthase and sporulation protein PAI2	600	3.459.483-3.460.082
	Cellulase	1269	3.482.955-3.484.223	525	rap 5	207	3.484.748-3.484.954	-		-	-	-	166	Sensory box/GGDEF family protein	2730	3.485.120-3.487.849	
	Transcriptional regulator	228	2.754-2.981	472	rap1 6	1080	3.453-4.532	+	BTB_9p	phr	147	4.529-4.675	+	120	Hypothetical protein	582	4.795-5.376
	Hypothetical protein	393	250.783-251.175	153	rapC 7	1095	251.328-252.422	+	BTB_502p	phr	150	252.422-252.571	+	55	Hypothetical protein	420	252.626-253.045
	Hfq protein	186	10.218-10.403	674	rapF 8	1113	11.208-12.320	-	BTB_78p	phr	135	11.077-11.211	-	218	Hypothetical protein	693	12.538-13.230
	Hypothetical protein	774	2.298-3.071	495	rapK 9	1095	3.566-4.660	-	BTB_6p	-	-	-	454	Putative 23,6 KDa protein (Rep protein)	615	5.114-5.728	
Bt HD-771	Sensory box/GGDEF family protein	2487	382.090-384.576	164	rap 1	219	384.740-384.958	+		-	-	-	471	Glycosidase, family 5	1311	385.429-386.739	
	Methyl-accepting chemotaxis protein	1983	395.995-397.977	151	rap 2	1116	398.289-399.404	-		phr	162	398.128-398.289	-	1161	CDP-glucose 4,6-dehydratase	1089	400.565-401.653
	Hypothetical protein	348	440.736-441.083	156	rap 3	1095	441.239-442.333	+		-	-	-	55	Hypothetical protein	132	442.388-442.519	
	Hypothetical protein (pyruvate kinase)	99	1.456.283-1.456.381	549	rap 4	1101	1.456.930-1.458.030	+	cromossomo	phr	159	1.458.030-1.458.188	+	596	Transcriptional repressor PagR	225	1.458.784-1.458.972
	Alpha/beta fold family hydrolase	873	3.071.733-3.072.605	239	rapA 5	1095	3.073.068-3.074.162	-		phr	225	3.072.844-3.073.068	-	275	Glyoxalase family protein	978	3.074.419-3.075.396
	Hypothetical protein	828	1.786.035-1.786.862	320	rap 6	1074	1.787.304-1.788.377	-		phr	126	1.787.182-1.787.307	+	191	Hypothetical protein (3-oxoacyl-ACP synthase)	348	1.788.568-1.788.915
	Hypothetical protein	1089	71.760-72.845	627	rap 7	1095	73.603-74.697	-	p01	phr	135	73.472-73.606	+	191	Hypothetical protein (ArsR family transcriptional regulator)	294	74.888-75.181
	Hypothetical protein	609	34.808-35.416	347	rap 8	1101	35.763-36.863	+	p03	phr	246	36.860-37.105	+	54	Hypothetical protein	237	37.159-37.395
Hypothetical protein	693	30.301-30.993	218	rap 9	1113	31.211-32.323	+	p04	phr	135	32.320-32.454	-	675	Hfq protein	186	33.129-33.314	
Bt HD-789	Glyoxylase	978	577.762-578.739	249	rapA 1	1095	578.988-580.082	+	cromossomo	phr	201	580.082-580.282	+	241	Alpha/beta fold family hydrolase	873	580.523-581.395
	Prolipoprotein diacylglyceryl transferase	681	1.238.038-1.238.718	124	rap 2	1095	1.238.979-1.240.073	-		phr	138	1.238.842-1.238.979	+	234	Hypothetical protein	87	1.240.307-1.240.393

	Hypothetical protein (3-oxoacyl-ACP synthase)	348	1.600.279-1.600.626	187	rap 3	1074	1.600.813-1.601.886	+		phr	126	1.601.883-1.602.008	+	320	Hypothetical protein	828	1.602.328-1.603.155
	Transcriptional repressor PagR	294	1.923.378-1.923.671	552	rap 4	1101	1.924.381-1.925.481	-		phr	159	1.924.223-1.024.381	+	1665	Hypothetical protein	81	1.927.146-1.927.226
	Hypothetical protein	132	2.883.371-2.883.502	55	rap 5	1098	2.883.557-2.884.654	-		-	-	-	-	156	Hypothetical protein	348	2.884.810-2.885.157
	CDP-glucose 4,6-dehydratase	1089	2.924.271-2.925.359	1163	rap 6	1116	2.926.522-2.927.637	+		phr	162	2.927.637-2.927.798	-	151	Methyl-accepting chemotaxis protein	1983	2.927.949-2.929.931
	Glycosidase, family 5	1311	2.937.000-2.938.310	471	rap 7	246	2.938.781-2.939.026	-		-	-	-	-	137	Sensory box/GGDEF family protein	2487	2.939.163-2.941.649
	Hypothetical protein	489	144.203-144.691	82	rap 8	1098	144.773-145.870	+	p02	-	-	-	-	515	Hypothetical protein	423	146.385-146.807
Bt MC 28	Histidine kinase	1311	216.001-217.311	361	rap 1	1095	217.830-218.924	-		phr	159	217.672-217.830	+	419	Immune inhibitor A	2478	219.343-221.820
	Glyoxalase/bleomycin resistance protein/ dioxygenase	993	246.657-247.649	253	rap 2	1095	247.902-248.996	+	cromossomo	phrA	138	248.996-249.178	+	261	Lipase, putative	852	249.439-250.290
	Hypothetical protein (selenoprotein O and cysteine-containing like protein)	1467	2.596.326-2.597.792	256	rap 3	1095	2.598.048-2.599.142	+		phr	189	2.599.139-2.599.327	+	74	Transcriptional repressor of sporulation and protease synthase	600	2.599.401-2.600.000
	Hypothetical protein	714	134.382-135.095	476	rapD 4	1089	135.705-136.793	-	pMC183	phrD	135	135.571-135.705	+	299	Hypothetical protein (pXO2-72)	222	137.092-137.313
	Hypothetical protein	213	22.088-22.300	64	rap 5	1233	22.476-23.708	-	pMC189	phr	132	22.364-22.495	+	193	Iniciator RepB protein	1212	23.901-24.112
	Cell division protein FtsW/RodA/SpoVE	1119	18.403-19.521	484	rap 6	1113	20.163-21.275	-		phr	159	20.005-20.163	+	439	Hypothetical protein	132	21.714-21.845
	Methyl-accepting chemotaxis protein	1983	34.328-36.310	266	rap 7	1131	36.737-37.867	-	pMC429	phr	162	36.576-36.737	-	353	Protein CgeD	1296	38.220-39.515
	Hypothetical protein	114	346.633-346.746	61	rap 8	1095	346.932-348.026	-		phr	129	346.807-346.935	+	241	ArsR family transcriptional regulator	345	348.267-348.611
	Hypothetical protein	627	23.528-24.154	294	rap 9	1137	24.448-25.584	+	pMC54	phrC	246	25.581-25.826	+	283	Hypothetical protein	237	25.867-26.103
Bt YBT-1518	Hypothetical protein	1197	796.965-798.161	731	rap 1	1104	798.892-799.995	+		-	-	-	-	729	Hypothetical protein	1200	800.724-801.923
	Hypothetical protein (cell division protein)	165	1.228.079-1.228.243	511	rap 2	1110	1.228.754-1.229.863	+		phr	138	1.229.860-1.229.997	-	625	Reverse transcriptase/maturase	1275	1.230.622-1.231.896
	Glyoxylase	978	1.265.489-1.266.466	256	rap 3	1095	1.266.722-1.267.816	+		phr	246	1.267.816-1.268.061	+	262	Lipase	852	1.268.323-1.269.174
	Prolipoprotein diacylglycerol transferase family protein	681	1.965.447-1.966.127	118	rap 4	1095	1.966.382-1.967.476	-	cromossomo	phr	138	1.966.245-1.966.382	+	234	Hypothetical protein (ATPase)	87	1.967.710-1.967.796
	Hypothetical protein (selenoprotein O and cysteine-containing like protein)	1467	3.750.373-3.751.839	256	rap 5	1095	3.752.095-3.753.189	+		phr	189	3.753.186-3.753.374	+	74	Protease synthase and sporulation negative regulatory PAI2	600	3.753.448-3.754.047
	Putative endoglucanase	1311	3.800.347-3.801.657	470	rap 6	219	3.802.127-3.802.345	-		-	-	-	-	166	Sensory box/GGDEF family protein	2487	3.802.511-3.804.997
	Hypothetical protein	444	39.833-40.276	201	rap 7	894	40.477-41.370	-	pBMB0231	-	-	-	-	549	{rap 120pb}/ Hypothetical protein	507	41.919-42.425
Bt chinensis CT-43	Hypothetical protein	147	1.007.717-1.007.863	266	rap 1	1110	1.008.129-1.009.238	+		-	-	-	-	441	S-layer protein	1734	1.009.679-1.011.412
	Glyoxalase family protein	978	1.047.535-1.048.512	256	rap 2	1095	1.048.768-1.049.862	+		phr	291	1.049.862-1.050.152	+	262	Lipase	852	1.050.414-1.051.265
	Prolipoprotein diacylglycerol transferase	681	1.695.052-1.695.732	118	rap 3	1095	1.695.987-1.697.081	-	cromossomo	phr	138	1.695.850-1.695.987	+	320	K+-transporting ATPase, A chain	144	1.697.258-1.697.401
	Hypothetical protein (selenoprotein O and cysteine containing homologs)	1467	3.442.599-3.444.065	256	rap 4	1095	3.444.321-3.445.415	+		phr	189	3.445.412-3.445.600	+	259	Transcriptional repressor of sporulation and protease synthase	600	3.445.674-3.446.273
	Cellulose	1269	3.469.147-3.470.415	525	rap 5	219	3.470.940-3.471.158	-		-	-	-	-	154	Sensory box/GGDEF family protein	2730	3.471.312-3.474.041
	Hypothetical protein (ATPase)	906	104.969-105.874	160	rap 6	1185	106.034-107.218	+	pCT127	phr	159	107.218-107.376	+	91	Hypothetical protein	360	107.467-107.826
	Hypothetical protein	249	25.892-26.140	345	rap 7	1095	26.485-27.579	+		phr	138	27.576-27.713	+	124	Hypothetical protein	165	27.837-28.001
	Hypothetical protein	1491	66.316-67.806	250	rap 8	1074	68.056-69.129	+	pCT281	phr	135	69.129-69.263	-	179	Sml1/Knr4	429	69.442-69.870
	Transcriptional repressor PagR	294	205.546-205.839	190	rap 9	1092	206.029-207.120	+		phrB	135	207.117-207.251	+	213	Transposase	282	207.464-207.745
	Hypothetical protein	600	26.010-	508	rap 10	1116	27.117-	+	pCT51	phr	246	28.232-	+	87	Hypothetical protein	237	28.531-

		26.609				28.232			28.477						28.767		
	Hypothetical protein	753	4.484-5.236	495	rap 11	1056	5.731-6.786	-	pCT6880	-	-	-	-	-	-		
	Hypothetical protein (type IV secretory system conjugative DNA transfer family protein)	693	48.607-49.299	218	rap 12	1113	49.517-50.629	+	pCT72	phr	135	50.626-50.760	-	674	Hfq protein	186	51.434-51.619
	Hypothetical protein	129	6.010-6.138	214	rap 13	1080	6.352-7.431	+	pCT8513	phr	147	7.428-7.574	+	120	Hypothetical protein	582	7.694-8.275
Bt finitimus YBT-020	Glyoxalase family protein	978	1.160.155-1.161.132	218	rap 1	1113	1.161.350-1.162.462	+		phr	132	1.162.459-1.162.590	-	231	Alpha/beta hydrolase fold family lipase	873	1.162.821-1.163.693
	Hypothetical protein (selenoprotein O and cysteine-containing like protein)	1467	3.345.658-3.347.124	256	rap 2	1095	3.347.380-3.348.474	+	cromossomo	phr	138	3.348.474-3.348.611	+	75	Transcriptional repressor of sporulation and protease synthase	600	3.348.686-3.349.285
	Group-specific protein (typeVII secretion protein)	1272	4.932.053-4.933.324	222	rap 3	1095	4.933.546-4.934.640	+		phr	285	4.934.640-4.934.924	+	294	Site-specific recombinase, phage integrase family protein	495	4.935.218-4.935.712
Bt kurstaki HD-73	Glyoxalase / bleomycin resistance	978	1.158.532-1.159.509	257	rap 1	1089	1.159.766-1.160.854	+		phr	192	1.160.854-1.161.045	-	241	Lipase	672	1.161.286-1.161.957
	Hypothetical protein (SAM-dependent methyltransferase)	216	2.650.539-2.650.754	353	rap 2	300	2.651.107-2.651.406	+		-	-	-		269	ATP-binding protein IstB	753	2.651.675-2.652.427
	Transposase A	1098	2.282.037-2.283.134	381	rap 3	1074	2.283.515-2.284.588	+		phr	126	2.284.585-2.284.710	+	297	Hypothetical protein	765	2.285.007-2.285.771
	Hypothetical protein	132	3.590.971-3.591.102	55	rap 4	1095	3.591.157-3.592.251	-	cromossomo	-	-	-		520	Transporter, sodium/bile acid symporter	969	3.592.771-3.593.739
	CDP-glucose 4,6-dehydratase	1089	3.624.548-3.625.636	1121	rap 5	1116	3.626.757-3.627.872	+		phr	162	3.627.872-3.628.033	+	162	Methyl-accepting chemotaxis protein	1983	3.628.195-3.630.177
	Hypothetical protein (glycosyl hydrolase family 5)	1269	3.642.927-3.644.195	513	rap 6	474	3.644.708-3.645.181	-		-	-	-		152	Sensory box/GGDEF	1863	3.645.333-3.647.195
	Hypothetical protein (glycerophosphoryl phosphodiesterase)	567	5.237.190-5.237.756	179	rap 7	1116	5.238.153-5.239.268	-		phr	219	5.237.935-5.238.153	+	204	Hypothetical protein (acetylmuramoyl-L-alanine amidase)	936	5.239.472-5.240.407
	Hypothetical protein	246	51.344-51.589	700	rapD 8	1074	52.289-53.362	+	pAW63	phrD	135	53.362-53.496	+	141	Hypothetical protein (transposase)	870	53.637-54.506
	Low similarity with known proteins (cardiolipin synthase, phospholipase D)	906	54.709-55.614	244	rap 9	1101	55.858-56.958	+	pHT77	phr	159	56.958-57.116	+	91	Hypothetical protein	360	57.207-57.566
	Hypothetical protein	123	1.230-1.352	214	rap 10	1080	1.566-2.645	+	pHT8_1	phr	147	2.642-2.788	+	90	Hypothetical protein	612	2.878-3.489
Bt thuringiensis IS 5056	Mobile element protein	1296	1.003.020-1.004.315	935	rap 1	1110	1.005.250-1.006.359	+		-	-	-		441	S-layer protein, putative	1734	1.006.800-1.008.533
	Glyoxalase family protein	978	1.044.654-1.045.631	256	rap 2	1095	1.045.887-1.046.981	+		phr	201	1.046.981-1.047.181	+	241	Lipase, alpha/beta hydrolase fold family	873	1.047.422-1.048.294
	Prolipoprotein diacylglycerol transferase family protein	681	1.701.604-1.702.284	118	rap 3	1095	1.702.539-1.703.633	-	cromossomo	phr	138	1.702.402-1.702.539	+	144	Hypothetical protein (ATPase)	177	1.703.777-1.703.953
	Selenoprotein O and cysteine-containing-like protein	1467	3.447.353-3.448.819	256	rap 4	1095	3.449.075-3.450.169	+		phr	189	3.450.166-3.450.354	+	74	Protease synthase and sporulation negative regulatory protein PAI2	600	3.450.428-3.451.027
	Putative endoglucanase	1311	3.475.576-3.476.886	473	rap 5	207	3.477.369-3.477.575	-		-	-	-		166	Sensory box/GGDEF family protein	2358	3.477.741-3.480.098
	Hypothetical protein	237	15.229-15.465	54	rap 6	1116	15.764-16.879	-	pIS56-39	phr	246	15.519-15.764	+	159	N-acetylmuramoyl-L-alanine amidase	135	17.038-17.172
	Hypothetical protein (ATPase)	849	80.077-80.925	301	rap 7	1101	81.226-82.326	+	pIS56-107	phr	159	82.326-82.484	+	618	Hypothetical protein	537	83.102-83.638
	Mobile element protein	753	10.033-10.785	280	rap 8	1101	11.065-12.165	+		phr	159	12.165-12.323	+	158	Hypothetical protein (transposase)	159	12.481-12.639
	Bacillolysin	216	111.577-111.792	299	rap 9	1095	112.091-113.185	+		phr	129	113.182-113.310	+	71	Hypothetical protein	162	113.381-113.542
	Hypothetical protein (aspartate phosphatase)	159	2.689-2.847	117	rap 10	1149	3.106-4.254	-	pIS56-16	phr	141	2.964-3.104	-	395	Rep protein	657	4.649-5.305
	Hypothetical protein	249	30.114-30.362	345	rap 11	1095	30.707-31.801	+		phr	138	31.798-31.935	+	124	Hypothetical protein	165	32.059-32.223
	Hypothetical protein (phoH-like family protein)	1491	70.539-72.029	250	rapD 12	1074	72.279-73.352	+	pIS56-285	phrD	135	73.352-73.486	-	179	Hypothetical protein (SMI1/KNR4 family protein)	429	73.665-74.093
	Transcriptional regulator, ArsR family	294	209.772-210.065	190	rap 13	1092	210.255-211.346	+		phrB	135	211.343-211.477	+	213	Hypothetical protein (transposase)	282	211.690-211.971
Transcriptional regulator, ArsR family	294	18.463-18.756	190	rap 14	1092	18.946-20.037	+	pIS56-328	phrB	135	20.034-20.168	+	470	Hypothetical protein (Isp1b protein/Vip1)	2355	20.638-22.992	

	Methyl-accepting chemotaxis protein	1983	61.850-63.832	88	rap 15	1116	64.081-65.196	-		phr	162	63.920-64.081	-	648	Hypothetical protein (SmpA/OmrA family protein)	897	65.844-66.740
	Hypothetical protein (Rep protein)	615	2.434-3.048	454	rap 16	1095	3.502-4.596	+	pIS56-6	phr	141	4.593-4.733	+	65	Hypothetical protein	267	4.798-5.064
	RNA-binding protein Hfq	189	15.954-16.139	674	rap 17	1113	16.944-18.056	-	pIS56-68	phr	135	16.813-16.947	-	218	Hypothetical protein (typeIV secretory system conjugative DNA transfer family protein)	693	18.274-18.966
Bt Al Hakam	Dihydroxyacetone kinase	183	1.004.078-1.004.260	926	rap 1	1173	1.005.186-1.006.358	+		-	-	-		785	HD domain-containing protein	1140	1.007.143-1.008.282
	Glyoxalase	993	1.042.477-1.043.469	206	rap 2	1146	1.043.675-1.044.820	+		phr	246	1.044.820-1.045.065	+	236	Alpha/beta hydrolase fold family, lipase	876	1.045.301-1.046.176
	2',3'-cyclic-nucleotide 2'-phosphodiesterase	1986	2.429.965-2.431.950	434	rap 3	1113	2.432.384-2.433.496	-	cromossomo	-	-	-		527	Transcriptional regulator, DNA binding protein	210	2.434.023-2.434.232
	Lacl family transcriptional regulator	990	3.104.452-3.105.441	230	rap 4	1131	3.105.671-3.106.801	+		-	-	-		2249	Peroxidase	588	3.109.050-3.109.637
	Hypothetical protein (selenoprotein O and cysteine-containing like protein)	1467	3.339.240-3.340.706	202	rap 5	1149	3.340.908-3.342.056	+		-	-	-		212	Transcriptional repressor of sporulation and protease synthase	654	3.342.268-3.342.921
Ba Ames	Glyoxylase family protein	978	1.008.127-1.009.104	224	rap 1	1113	1.009.328-1.010.440	+		phr	132	1.010.437-1.010.568	+	228	Alpha/beta hydrolase family protein	876	1.010.796-1.011.671
	Pseudogene	1952	2.343.235-2.345.186	274	rap 2	1101	2.345.618-2.346.718	-	cromossomo	phr	159	2.345.460-2.345.618	+	540	DNA-binding protein	201	2.347.258-2.347.458
	Sugar-binding transcriptional regulator, Lacl family	990	3.020.027-3.021.016	237	rap 3	1116	3.021.253-3.022.368	+		phr	162	3.022.368-3.022.529	+	1642	Hypothetical protein (peroxidase)	549	3.024.171-3.024.719
Ba Sterne	Glyoxylase	978	1.008.022-1.008.999	224	rap 1	1113	1.009.223-1.010.335	+		-	-	-		356	Alpha/beta hydrolase	876	1.010.691-1.011.566
	Pseudogene	1953	2.343.287-2.345.239	274	rap 2	1113	2.345.671-2.346.783	-	cromossomo	phr	159	2.345.513-2.345.671	+	519	DNA-binding protein	210	2.347.302-2.347.511
	Lacl family transcriptional regulator	990	3.020.718-3.021.707	237	rap 3	1116	3.021.944-3.023.059	+		-	-	-		1803	Hypothetical protein (peroxidase)	549	3.024.862-3.025.410
Ba 'Ames Ancestor'	Glyoxylase	978	1.008.128-1.009.105	224	rap 1	1113	1.009.329-1.010.441	+		-	-	-		356	Alpha/beta hydrolase	876	1.010.797-1.011.672
	Pseudogene	1952	2.343.359-2.345.310	274	rap 2	1101	2.345.742-2.346.842	-	cromossomo	phr	159	2.345.584-2.345.742	+	540	DNA-binding protein	201	2.347.382-2.347.582
	Lacl family transcriptional regulator	990	3.020.155-3.021.144	237	rap 3	1116	3.021.381-3.022.496	+		phr	162	3.022.496-3.022.657	+	1642	Hypothetical protein (peroxidase)	549	3.024.299-3.024.847
Ba A0248	Glyoxylase family protein	978	1.008.028-1.009.005	224	rap 1	1113	1.009.229-1.010.341	+		phr	132	1.010.338-1.010.469	+	228	Hydrolase, alpha/beta fold family	876	1.010.697-1.011.572
	Pseudogene	1074	2.344.137-2.345.210	274	rap 2	1101	2.345.642-2.346.742	-	cromossomo	phr	159	2.345.484-2.345.642	+	540	DNA-binding protein	201	2.347.282-2.347.482
	Sugar-binding transcriptional regulator, Lacl family	990	3.020.055-3.021.044	237	rap 3	1116	3.021.281-3.022.396	+		phr	162	3.022.396-3.022.557	+	1642	Hypothetical protein (peroxidase)	549	3.024.199-3.024.747
Ba A16	Lactoylglutathione lyase	978	1.007.959-1.008.936	224	rap 1	1113	1.009.160-1.010.272	+		phr	132	1.010.269-1.010.400	+	228	Putative hydrolase or acyltransferase (alpha/beta hydrolase superfamily)	876	1.010.628-1.011.503
	5'-nucleotidase/2',3'-cyclic phosphodiesterase	720	2.344.474-2.345.193	274	rap 2	1101	2.345.625-2.346.725	-	cromossomo	phr	159	2.345.467-2.345.625	+	540	Hypothetical protein (DNA-binding protein)	201	2.347.265-2.347.465
	Transcriptional regulator	990	3.020.667-3.021.656	237	rap 3	1116	3.021.893-3.023.008	+		phr	162	3.023.008-3.023.169	+	1642	Hypothetical protein (peroxidase)	549	3.024.811-3.025.359
Ba A16R	Lactoylglutathione lyase	978	1.007.913-1.008.890	224	rap 1	1113	1.009.114-1.010.226	+		phr	132	1.010.223-1.010.354	+	228	Putative hydrolase or acyltransferase (alpha/beta hydrolase superfamily)	876	1.010.582-1.011.457
	5'-nucleotidase/2',3'-cyclic phosphodiesterase	720	2.343.963-2.344.682	274	rap 2	1101	2.345.114-2.346.214	-	cromossomo	phr	159	2.344.956-2.345.114	+	540	Hypothetical protein (DNA-binding protein)	201	2.346.754-2.346.954
	Transcriptional regulator	990	3.020.155-3.021.144	237	rap 3	1116	3.021.381-3.022.496	+		phr	162	3.022.496-3.022.657	+	1642	Hypothetical protein (peroxidase)	549	3.024.299-3.024.847
Ba CDC684	Hypothetical protein (peroxidase)	549	1.213.915-1.214.463	1642	rap 1	1116	1.216.266-1.217.381	-		phr	162	1.216.105-1.216.266	+	237	Sugar-binding transcriptional regulator, Lacl	990	1.217.618-1.218.607
	DNA-binding protein	201	1.891.515-1.891.715	540	rap 2	1101	1.892.255-1.893.355	+	cromossomo	phr	159	1.893.355-1.893.513	+	274	2',3'-cyclic-nucleotide 2'-phosphodiesterase	1074	1.893.787-1.894.860
	Alpha/beta fold family hydrolase	876	3.227.275-3.228.150	228	rap 3	1113	3.228.506-3.229.618	-		phr	132	3.228.378-3.228.509	+	224	Glyoxylase	978	3.229.842-3.230.819
Ba H9401	Glyoxalase	993	999.100-1.000.092	224	rap 1	1113	1.000.316-1.001.428	+	cromossomo	phr	132	1.001.425-1.001.556	+	252	Alpha/beta fold family hydrolase	852	1.001.808-1.002.659
	2',3'-cyclic-nucleotide 2'-phosphodiesterase	1074	2.334.401-2.335.474	274	rap 2	1113	2.335.906-2.337.018	-		phr	159	2.335.748-2.335.906	+	519	DNA-binding protein	210	2.337.537-2.337.746

	Lacl family sugar-binding transcriptional regulator	990	3.010.620-3.011.609	300	rap 3	1053	3.011.909-3.012.961	+		phr	162	3.012.961-3.013.122	+	1642	Peroxidase	591	3.014.764-3.015.354
	Hypothetical protein (histidine kinase)	675	3.264.439-3.265.113	76	rap 4	345	3.265.189-3.265.533	+		phr	192	3.265.530-3.265.721	+	74	Transcriptional repressor of sporulation and protease synthase	600	3.265.795-3.266.394
	Glyoxalase	993	995.350-996.342	224	rap 1	1113	996.566-997.678	+		phr	123	997.683-997.805		252	Alpha/beta fold family hydrolase	852	998.057-998.908
	2',3'-cyclic-nucleotide 2'-phosphodiesterase	1953	2.329.364-2.331.316	274	rap 2	141	2.331.748-2.331.888	-		phr	159	2.331.590-2.331.748	+	519	DNA-binding protein	210	2.333.378-2.333.587
Ba SVA 11	Lacl family sugar-binding transcriptional regulator	990	3.005.601-3.006.590	300	rap 3	987	2.331.873-2.332.859	-	cromossomo	phr	162	3.007.942-3.008.103	+	1642	Peroxidase	591	3.009.745-3.010.335
	Hypothetical protein (histidine kinase)	675	3.259.483-3.260.157	76	rap 4	1053	3.006.890-3.007.942	+		phr	192	3.260.574-3.260.765	+	74	Transcriptional repressor of sporulation and protease synthase	600	3.260.839-3.261.438
					rap 5	345	3.260.233-3.260.577	+		phr							

Tam.: Tamanho; Dist.: distância entre o sistema *rap-phr* e o gene localizado mais proximamente; Bc: *Bacillus cereus*, Bt: *B. thuringiensis*, Ba: *B. anthracis*; **rap**: genes *rap* localizados por '*rap*', **rap**: genes *rap* encontrados como '*response regulator aspartate phosphatase*', **rap**: genes *rap* encontrados por similaridade no BLAST; **phr**: *phr* anotado como *hypothetical protein*, *group-specific protein*, *putative exported peptide*, *lipoprotein*; **phr**: *phr* anotado como: *phr*, *response regulator aspartate phosphatase inhibitor*.

Tabela 3 – Relação dos plasmídeos contidos nas linhagens de *B. cereus* (Bc), *B. thuringiensis* (Bt) e *B. anthracis* (Ba) empregadas neste estudo, bem como a localização do gene *rap* dentro do plasmídeo.

Linhagem	Plasmídeo	Nº de acesso	Tamanho (Kb)	%GC	Localização do <i>rap</i>	
Bc ATCC 14579	pBClin15	NC_004721.2	15	38,0	-	-
Bc 03BB102	p03BB102_179	NC_012473.1	180	32,2	-	-
Bc AH187	pAH187_3	NC_011657.1	3,1	34,9	-	-
	pAH187_12	NC_011654.1	12	31,1	-	-
	pAH187_45	NC_011656.1	45	35,5	-	-
	pAH187_270	NC_011655.1	270	34,2	-	-
Bc AH820	pAH820_3	NC_011776.1	3,1	34,9	-	-
	pAH820_10	NC_011771.1	11	33,6	-	-
	pAH820_272	NC_011777.1	272	33,6	264.817-264.993	-
Bc ATCC 10987	pBc10987	NC_005707.1	208	33,4	-	-
Bc B4264	-	-	-	-	-	-
Bc E33L	pE33L5	NC_007104.1	5,1	30,9	-	-
	pE33L8	NC_007106.1	8,2	31,9	-	-
	pE33L9	NC_007107.1	9,1	31,0	-	-
	pE33L54	NC_007105.1	53	31,9	-	-
	pE33L466	NC_007103.1	466	33,1	333.829-334.743	-
Bc F837/76	pF837_10	NC_016780.1	10	31,1	-	-
	pF837_55	NC_016794.1	55	36,2	-	-
Bc FRI-35	p04	NC_018494.1	3,1	34,9	-	-
	p03	NC_018499.1	36	30,7	-	-
	p02	NC_018493.1	41	32,5	-	-
	p01	NC_018492.1	219	33,4	-	-
Bc G9842	pG9842_140	NC_011774.1	140	32,9	-	-
	pG9842_209	NC_011775.1	209	30,0	89.302-90.396	-
Bc NC7401	pNC4	NC_016774.1	3,1	34,9	-	-
	pNC3	NC_016793.1	3,9	34,7	-	-
	pNC2	NC_016773.1	5,4	34,1	-	-
	pNC1	NC_016772.1	48	36,5	-	-
	pNCcld	NC_016792.1	270	34,2	-	-
Bc Q1	pBc53	NC_011971.1	53	35,1	-	-
	pBc239	NC_011973.1	239	33,5	155.498-156.616	-
Bc anthracis CI	pBAslCI14	NC_014333.1	14	37,9	-	-
	pCIXO2	NC_014332.1	182	33,1	-	-
	pCIXO1	NC_014331.1	94	32,5	-	-
Bt konkukian 97-27	pBT9727	NC_006578.1	77	32,6	-	-
Bt BMB171	pBMB171	NC_014172.1	313	33,3	-	159.318-160.412 209.870-209.977
Bt 407	BTB_2p	NC_018885.1	2,1	34,8	-	-
	BTB_5p	NC_018884.1	5,5	31,6	-	-
	BTB_6p	NC_018883.1	6,9	31,8	-	3.566-4.660
	BTB_7p	NC_018882.1	7,6	32,2	-	-
	BTB_8p	NC_018881.1	8,2	29,7	-	-
	BTB_9p	NC_018886.1	8,5	30,8	-	3.453-4.532
	BTB_15p	NC_018880.1	151	35,7	-	-
	BTB_78p	NC_018879.1	78	32,3	-	11.208-12.320
	BTB_502p	NC_018878.1	502	31,5	-	251.328-252.422
Bt HD-771	p04	NC_018488.1	65	32,0	-	31.211-32.323
	p08	NC_018490.1	8	30,1	-	-
	p07	NC_018503.1	9	33,5	-	-
	p06	NC_018489.1	14	35,2	-	-
	p01	NC_018486.1	171	33,0	-	73.603-74.697
	p05	NC_018502.1	45	36,0	-	-
	p03	NC_018487.1	70	34,3	cry	35.763-36.863
	p02	NC_018501.1	169	33,2	-	-
Bt HD-789	p06	NC_018512.1	6,8	36,0	-	-
	p05	NC_018511.1	7,7	35,3	-	-

	p04	NC_018517.1	15	39,7	-	-	
	p03	NC_018510.1	225	33,1	cry	-	
	p02	NC_018509.1	235	36,6	-	144.773-145.870	
	p01	NC_018516.1	350	33,4	-	-	
Bt MC 28	pMC8	NC_018684.1	7,8	32,1	-	-	
	pMC54	NC_018694.1	54	34,7	-	24.448-25.584	
	pMC95	NC_018685.1	95	34,0	cry	-	
	pMC183	NC_018686.1	183	32,8	cry	135.705-136.793	
	pMC189	NC_018687.1	190	33,4	cry	22.476-23.708	
	pMC319	NC_018688.1	320	32,5	cry	-	
						20.163-21.275	
		pMC429	NC_018689.1	430	32,6	-	36.737-37.867 346.932-348.026
Bt YBT-1518	pBMB0228	NC_020124.1	18	29,5	-	-	
	pBMB0229	NC_022874.1	45	36,6	-	-	
	pBMB0230	NC_022875.1	49	33,7	-	-	
	pBMB0231	NC_022876.1	146	33,9	-	40.477-41.370	
	pBMB0232	NC_022877.1	172	33,8	-	-	
	pBMB0233	NC_022882.1	241	35,0	-	-	
Bt chinensis CT-43	pCT6880	NC_017210.1	6,9	31,8	-	5.731-6.786	
	pCT8252	NC_017211.1	8,2	32,4	-	-	
	pCT8513	NC_017207.1	8,5	30,8	-	6.352-7.431	
	pCT9547	NC_017212.1	9,5	33,1	-	-	
	pCT14	NC_017209.1	15	31,5	-	-	
	pCT51	NC_017204.1	51	35,0	-	27.117-28.232	
	pCT72	NC_017205.1	72	32,0	-	49.517-50.629	
	pCT83	NC_017206.1	84	33,2	-	-	
	pCT127	NC_017202.1	128	32,1	cry	106.034-107.218 26.485-27.579	
		pCT281	NC_017203.1	281	33,0	cry	68.056-69.129 206.029-207.120
Bt finitimus YBT-020	pBMB28	NC_017199.1	139	33,9	cry	-	
	pBMB26	NC_017201.1	188	33,1	cry	-	
Bt kurstaki HD-73	pHT7	NC_020243.1	7,6	32,2	-	-	
	pHT8_2	NC_020242.1	8,2	29,7	-	-	
	pHT8_1	NC_020241.1	8,5	30,8	-	1.566-2.645	
	pHT11	NC_020250.1	12	31,8	-	-	
	pAW63	NC_020240.1	72	33,8	-	52.289-53.362	
	pHT77	NC_020239.1	76	30,8	-	55.858-56.958	
	pHT73	NC_020249.1	77	34,7	cry	-	
Bt thuringiensis IS 5056	pIS56-6	NC_020390.1	6,9	31,8	-	3.502-4.596	
	pIS56-8	NC_020377.1	8,2	32,4	-	-	
	pIS56-9	NC_020378.1	9,7	33,0	-	-	
	pIS56-11	NC_020391.1	11	31,6	-	-	
	pIS56-15	NC_020379.1	15	35,7	-	-	
	pIS56-16	NC_020380.1	16	33,3	-	3.106-4.254	
	pIS56-39	NC_020381.1	40	34,9	-	15.764-16.879	
	pIS56-63	NC_020392.1	64	34,7	cry	-	
	pIS56-68	NC_020382.1	69	31,8	-	16.944-18.056	
	pIS56-85	NC_020383.1	85	33,2	-	-	
	pIS56-107	NC_020393.1	107	31,0	cry	81.226-82.326	
		pIS56-233	NC_020394.1	234	32,7	-	11.065-12.165 112.091-113.185
							30.707-31.801 72.279-73.352 210.255-211.346
		pIS56-285	NC_020384.1	285	33,0	cry	-
						18.946-20.037 64.081-65.196	
Bt Al Hakam	pALH1	NC_008598.1	56	36,2	-	-	
Ba Ames	-						
Ba Sterne	-						
Ba 'Ames Ancestor'	pXO2	NC_007323.3	95	33,0	-	-	
	pXO1	NC_007322.2	182	32,5	-	-	

Ba A0248	pXO2	NC_012655.1	95	33,0	-
	pXO1	NC_012656.1	182	32,5	-
Ba A16	pXO2	CP001972.1	95	33,0	-
	pXO1	CP001971.1	182	32,5	-
Ba A16R	pXO1	CP001975.1	182	32,5	-
Ba CDC684	pXO2	NC_012577.1	95	33,1	-
	pXO1	NC_012579.1	182	32,5	-
Ba H9401	BAP2	NC_017727.1	95	33,1	-
	BAP1	NC_017726.1	182	32,5	-
Ba SVA 11	pXO2	CP006744.1	95	33,0	-
	pXO1	CP006743.1	182	32,5	-

	Tam prot	381aa	381aa	381aa	381aa	381aa	381aa	381aa
rapG	Local.	4140260-4141357	3963013-3964110	4136253-4137350	3944908-3946005	4071492-4072589	74887-75984	7510123-7511220
	Frame	+	+	+	+	+	-	+
	Tam gene	1098pb	1098pb	1098pb	1098pb	1098pb	1098pb	1098pb
	Tam prot	365aa	365aa	354aa	365aa	365aa	365aa	365aa

	Bs XF-1	Bs subtilis 6051-HGW	Bs subtilis BAB-1	Bs BEST7003	Bs PY79	Bs subtilis JH642 s. AG174	Bs subtilis AG1839
rapJ	Local.	297506-298627	304429-305550	304427-305548	304435-305556	304438-305559	304438-305559
	Frame	+	+	+	+	+	+
	Tam gene	1122pb	1122pb	1122pb	1122pb	1122pb	1122pb
	Tam prot	373aa	373aa	373aa	373aa	373aa	373aa
rapC	Local.	422095-423261	428829-429977	428806-429975	428823-429971	428826-429974	428826-429974
	Frame	+	+	+	+	+	+
	Tam gene	1167pb	1149pb	1070pb	1149pb	1149pb	1149pb
	Tam prot	388aa	382aa	389aa	382aa	382aa	382aa
rapI	Local.		547303-548478			530852-532027	530852-532027
	Frame		+			+	+
	Tam gene		1176pb			1176pb	1176pb
	Tam prot		391aa			391aa	391aa
rapH	Local.		750956-752086	722645-723775	726194-727324	734507-735637	734507-735637
	Frame		+	+	+	+	+
	Tam gene		1131pb	1131pb	1131pb	1131pb	1131pb
	Tam prot		376aa	376aa	376aa	376aa	376aa
rapA	Local.	1319005-1320135	1315866-1317002	1291110-1292240	1278909-1280039	1299423-1300553	1299423-1300553
	Frame	+	+	+	+	+	+
	Tam gene	1131pb	1137pb	1131pb	1131pb	1131pb	1131pb
	Tam prot	376aa	378aa	376aa	376aa	376aa	376aa
rapK	Local.		2062149-2063264	2028344-2029459	2014405-2015520	2034918-2036033	2034918-2036033
	Frame		+	+	+	+	+
	Tam gene		1116pb	1116pb	1116pb	1116pb	1116pb
	Tam prot		371aa	371aa	371aa	371aa	371aa
rapE	Local.		2659196-2660323	2491005-2492132	24770067-2478194	2631970-2633106	2631970-2633106
	Frame		+	+	+	+	+
	Tam gene		1128pb	1128pb	1128pb	1137pb	1137pb
	Tam prot		375aa	375aa	375aa	378aa	378aa
rapD	Local.	3581362-3582426	3744355-3745419	3571902-3572966	3562203-3563267	3717116-3718180	3722387-3723451
	Frame	+	+	+	+	+	+
	Tam gene	1065pb	1065pb	1065pb	1065pb	1065pb	1065pb
	Tam prot	376aa	354aa	354aa	354aa	354aa	354aa
rapB	Local.	3609878-3611011	3771005-3772138	2487803-2488930	3598554-3599687	3588855-3589988	3743768-3744901 3750172

	Frame	-	-	-	-	-	-
	Tam gene	1134pb	1134pb	1128pb	1134pb	1134pb	1134pb
	Tam prot	377aa	377aa	375aa	377aa	377aa	377aa
	Local.	3684941-3686086	3846006-3847151	3673555-3674700	3663856-3665001	3818768-3819913	3824039-3825184
rapF	Frame	+	+	+	+	+	+
	Tam gene	1146pb	1146pb	1146pb	1146pb	1146pb	1146pb
	Tam prot	381aa	381aa	381aa	381aa	381aa	381aa
	Local.		4140264-4141361	3967695-3968792	3958112-3959209	4113022-4114119	4118293-4119390
rapG	Frame		+	+	+	+	+
	Tam gene		1098pb	1098pb	1098pb	1098pb	1098pb
	Tam prot		365aa	365aa	365aa	365aa	365aa

Tabela 7 – Relação dos domínios encontrados nos três bancos de dados para cada um dos genes *rap* identificados.

Linhagem	Nº rap	Tamanho proteína (aa)	Domínios		
			CDD	Pfam	Interpro
Bc ATCC 14579	rap 1	369	NrfG TPR_12 TPR_11	TPR11	TLH
	rap 2	370	NrfG TPR_12 TPR_11	N.I	TLH TRCD TR
	rap 3	357	NrfG TPR_12 TPR_11	TPR11	TLH TRCD
	rap 4	365	NrfG TPR_12 TPR_19	N.I	TLH TRCD TR
	rap 5	371	NrfG TPR_12	N.I	TLH TR
Bc 03BB102	rap 1	369	TPR_12	N.I	TLH
	rap 2	364	NrfG TPR_12 TPR_11	TPR1	TLH TRCD TR TPR1
	rap 3	366	NrfG TPR_12	N.I	TLH TRCD
	rap 4	371	NrfG TPR_12	N.I	TLH TR
	rap 5	364	NrfG TPR_12	TPR12/TPR8	TLH TRCD TR
Bc AH187	rap 1	364	NrfG TPR_12	TPR2	TLH TPR2
	rap 2	364	NrfG TPR_12	TPR12	TLH TRCD TR
Bc AH820	rapA 1	362	NrfG TPR_12 TPR_11	TPR11	TLH TRCD TR
	rap 2	366	NrfG TPR_12	N.I	TLH TRCD
	rap 3	371	NrfG TPR_12	TPR1	TLH TR TPR1
	rap 4	364	NrfG TPR_12 TPR_11	TPR1	TLH TR TPR1
	rapB 5	58	N.I	N.I	N.I
Bc ATCC 10987	rap 1	369	TPR_12	N.I	TLH
	rap 2	364	NrfG TPR_12	TPR12	TLH TRCD TR
Bc B4264	rapA 1	364	NrfG TPR_12 TPR_11	TPR2	TLH TR TPR2
	rapC 2	336	NrfG TPR_12	N.I	TLH
	rap 3	364	NrfG TPR_12 TPR_11	TPR1	TLH TRCD TR TPR1
	rap 4	371	NrfG TPR_12	N.I	TLH TR
Bc E33L	rapA 1	364	NrfG TPR_12	N.I	TLH
	rap 2	370	NrfG TPR_12	N.I	TLH TRCD
	rap 3	364	NrfG TPR_12	TPR12	TLH TRCD TR
	rapA 4	364	NrfG TPR_12	N.I	TLH TRCD
	rapD 5	304	NrfG TPR_12 TPR_11	TPR11	TLH TRCD TR
Bc F837/76	rap 1	369	TPR_12	N.I	TLH
	rap 2	364	NrfG TPR_12 TPR_11	N.I	TLH
	rap 3	366	NrfG TPR_12	N.I	TLH TRCD
	rap 4	371	TPR_1 NrfG TPR_12	N.I	TLH
	rap 5	364	NrfG TPR_12	TPR12/TPR8	TLH TRCD TR
Bc FRI-35	rap 1	364	NrfG TPR_12	N.I	TLH TR
	rap 2	364	NrfG TPR_12	N.I	TLH TRCD TR
	rap 3	364	NrfG TPR_12	TPR1	TLH TRCD TR TPR1
	rap 4	363	NrfG TPR_12 TPR_11	TPR1	TLH TRCD TPR1
	rap 5	366	NrfG TPR_12	N.I	TLH TRCD
Bc G9842	rapA 1	364	NrfG TPR_12 TPR_16	N.I	TLH TRCD TR
	rap 2	364	NrfG TPR_12	N.I	TLH TR
	rap 3	357	NrfG TPR_12 TPR_11	TPR11	TLH TRCD
	rap 4	366	NrfG TPR_12	N.I	TLH TRCD
	rap 5	364	NrfG TPR_12 TPR_19	N.I	TLH TRCD TR
	rap 6	371	NrfG TPR_12	N.I	TLH TR
	rap 7	76	Outro	N.I	N.I
	rap 8	364	NrfG TPR_12	N.I	TLH TR
Bc NC7401	rap 1	364	NrfG TPR_12	TPR2	TLH TPR2
	rap 2	364	NrfG TPR_12	TPR12	TLH TRCD TR
Bc Q1	rapA 1	364	NrfG TPR_12	N.I	TLH
	rapA 2	364	NrfG TPR_12	TPR2	TLH TPR2
	rapE 3	365	NrfG TPR_12/_11/_19	TPR12/TPR12	TLH TRCD TR
	rap 4	372	NrfG TPR_12	TPR11	TLH
Bc anthracis CI	rap 1	362	NrfG TPR_12 TPR_11	TPR11	TLH TRCD TR
	rapA 2	366	NrfG TPR_12	N.I	TLH TRCD
	rap 3	389	NrfG TPR_12	N.I	TLH TRCD TR
Bt konkukian 97-27	rap 1	364	NrfG TPR_12 TPR_11	TPR1	TLH TRCD TPR1
	rapB 2	364	NrfG TPR_12	TPR12	TLH TR
	rapA 3	370	NrfG TPR_12	N.I	TLH TRCD
	rapE 4	365	NrfG TPR_12 TPR_19	N.I	TLH TRCD TR
	rap 5	364	NrfG TPR_12	N.I	TLH TRCD TR
Bt BMB171	rap 1	364	NrfG TPR_12	N.I	TLH

	rap 2	364	NrfG TPR_12	N.I	TLH TRCD
	rap 3	364	NrfG TPR_12	N.I	TLH TR
	rap 4	357	NrfG TPR_12 TPR_11	TPR11	TLH TRCD
	rap 5	364	NrfG TPR_12	N.I	TLH TRCD
	rap 6	371	NrfG TPR_12	N.I	TLH TR
	rap 7	157	NrfG TPR_12 TPR_11	TPR11	TLH
	rap 8	364	NrfG TPR_12	TPR11	TLH TRCD TR
	rap 9	35	N.I	N.I	N.I
Bt 407	rapF1 1	369	TPR_12	N.I	TLH
	rapI 2	364	NrfG TPR_12	N.I	TLH
	rap 3	364	NrfG TPR_12	N.I	TLH TR
	rapF2 4	364	NrfG TPR_12	TPR12	TLH TRCD TR
	rap 5	68	N.I	N.I	N.I
	rapI 6	359	NrfG TPR_12	N.I	TLH
	rapC 7	364	NrfG TPR_12 TPR_12	N.I	TLH TR
	rapF 8	370	NrfG TPR_12	N.I	TLH TR
	rapK 9	364	NrfG TPR_12 TPR_11	TPR1	TLH TRCD TR TPR1
Bt HD-771	rap 1	72	N.I	N.I	N.I
	rap 2	371	NrfG TPR_12	N.I	TLH TR
	rap 3	364	NrfG TPR_12 TPR_19	N.I	TLH TRCD TR
	rap 4	366	NrfG TPR_12	N.I	TLH TRCD
	rapA 5	364	NrfG TPR_12	TPR2	TLH TR TPR2
	rap 6	357	NrfG TPR_12	TPR11	TLH TRCD
	rap 7	364	NrfG TPR_12	N.I	TLH TRCD TR TPR1
	rap 8	366	NrfG TPR_12	N.I	TLH
	rap 9	370	NrfG TPR_12 TPR_11	N.I	TLH TR
Bt HD-789	rapA 1	364	NrfG TPR_12 TPR_16	N.I	TLH TRCD TR
	rap 2	364	NrfG TPR_12	N.I	TLH TR
	rap 3	357	NrfG TPR_12/_11/_10	TPR11	TLH TR
	rap 4	366	NrfG TPR_12	N.I	TLH TRCD
	rap 5	365	NrfG TPR_12 TPR_19	N.I	TLH TRCD TR
	rap 6	371	NrfG TPR_12	N.I	TLH TR
	rap 7	81	N.I	N.I	N.I
	rap 8	365	NrfG TPR_12	N.I	TLH TR
Bt MC 28	rap 1	364	NrfG TPR_12	N.I	TLH TRCD
	rap 2	364	NrfG TPR_12	N.I	TLH
	rap 3	364	NrfG TPR_12 TPR_11	TPR1	TLH TR TPR1
	rapD 4	362	NrfG TPR_12	TPR12	TLH TR
	rap 5	410	NrfG TPR_12	N.I	TLH
	rap 6	370	NrfG TPR_12	N.I	TLH TRCD
	rap 7	376	NrfG TPR_12	N.I	TLH TR
	rap 8	364	TPR_12	N.I	TLH TRCD
	rap 9	378	NrfG TPR_12	N.I	TLH
Bt YBT-1518	rap 1	367	NrfG TPR_12	N.I	TLH
	rap 2	369	TPR_12	N.I	TLH
	rap 3	364	NrfG TPR_12	N.I	TLH
	rap 4	364	NrfG TPR_12	N.I	TLH TR
	rap 5	364	NrfG TPR_12	TPR12	TLH TRCD TR
	rap 6	72	N.I	N.I	N.I
	rap 7	297	NrfG TPR_12	TPR1	TLH TPR1
Bt chinensis CT-43	rap 1	369	TPR_12	N.I	TLH
	rap 2	364	NrfG TPR_12	N.I	TLH
	rap 3	364	NrfG TPR_12	N.I	TLH TR
	rap 4	364	NrfG TPR_12	TPR12	TLH TRCD TR
	rap 5	72	N.I	N.I	N.I
	rap 6	394	NrfG TPR_12/_19/_11	TPR2	TLH TRCD TRP2
	rap 7	364	NrfG TPR_12	N.I	TLH
	rap 8	357	NrfG TPR_12 TPR_11	TPR12	TLH
	rap 9	363	NrfG TPR_12	N.I	TLH TRCD TR
	rap 10	371	NrfG TPR_12 TPR_11	TPR1	TLH TPR1
	rap 11	351	NrfG TPR_12 TPR_11	TPR1	TLH TRCD TR TPR1
	rap 12	370	NrfG TPR_12 TPR_11	N.I	TLH TR
	rap 13	359	NrfG TPR_12	N.I	TLH
Bt finitimus YBT-020	rap 1	370	NrfG TPR_12 TPR_11	N.I	TLH TRCD TR
	rap 2	364	NrfG TPR_12	TPR12/TPR8	TLH TRCD TR
	rap 3	364	NrfG TPR_12	N.I	TLH TRCD
Bt kurstaki	rap 1	362	NrfG TPR_12/11/16/19	TPR11	TLH TRCD TR

HD-73	rap 2	99	NrfG TPR_12	N.I	N.I
	rap 3	357	NrfG TPR_12 TPR_11	TPR11	TLH TRCD
	rap 4	364	NrfG TPR_12 TPR_19	N.I	TLH TRCD TR
	rap 5	371	NrfG TPR_12	N.I	TLH TR
	rap 6	157	NrfG TPR_12 TPR_11	TPR11	TLH
	rap 7	371	NrfG TPR_12	N.I	TLH TRCD TR
	rapD 8	357	NrfG TPR_12 TPR_11	TPR12	TLH TR
	rap 9	366	NrfG TPR_12/_11/_19	TPR2	TLH TRCD TPR2
	rap 10	359	NrfG TPR_12	N.I	TLH
	Bt thuringiensis IS 5056	rap 1	369	TPR_12	N.I
rap 2		364	NrfG TPR_12	N.I	TLH
rap 3		364	NrfG TPR_12	N.I	TLH TR
rap 4		364	NrfG TPR_12	TPR12	TLH TRCD TR
rap 5		68	N.I	N.I	N.I
rap 6		371	NrfG TPR_12	TPR1	TLH TPR1
rap 7		366	NrfG TPR_12/_11/_19	TPR2	TLH TRCD TPR2
rap 8		366	NrfG TPR_12 TPR_11	N.I	TLH TRCD
rap 9		364	TPR_12	N.I	TLH TRCD
rap 10		382	NrfG TPR_12	N.I	TLH TRCD
rap 11		364	NrfG TPR_12	N.I	TLH
rapD 12		357	NrfG TPR_12 TPR_11	TPR12	TLH
rap 13		363	NrfG TPR_12	N.I	TLH TRCD TR
rap 14		363	NrfG TPR_12 TPR_11	N.I	TLH TRCD
rap 15		371	NrfG TPR_12	TPR1	TLH TR TPR1
rap 16		364	NrfG TPR_12 TPR_11	TPR1	TLH TRCD TR TPR1
rap 17		370	NrfG TPR_12	N.I	TLH TR
Bt Al Hakam	rap 1	390	TPR_12	N.I	TLH
	rap 2	381	NrfG TPR_12 TPR_11	TPR1	TLH TRCD TR TPR1
	rap 3	370	NrfG TPR_12	N.I	TLH TRCD
	rap 4	376	NrfG TPR_12	N.I	TLH
	rap 5	382	NrfG TPR_12	TPR12/TPR8	TLH TRCD TR
Ba Ames	rap 1	370	NrfG TPR_12 TPR_11	N.I	TLH TRCD TR
	rap 2	366	NrfG TPR_12	N.I	TLH TRCD
	rap 3	371	NrfG TPR_12	TPR1	TLH TR TPR1
Ba Sterne	rap 1	370	NrfG TPR_12 TPR_11	N.I	TLH TRCD TR
	rap 2	370	NrfG TPR_12	N.I	TLH TRCD
	rap 3	371	NrfG TPR_12	TPR1	TLH TR TPR1
Ba 'Ames Ancestor'	rap 1	370	NrfG TPR_12 TPR_11	N.I	TLH TRCD TR
	rap 2	366	NrfG TPR_12	N.I	TLH TRCD
	rap 3	371	NrfG TPR_12	TPR1	TLH TR TPR1
Ba A0248	rap 1	370	NrfG TPR_12 TPR_11	N.I	TLH TRCD TR
	rap 2	366	NrfG TPR_12	N.I	TLH TRCD
	rap 3	371	NrfG TPR_12	TPR1	TLH TR TPR1
Ba A16	rap 1	370	NrfG TPR_12 TPR_11	N.I	TLH TRCD TR
	rap 2	366	NrfG TPR_12	N.I	TLH TRCD
	rap 3	371	NrfG TPR_12	TPR1	TLH TR TPR1
Ba A16R	rap 1	370	NrfG TPR_12 TPR_11	N.I	TLH TRCD TR
	rap 2	366	NrfG TPR_12	N.I	TLH TRCD
	rap 3	371	NrfG TPR_12	TPR1	TLH TR TPR1
Ba CDC684	rap 1	371	NrfG TPR_12	TPR1	TLH TR TPR1
	rap 2	366	NrfG TPR_12	N.I	TLH TRCD
	rap 3	370	NrfG TPR_12	N.I	TLH TRCD TR
Ba H9401	rap 1	370	NrfG TPR_12 TPR_11	N.I	TLH TRCD TR
	rap 2	370	NrfG TPR_12	N.I	TLH TRCD
	rap 3	350	NrfG TPR_12	TPR1	TLH TR TPR1
	rap 4	114	N.I	N.I	N.I
Ba SVA 11	rap 1	370	NrfG TPR_12 TPR_11	N.I	TLH TRCD TR
	rap 2	46	N.I	N.I	N.I
	rap 3	328	NrfG TPR_12	N.I	TLH TRCD
	rap 4	350	NrfG TPR_12	TPR1	TLH TR TPR1
	rap 5	114	N.I	N.I	N.I

Bc: *Bacillus cereus*, Bt: *B. thuringiensis*, Ba: *B. anthracis*; rap: genes rap localizados por 'rap', rap: genes rap encontrados como 'response regulator aspartate phosphatase', rap: genes rap encontrados por similaridade no BLAST; NI: não identificado, TPR: tetratricopeptide repeat domain, NrfG: TPR repeat, TLH: tetratricopeptide-like helical domain, TRCD: tetratricopeptide repeat-containing domain, TR: tetratricopeptide repeat, N.I: não identificado.

Quadro 1 – Resultado do BLASTp entre as proteínas Rap de *Bacillus subtilis* str.168

<i>query</i> <i>subject</i>	Rap A	Rap B	Rap C	Rap D	Rap E	Rap F	Rap G	Rap H	Rap I	Rap J	Rap K
Rap A		98% cob 51% ident	98% cob 45% ident	95% cob 25% ident	98% cob 47% ident	99% cob 42% ident	98% cob 26% ident	98% cob 48% ident	98% cob 37% ident	97% cob 44% ident	96% cob 28% ident
Rap B	98% cob 51% ident		100% cob 43% ident	90% cob 25% ident	98% cob 40% ident	98% cob 43% ident	97% cob 30% ident	98% cob 46% ident	98% cob 39% ident	96% cob 43% ident	94% cob 29% ident
Rap C	98% cob 45% ident	99% cob 43% ident		91% cob 27% ident	98% cob 41% ident	99% cob 57% ident	97% cob 25% ident	98% cob 41% ident	96% cob 38% ident	95% cob 41% ident	97% cob 27% ident
Rap D	95% cob 25% ident	97% cob 25% ident	96% cob 27% ident		93% cob 27% ident	96% cob 26% ident	78% cob 26% ident	64% cob 25% ident	95% cob 23% ident	99% cob 25% ident	78% cob 26% ident
Rap E	99% cob 47% ident	99% cob 40% ident	99% cob 41% ident	86% cob 27% ident		99% cob 38% ident	75% cob 28% ident	98% cob 42% ident	97% cob 36% ident	97% cob 39% ident	98% cob 27% ident
Rap F	99% cob 42% ident	98% cob 43% ident	99% cob 57% ident	91% cob 26% ident	98% cob 38% ident		96% cob 26% ident	99% cob 40% ident	95% cob 39% ident	96% cob 42% ident	96% cob 26% ident
Rap G	100% cob 26% ident	99% cob 30% ident	98% cob 25% ident	87% cob 26% ident	98% cob 28% ident	97% cob 26% ident		99% cob 26% ident	99% cob 27% ident	97% cob 28% ident	100% cob 42% ident
Rap H	99% cob 48% ident	98% cob 46% ident	99% cob 41% ident	66% cob 25% ident	98% cob 42% ident	99% cob 40% ident	98% cob 26% ident		98% cob 35% ident	97% cob 41% ident	98% cob 27% ident
Rap I	95% cob 37% ident	96% cob 39% ident	93% cob 38% ident	89% cob 23% ident	93% cob 36% ident	92% cob 39% ident	94% cob 27% ident	93% cob 35% ident		93% cob 41% ident	94% cob 30% ident
Rap J	98% cob 44%	97% cob 43%	97% cob 41%	96% cob 25%	97% cob 39%	97% cob 42%	97% cob 28%	98% cob 41%	98% cob 41%		98% cob 28%

	ident	ident	ident	ident	ident	ident	ident	ident	ident	ident	ident
Rap K	97% cob 28% ident	95% cob 29% ident	98% cob 27% ident	79% cob 26% ident	99% cob 27% ident	98% cob 26% ident	99% cob 42% ident	98% cob 27% ident	99% cob 30% ident	95% cob 28% ident	

Cob.: cobertura, ident.: identidade.

Quadro 2 – Resultado do BLASTp entre as proteínas Rap de *Bacillus cereus* ATCC 14579

<i>query</i> <i>subject</i>	Rap 1	Rap 2	Rap 3	Rap 4	Rap 5
Rap 1		98% cob 47% ident	100% cob 42% ident	100% cob 47% ident	100% cob 53% ident
Rap 2	98% cob 47% ident		98% cob 39% ident	99% cob 44% ident	99% cob 48% ident
Rap 3	100% cob 42% ident	99% cob 39% ident		100% cob 41% ident	100% cob 40% ident
Rap 4	99% cob 47% ident	99% cob 44% ident	99% cob 41% ident		99% cob 44% ident
Rap 5	100% cob 53% ident	99% cob 48% ident	100% cob 40% ident	100% cob 44% ident	

Cob.: cobertura, ident.: identidade; Numeração do Rap é correspondente à identificada na Tabela 2

Quadro 3 – Resultado do BLASTp entre as proteínas Rap de *Bacillus thuringiensis* kurstaki HD-73

<i>query</i> <i>subject</i>	Rap 1	Rap 3	Rap 4	Rap 5	Rap 7	Rap 8	Rap 9	Rap 10
Rap 1		100% cob 38% ident	100% cob 44% ident	100% cob 44% ident	99% cob 47% ident	100% cob 43% ident	100% cob 43% ident	97% cob 42% ident
Rap 3	100% cob 38% ident		100% cob 40% ident	100% cob 40% ident	99% cob 41% ident	100% cob 52% ident	100% cob 40% ident	98% cob 46% ident
Rap 4	100% cob 44% ident	100% cob 40% ident		100% cob 44% ident	99% cob 45% ident	100% cob 39% ident	100% cob 41% ident	100% cob 45% ident
Rap 5	100% cob 44% ident	100% cob 40% ident	100% cob 44% ident		99% cob 54% ident	100% cob 43% ident	100% cob 43% ident	100% cob 43% ident
Rap 7	100% cob 47% ident	99% cob 41% ident	99% cob 45% ident	99% cob 54% ident		99% cob 45% ident	99% cob 49% ident	99% cob 45% ident
Rap 8	100% cob 43% ident	100% cob 52% ident	100% cob 39% ident	100% cob 43% ident	99% cob 45% ident		100% cob 42% ident	100% cob 49% ident
Rap 9	100% cob 43% ident	100% cob 40% ident	100% cob 41% ident	100% cob 43% ident	99% cob 49% ident	100% cob 42% ident		100% cob 45% ident
Rap 10	98% cob 42% ident	98% cob 46% ident	100% cob 45% ident	100% cob 43% ident	99% cob 45% ident	100% cob 49% ident	100% cob 45% ident	

Cob.: cobertura, ident.: identidade; Numeração do Rap é correspondente à identificada na Tabela 2.

Quadro 4 – Resultado do BLASTp entre os peptídeos Phr de *B. subtilis* str. 168.

<i>subject</i> <i>query</i>	PhrC	PhrI	PhrH	PhrA	PhrK	PhrE	PhrF	PhrG
PhrC		X	X	85% cob. $2e^{-5}$ 29% ident.	X	37% cob. 0,007 60% ident.	97% cob. 0,003 49% ident.	57% cob. 0,18 26% ident.
PhrI	X		X	X	X	X	X	79% cob. $1e^{-5}$ 42% ident.
PhrH	X	X		X	22% cob. 0,020 31% ident.	X	X	X
PhrA	43% cob. $7e^{-5}$ 42% ident.	X	X		X	X	X	X
PhrK	X	X	32% cob. 0,020 31% ident.	X		X	X	97% cob. $2e^{-7}$ 36% ident.
PhrE	15% cob. 0,014 71% ident.	X	X	X	X		X	X
PhrF	97% cob. $1e^{-7}$ 49% ident.	X	X	X	X	X		X
PhrG	15% cob. 3,0 50% ident.	76% cob. $1e^{-5}$ 42%	X	X	100% cob. $2e^{-7}$ 36% ident.	X	X	

Cob.: cobertura; ident.: identidade; X: nenhuma similaridade

Quadro 5 – Resultado do BLASTp entre os peptídeos Phr de *B. thuringiensis* HD-73.

<i>subject</i> <i>query</i>	Phr7	Phr1	Phr3	Phr5	Phr #	Phr9	Phr8	Phr10
Phr7		68% cob. 0,001 33% ident.	23% cob. 0,11 31% ident.	91% cob. 0,002 30% ident.	11% cob. 4,1 50% ident.	18% cob. $5e^{-4}$ 46% ident.	X	X
Phr1	66% cob. 0,001 33% ident.		X	84% cob. $5e^{-6}$ 36% ident.	X	65% cob. $2e^{-4}$ 31% ident.	X	38% cob. $2e^{-4}$ 31% ident.
Phr3	31% cob. 0,11 31% ident.	X		X	X	X	100% cob. $2e^{-5}$ 27% ident.	X
Phr5	94% cob. 0,002 30% ident.	98% cob. $5e^{-6}$ 36% ident.	X		X	X	X	62% cob. $3e^{-4}$ 39%
Phr #	8% cob. 4,1 50% ident.	X	X	X		47% cob. 0,016 30% ident.	X	X
Phr9	25% cob. $5e^{-4}$ 46% ident.	86% cob. $2e^{-4}$ 31% ident.	X	X	57% cob. 0,016 30% ident.		X	X
Phr8	X	X	100% cob. $2e^{-5}$ 27% ident.	X	X	X		100% cob. $4e^{-6}$ 46% ident.
Phr10	X	50% cob. $2e^{-4}$ 42% ident.	X	60% cob. $3e^{-4}$ 39% ident.	X	X	97% cob. $4e^{-6}$ 46% ident.	

#: Phr não ligado a um Rap; cob.: cobertura; ident.: identidade; X: nenhuma similaridade; Numeração do Phr é correspondente à identificada na Tabela 2.

– Agrupamentos das proteínas Rap por similaridade.

Grupo 1 (25)	Grupo 2 (22)	Grupo 3 (15)	Grupo 4 (14)	Grupo 5 (12)
3) Bc 03BB102 2) Bc AH820 2) Bc E33L 3) Bc F837/76 5) Bc FRI-35 4) Bc G9842 2) Bc CI 3) Bt 97-27 4) Bt HD-771 4) Bt HD-789 6) Bt MC28 p 6) Bt CT-43 p 9) Bt HD-73 p 7) Bt IS5056 p 8) Bt IS5056 p 3) Bt Al Hakam 2) Ba Ames 2) Ba Sterne 2) Ba A. Ancestor 2) Ba A0248 2) Ba A16 2) Ba A16R 2) Ba CDC684 2) Ba H9401 3) Ba SVA11	5) Bc ATCC 14579 4) Bc 03BB102 3) Bc AH820 4) Bc B4264 4) Bc F837/76 6) Bc G9842 6) Bt BMB171 2) Bt HD-771 6) Bt HD-789 7) Bt MC28 p 5) Bt HD-73 15) Bt IS5056 p 4) Bt Al Hakam 3) Ba Ames 3) Ba Sterne 3) Ba A. Ancestor 3) Ba A0248 3) Ba A16 3) Ba A16R 1) Ba CDC684 3) Ba H9401 4) Ba SVA11	2) Bc ATCC 14579 1) Bt YBT-020 1) Ba Ames 1) Ba Sterne 1) Ba A. Ancestor 1) Ba A0248 1) Ba A16 1) Ba A16R 3) Ba CDC684 1) Ba H9401 1) Ba SVA11 8) Bt 407 p 9) Bt HD-771 p 12) Bt CT-43 p 17) Bt IS5056 p	1) Bc E33L 1) Bt BMB171 2) Bt 407 2) Bt MC28 3) Bt YBT-1518 2) Bt CT-43 2) Bt IS5056 1) Bc B4264 5) Bt HD-771 1) Bc AH820 1) Bc CI 1) Bt HD-73 2) Bc F837/76 1) Bc FRI-35	2) Bc AH187 4) Bc AH820 2) Bc ATCC 10987 3) Bc B4264 3) Bc E33L 3) Bc FRI-35 2) Bc NC7401 4) Bt 407 3) Bt MC28 5) Bt YBT-1518 4) Bt CT-43 4) Bt IS5056
Grupo 6 (8)	Grupo 7 (7)	Grupo 8 (6)	Grupo 9 (6)	Grupo 10 (5)
1) Bc 03BB102 1) Bc ATCC 10987 1) Bc F837/76 1) Bt 407 2) Bt YBT-1518 1) Bt CT-43 1) Bt IS5056 1) Bt Al Hakam	2) Bc G9842 3) Bt BMB171 3) Bt 407 2) Bt HD-789 4) Bt YBT-1518 3) Bt CT-43 3) Bt IS5056	4) Bc ATCC 14579 5) Bc G9842 4) Bt 97-27 3) Bt HD-771 5) Bt HD-789 4) Bt HD-73	3) Bc ATCC 14579 3) Bc G9842 4) Bt BMB171 6) Bt HD-771 3) Bt HD-789 3) Bt HD-73	2) Bt BMB171 9) Bt 407 p 11) Bt CT-43 p 10) Bt IS5056 p 16) Bt IS5056 p
Grupo 11 (5)	Grupo 12 (4)	Grupo 13 (4)	Grupo 14 (4)	Grupo 15 (4)
4) Bc FRI-35 7) Bt HD-771 p 9) Bt CT-43 p 13) Bt IS5056 p 14) Bt IS5056 p	8) Bc G9842 p 2) Bt 97-27 7) Bt CT-43 p 11) Bt IS5056 p	5) Bt BMB171 4) Bc E33L 1) Bc Q1 3) Bt YBT-020	4) Bt MC28 p 8) Bt CT-43 p 8) Bt HD-73 p 12) Bt IS5056 p	5) Bc 03BB102 5) Bc F837/76 2) Bt YBT-020 5) Bt Al Hakam
Grupo 16 (3)	Grupo 17 (3)	Grupo 18 (3)	Grupo 19 (3)	Grupo 20 (3)
2) Bc 03BB102 1) Bt 97-27 2) Bt Al Hakam	3) Bc CI 5) Bt 97-27 2) Bc FRI-35	6) Bt 407 p 13) Bt CT-43 p 10) Bt HD-73 p	2) Bc B4264 8) Bt HD-771 p 9) Bt MC28 p	10) Bt CT-43 p 7) Bt HD-73 6) Bt IS5056 p
Grupo 21 (3)	Grupo 22 (2)	Grupo 23 (2)	Grupo 24 (2)	Grupo 25 (2)
1) Bc AH187 1) Bc NC7401 2) Bc Q1	5) Bc E33L p 8) Bt BMB171 p	1) Bc G9842 1) Bt HD-789	8) Bt MC28 p 9) Bt IS5056 p	1) Bc ATCC 14579 4) Bc Q1 p
Sem grupo				
3) Bc Q1 7) Bt 407 p		8) Bt HD-789 p 1) Bt MC28	5) Bt MC28 p 1) Bt YBT-1518	

A numeração corresponde a descrita na Tabela 2; o número entre parenteses na frente do número do grupo se refere ao número de componentes de cada grupo. O critério para o agrupamento foi alinhamento entre as sequências com e-value menor do que e^{-3} , e dos subgrupos foi identidade menor que 75%.