UNIVERSIDAD DE INVESTIGACIÓN DE TECNOLOGÍA EXPERIMENTAL YACHAY

Escuela de Ciencias Biológicas e Ingeniería

TÍTULO: Diversity and Evolutionary Dynamics of Spore Coat Proteins on Spore-forming Species of Bacillales

Trabajo de integración curricular presentado como requisito para la obtención del título de Biólogo

Autor:

Henrry Patricio Secaira Morocho

Tutor:

Ph.D. José Antonio Castillo

Urcuquí, marzo 2020

Urcuquí, 3 de marzo de 2020

SECRETARÍA GENERAL
(Vicerrectorado Académico/Cancillería)
ESCUELA DE CIENCIAS BIOLÓGICAS E INGENIERÍA
CARRERA DE BIOLOGÍA
ACTA DE DEFENSA No. UITEY-BIO-2020-00006-AD

En la ciudad de San Miguel de Urcuquí, Provincia de Imbabura, a los 3 días del mes de marzo de 2020, a las 10:00 horas, en el Aula CHA-01 de la Universidad de Investigación de Tecnología Experimental Yachay y ante el Tribunal Calificador, integrado por los docentes:

Presidente Tribunal de Defensa	Dr. GONZALES ZUBIATE, FERNANDO ALEXIS , Ph.D.
Miembro No Tutor	Dr. ALVAREZ BOTAS, FRANCISCO JAVIER , Ph.D.
Tutor	Dr. CASTILLO MORALES, JOSE ANTONIO , Ph.D.

Se presenta el(la) señor(ita) estudiante **SECAIRA MOROCHO, HENRRY PATRICIO**, con cédula de identidad No. **1719837443** , de la **ESCUELA DE CIENCIAS BIOLÓGICAS E INGENIERÍA**, de la Carrera de **BIOLOGÍA**, aprobada por el Consejo de Educación Superior (CES), mediante Resolución **RPC-SO-37-No.438-2014**, con el objeto de rendir la sustentación de su trabajo de titulación denominado: **Diversity and Evolutionary Dynamics of Spore Coat Proteins on Spore-forming Species of Bacillales**, previa a la obtención del título de **BIÓLOGO/A**.

El citado trabajo de titulación, fue debidamente aprobado por el(los) docente(s):

Tutor	Dr. CASTILLO MORALES, JOSE ANTONIO , Ph.D.

Y recibió las observaciones de los otros miembros del Tribunal Calificador, las mismas que han sido incorporadas por el(la) estudiante.

Previamente cumplidos los requisitos legales y reglamentarios, el trabajo de titulación fue sustentado por el(la) estudiante y examinado por los miembros del Tribunal Calificador. Escuchada la sustentación del trabajo de titulación, que integró la exposición de el(la) estudiante sobre el contenido de la misma y las preguntas formuladas por los miembros del Tribunal, se califica la sustentación del trabajo de titulación con las siguientes calificaciones:

Tipo	Docente	Calificación
Miembro Tribunal De Defensa	Dr. ALVAREZ BOTAS, FRANCISCO JAVIER , Ph.D.	9.7
Presidente Tribunal De Defensa	Dr. GONZALES ZUBIATE, FERNANDO ALEXIS , Ph.D.	10.0
Tutor	Dr. CASTILLO MORALES, JOSE ANTONIO , Ph.D.	10.0

Lo que da un promedio de: **9.9 (Nueve punto Nueve)**, sobre 10 (diez), equivalente a: **APROBADO**

Para constancia de lo actuado, firman los miembros del Tribunal Calificador, el/la estudiante y el/la secretario ad-hoc.

Henry Secaira
SECAIRA MOROCHO, HENRRY PATRICIO
Estudiante

Dr. GONZALES ZUBIATE, FERNANDO ALEXIS , Ph.D.
Presidente Tribunal de Defensa

Dr. CASTILLO MORALES, JOSE ANTONIO , Ph.D.
Tutor

UNIVERSIDAD YACHAY TECH

Dr. ALVAREZ BOTAS, FRANCISCO JAVIER , Ph.D.
Miembro No Tutor

ALARCON FELIX, KARLA ESTEFANIA
Secretario Ad-hoc

AUTORÍA

Yo, **Henrry Patricio Secaira Morocho**, con cédula de identidad 1719837443, declaro que las ideas, juicios, valoraciones, interpretaciones, consultas bibliográficas, definiciones y conceptualizaciones expuestas en el presente trabajo; así cómo, los procedimientos y herramientas utilizadas en la investigación, son de absoluta responsabilidad de el/la autora (a) del trabajo de integración curricular. Así mismo, me acojo a los reglamentos internos de la Universidad de Investigación de Tecnología Experimental Yachay.

Urcuquí, marzo de 2020.

Henry Secaira

Henrry Patricio Secaira Morocho
CI: 1719837443

AUTORIZACIÓN DE PUBLICACIÓN

Yo, **Henrry Patricio Secaira Morocho,** con cédula de identidad 1719837443, cedo a la Universidad de Tecnología Experimental Yachay, los derechos de publicación de la presente obra, sin que deba haber un reconocimiento económico por este concepto. Declaro además que el texto del presente trabajo de titulación no podrá ser cedido a ninguna empresa editorial para su publicación u otros fines, sin contar previamente con la autorización escrita de la Universidad.

Asimismo, autorizo a la Universidad que realice la digitalización y publicación de este trabajo de integración curricular en el repositorio virtual, de conformidad a lo dispuesto en el Art. 144 de la Ley Orgánica de Educación Superior

Urcuquí, marzo de 2020.

Henry Secaira

Henrry Patricio Secaira Morocho
CI: 1719837443

Agradecimiento

Me gustaría agradecer a mi mentor de tesis, el Dr. José Antionio Castillo, por su inestimable apoyo y orientación durante esta investigación y redacción de tesis.

Además, me gustaría agradecer a mis profesores en Yachay Tech por todas las clases, lecciones y consejos durante mi carrera.

Estoy profundamente agradecido con mis amigos, compañeros de clase y familia, especialmente con mi madre y su apoyo constante durante estos años.

Henrry Patricio Secaira Morocho

Resumen

Entre los miembros del orden de Bacillales, hay varias especies capaces de formar una estructura llamada endospora. Las endosporas permiten que las bacterias sobrevivan bajo condiciones de crecimiento desfavorables. Además, las endosporas promueven la germinación cuando las condiciones ambientales son favorables. Varias proteínas necesarias para el ensamblaje de la capa, la síntesis de la corteza y la germinación se conocen colectivamente como proteínas de la capa de esporas. Este proyecto tiene como objetivo determinar la diversidad y los procesos evolutivos de los genes de la capa de esporas en varias especies de Bacillales formadoras de esporas. Los métodos de BLASTp, Clustering y KEGG Orthology se han utilizado para determinar la existencia y diversidad de genes de la capa en ciento 141 genomas de especies formadoras de esporas. Además, las fuerzas de selección que actúan sobre los genes de la capa de esporas se han estimado utilizando los métodos Tajima's D, dN / dS, MEME y BUSTED. Por último, los genes de la capa de la espora sometidos a eventos de transferencia horizontal de genes (HGT) han sido identificados por la reconciliación filogenética entre árboles filogenéticos de especies y genes. Además, los resultados de HGT se han confirmado escaneando los genomas para encontrar rastros de secuencias de inserción, utilizando la base de datos ICEberg. Los resultados sugieren que los genes mejor conservados entre las diferentes especies son aquellos con los roles más importantes en el ensamblaje de la capa y permiten la adaptación específica a un nicho. Del mismo modo, los resultados de HGT confirman que la esporulación es una característica ancestral en *Bacillus*.

Palabras clave: proteínas de la capa de espora, Bacillales, *Bacillus*, proteínas morfogenéticas de la capa, selección positiva / negativa, transferencia horizontal de genes.

Abstract

Among members of the Bacillales order, there are several species capable of forming a structure called endospore. Endospores enable bacteria to survive under unfavorable growth conditions. Moreover, endospores promote germination when environmental conditions are favorable again. Several proteins necessary for coat assembly, cortex synthesis, and germination are collectively known as spore coat proteins. This project aims to determine the diversity and evolutionary processes of spore coat genes on various spore-forming species of Bacillales. The methods of BLASTp, Clustering, and KEGG Orthology have been used to determine the existence and diversity of coat genes across one hundred forty-one genomes of spore-forming species. Furthermore, selection pressures that act over spore coat genes have been estimated using the methods of Tajima's D, dN/dS ratio, MEME, and BUSTED. Finally, spore coat genes subjected to horizontal gene transfer (HGT) events have been identified by phylogenetic reconciliation between gene-species phylogenetic trees. Additionally, HGT results have been further confirmed by scanning the genomes to find traces of insertion sequences, using the ICEberg database. The results suggest that the best-conserved genes among different species are those with the most important roles in coat assembly and enable niche-specific adaptation. Likewise, HGT results confirm that sporulation is an ancestral feature in *Bacillus*.

Key Words:

Spore coat proteins, Bacillales, *Bacillus*, morphogenetic coat proteins, positive/negative selection, horizontal gene transfer.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

ABBREVIATIONS

HGT	Horizontal Gene Transfer
MEME	Mixed Effects Model of Evolution
BUSTED	Branch-Site Unrestricted Statistical Test for Episodic Diversification
dN	Non-synonymous mutations
dS	Synonymous mutations
GEIs	Genomic Islands

CHAPTER 1. INTRODUCTION

1.1 *Bacillales* and the spore

The Bacillales order has a great taxonomic and phylogenetic diversity and can thrive in all environments (Maayer, Aliyu, & Cowan, 2019). Some spore formers contribute to the human and mammal gut sporobiota (Paul et al., 2019; Suitso et al., 2010), while others are pathogens that cause foodborne diseases (Wells-Bennik et al., 2016) or important human pathogens (Barák, Ricca, & Cutting, 2005; Kotiranta, Lounatmaa, & Haapasalo, 2000; Mock & Fouet, 2001; Stenfors Arnesen, Fagerlund, & Granum, 2008). A striking feature of members of this order is their ability to form an extremely resistant cell type, the spore (Driks & Eichenberger, 2016; Maayer et al., 2019; Qin & Driks, 2013; Setlow, 2014b). The spore can survive a wide range of extreme environmental conditions, such as microbial predation, desiccation, heat, UV radiation, and toxic chemicals (Beladjal, Gheysens, Clegg, Amar, & Mertens, 2018; Driks & Eichenberger, 2016; Klobutcher, Ragkousi, & Setlow, 2006; Nicholson, Munakata, Horneck, Melosh, & Setlow, 2000; Setlow, 2014b). The spore is metabolically dormant and can remain in this state for hundreds of years (Qin & Driks, 2013; Setlow, Wang, & Li, 2017). In addition, the spore can sense its surrounding environment. When growth conditions are favorable again, they germinate to generate a vegetative form of bacteria (Moir & Cooper, 2015; Qin & Driks, 2013; Setlow, 2014a; Setlow et al., 2017).

1.2 Sporulation

To survive under such stress conditions, the bacterial cell undergoes an evolutionarily conserved process called sporulation to produce the spore (McKenney, Driks, & Eichenberger, 2013). Bacillus subtilis has been the most studied organism for this process (Abecasis et al., 2013; McKenney et al., 2013; Qin & Driks, 2013). Nevertheless, most processes of sporulation appear to be highly similar among all spore-forming species (Onyenwoke, Brill, Farahi,

& Wiegel, 2004). Studies from laboratories report that sporulation begins in the stationary phase when nutrients start to scarce (McKenney et al., 2013). Histidine sensor kinases (KinA, KinB, and KinC) trigger sporulation by phosphorylation of the master regulator of sporulation, Spo0A (McKenney et al., 2013). Phosphorylated Spo0A exert control over a large group of genes involved in asymmetric cell division and activation of the sporulation-specific sigma factors (McKenney et al., 2013; Molle et al., 2003). The sporulating cell undergoes an asymmetric cell division that produces a sporangium, which contains the larger mother cell and the smaller forespore (the future spore) (Driks & Eichenberger, 2016; McKenney et al., 2013). After that, the mother cell engulfs the forespore in a process similar to phagocytosis (Driks & Eichenberger, 2016; McKenney et al., 2013). After the engulfment, the forespore becomes a double membrane-bound cell within the mother cell (Driks & Eichenberger, 2016; McKenney et al., 2013). Coat assembly begins after the initiation of engulfment and continues through sporulation (Driks & Eichenberger, 2016; McKenney et al., 2013). The forespore then is composed of two external protective structures: the cortex, that is assembled between the inner and outer forespore membranes, and the proteinaceous coat (Henriques & Moran, 2007; McKenney et al., 2013). In the spore, the genome is contained in the partially dehydrated core (McKenney et al., 2013). In the final step, the mother cell releases the partially mature spore by lysis (McKenney et al., 2013). Maturation of the spore continues, specifically cross-linking of coat proteins, after the lysis of the cell mother (Abhyankar et al., 2015; Driks & Eichenberger, 2016; Ragkousi & Setlow, 2004).

1.3 Spore-related applications

Spores have several applications mostly based on different surface spore coat properties (Barák et al., 2005; McKenney et al., 2013). For example, spore-based therapies are used as oral vaccines to deliver probiotics to mucosal surfaces. These therapies exploit different surface properties to display heterologous antigens (Barák et al., 2005; Cutting, Hong, Baccigalupi, & Ricca, 2009; McKenney et al., 2013; Paul et al., 2019). Similarly, the surface properties can be used for the delivery of therapeutic agents, to produce biocatalysts, and biosorbents (McKenney

et al., 2013; Paul et al., 2019; Wu, Mulchandani, & Chen, 2008). Several Bacillus species are known as plant growth-promoting rhizobacteria, which are able to promote plant growth by either competing with plant pathogens or enhancing plant-microbe symbioses (Bhattacharyya & Jha, 2012; Glick, 2012; Paul et al., 2019).

1.4 Spore coat structure

There is a high diversity in spore coat morphology among spore-forming species (Driks & Eichenberger, 2016; McKenney et al., 2013). The spore coat structure has been extensively studied in B. subtilis by transmission electron microscopy (TEM) (Driks & Eichenberger, 2016). The innermost layer of the spore coat is called basement layer (Driks & Eichenberger, 2016). The basement layer contains spore coat proteins necessary to initiate coat assembly, SpoIVA, SpoVM, SpoVID (Driks & Eichenberger, 2016; McKenney et al., 2013). Then, the inner layer is located at the top of the basement layer (Driks & Eichenberger, 2016). The outer coat surrounds the inner coat, whereas the crust surrounds the outer (Driks & Eichenberger, 2016). Figure 1.1 shows the four layers (basement layer, inner layer, outer layer, crust) that compose the spore coat. Other spore-forming species, such as *Bacillus anthracis, Bacillus thuringiensis*, and *Bacillus cereus* also possess an exosporium (Driks & Eichenberger, 2016; McKenney et al., 2013; Qin & Driks, 2013; Waller, Fox, Fox, Fox, & Price, 2004). The exosporium is the outermost layer that surrounds the mature spore (Bozue, Welkos, & Cote, 2015). It is composed of fine hair-like projections that may be involved in infection for *B. anthracis* (Bozue et al., 2015).

1.5 Spore coat functions

The spore coat has various functions ranging from protection to interaction with the environment (Driks & Eichenberger, 2016). However, the exact mechanisms for resistance properties have not been elucidated completely (Driks & Eichenberger, 2016; Setlow, 2014b). Spore architecture has been described to influence the resistance of spores to extreme environmental conditions (McKenney et al., 2013). For example, the spore coat has a key role as

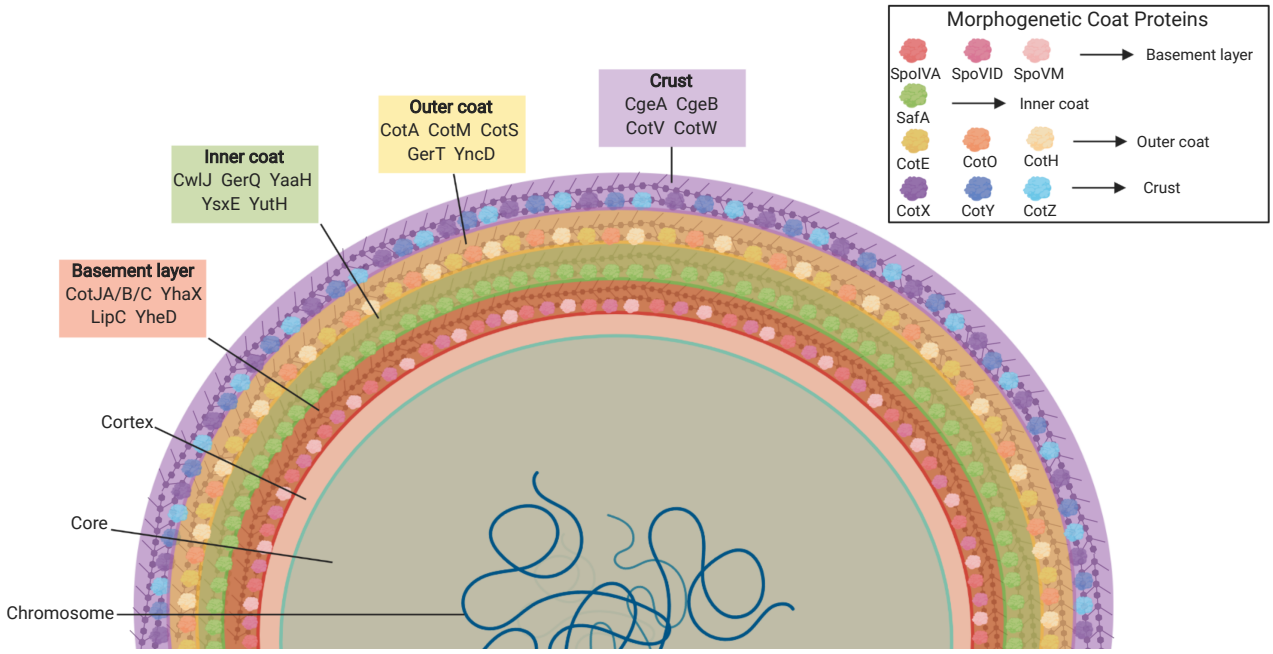


Figure 1.1. Model of spore coat structure. Assembly of each layer depend on the multimerization of a morphogenetic coat protein and its dependent individual coat proteins. Four layers with its morphogenetic and morphogenetic-dependent coat proteins are shown: basement layer (red), inner layer (green), outer layer (yellow), and crust (purple). Modified from McKenney et al. (2013).

a barrier in predation from bacteriovores, excluding degradative enzymes and toxic molecules (Driks & Eichenberger, 2016; Klobutcher et al., 2006; McKenney et al., 2013; Setlow, 2014b). However, since germination requires the passage of germinants (small molecules necessary for germination) through the coat, it has been suggested that the coat is not impermeable but rather porous (Driks & Eichenberger, 2016; Moir & Cooper, 2015). This porosity allows germinants, such as sugars, amino acids, peptidoglycan fragments, and ions to reach their receptors in the inner spore membrane, by passing through the spore coat (Driks & Eichenberger, 2016; McKenney et al., 2013; Setlow, 2014a).

Spores can resist extreme heat and desiccation by partial dehydration of the spore core, functional cortex, RecA-dependent DNA repair machinery, and various small acid-soluble proteins that bind to the spore core (McKenney et al., 2013; Setlow, 2007). Low water content

in the dormant spore core leads to reduced molecular mobility, which in turn confers elevated protein resistance to thermal inactivation (Setlow, 2014b).

1.6 Spore coat assembly and genetic interaction network

Spore coat synthesis, assembly, and maturation is a complex multiprotein process that requires several hours to be complete (Driks & Eichenberger, 2016). Assembly of coat layers depends on morphogenetic coat proteins, such as SpoIVA, SpoVM, SpoVID, CotE, CotH, CotO, CotX, CotY, CotZ, SafA, and coat proteins dependent on these morphogenetic proteins (Driks & Eichenberger, 2016; McKenney et al., 2013). SpoIVA and SpoVM are responsible for spore cortex formation, coat assembly, anchoring of the coat to the spore surface, and spore encasement, whereas, SpoVID is responsible for spore encasement (Driks & Eichenberger, 2016; McKenney et al., 2013; McKenney & Eichenberger, 2012). CotE is the most critical protein responsible for the assembly of the outer coat, and SafA is responsible for the assembly of the inner coat (Bauer, Little, Stöver, & Driks, 1999; Driks & Eichenberger, 2016; McKenney et al., 2013; Ozin, Henriques, Yi, & Moran, 2000; Zilhão et al., 1999). Several studies demonstrated the existence of a network of genetic interactions that consist of three independent modules: SpoIVA-dependent subnetwork, CotE-dependent subnetwork, and SafA-dependent subnetwork (Driks & Eichenberger, 2016; Krajčíková, Forgáč, Szabo, & Barák, 2017; McKenney et al., 2013, 2010), as shown in Figure 1.2

1.7 Problem statement

The spore coat structure has been extensively studied in *B. subtilis, B. anthracis*, and other model organisms by classical methods, such as TEM (Driks & Eichenberger, 2016). Nevertheless, using TEM, researchers cannot easily distinguish the different layers that compose the spore coat structure (Driks & Eichenberger, 2016). Besides, the spore coat is composed of several coat proteins localized in different layers (Driks & Eichenberger, 2016; McKenney et al., 2013). Previous studies have reported the existence of a genomic signature on a spore-forming species of Firmicutes (Abecasis et al., 2013; Galperin, 2013; Galperin et al., 2012). However,

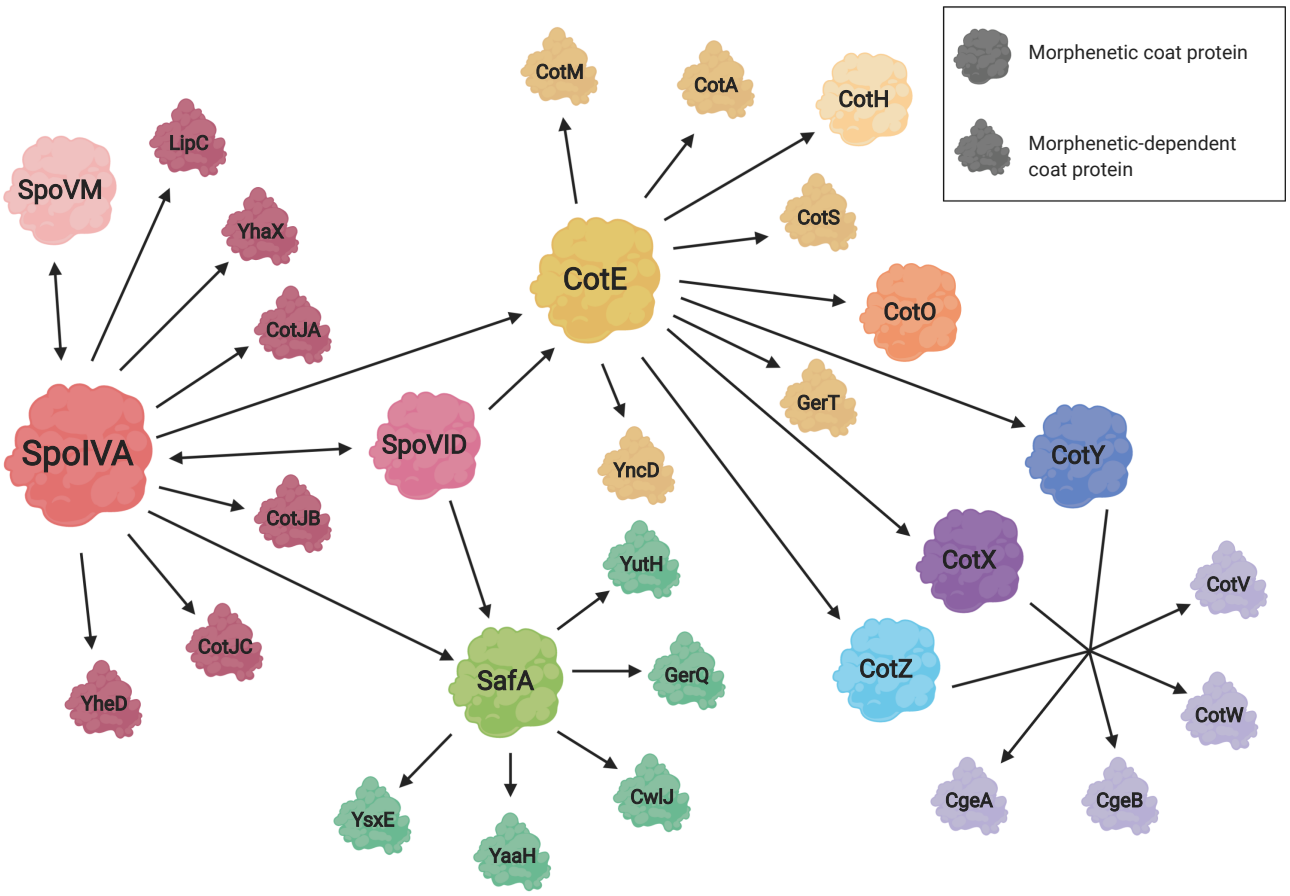


Figure 1.2. Spore coat protein interaction network. Morphogenetic and morphogenetic-dependent coat proteins interact with each other to form the four layers (basement layer, inner layer, outer layer, crust) of the spore coat. Recruitment of the morphogenetic coat proteins SafA and CotE depend on SpoIVA, whereas recruitment of CotO and CotX/Y/Z depend on CotE, the interaction network is highly hierarchical. SpoIVA, SpoVID, and SpoVM are required to initiate coat assembly and are at the top of the spore assembly hierarchy. Modified from McKenney et al. (2013).

the spore coat structure and its genomic signature of many non-model organisms, members of the Bacillales order, that can form spores are still unknown (Galperin, 2013; Galperin et al., 2012; Onyenwoke et al., 2004).

Despite the existence of more than eighty different spore coat proteins, studies have demonstrated that not all of them are required for coat synthesis, assembly, maturation, and

spore germination (Abecasis et al., 2013; Driks & Eichenberger, 2016; Galperin, 2013; Galperin et al., 2012; McKenney et al., 2013; Onyenwoke et al., 2004). Indeed, not all coat protein gene mutations are phenotypically silent, exceptions are morphogenetic coat proteins that control the assembly of other coat proteins (Driks & Eichenberger, 2016). Furthermore, little is known about the selective pressures and evolutionary histories acting upon those morphogenetic spore coat proteins and their dependent proteins among Bacillus species (Qin & Driks, 2013).

1.8 General and specific objectives

Determine the diversity and evolutionary dynamics of spore coat proteins among spore-forming species belonging to the *Bacillales* order.

- Identify which spore coat proteins of *Bacillus subtilis* are conserved among different *Bacillales* species, by employing three different methods: local BLASTp, Clustering, and KEGG Orthology
- Measure the degree of selection pressures acting on spore coat proteins in *Bacillus* species, by employing five summary statistics: Tajima's D, BUSTED, MEME, dN/dS (branch and site models)
- Analyze HGT events of spore coat proteins across *Bacillus* species

CHAPTER 2. METHODS

2.1 Sequence data and diversity analyses of spore coat genes in spore-forming species of *Bacillales*

Based on a thorough literature review, we identified ninety-seven genes that encode spore coat proteins in the *Bacillales* order, see Table 2.1. Each gene sequence was downloaded from the SubtiWiki server (`http://subtiwiki.uni-goettingen.de/`) (Zhu & Stülke, 2018) in October and November 2018, and they were located in the annotated genome of *Bacillus subtilis* 168 (NC_000964). One hundred and forty-one genomes of Bacillales order were retrieved from the NCBI's FTP server (`https://www.ncbi.nlm.nih.gov/genome/microbes/`). We selected forty genomes of the genus *Bacillus* and one hundred and one genomes of non-Bacillus genera belonging to *Bacillales*, see Table 2.2. All the genomes were downloaded as FASTA (.fasta, .fna, .fas) and GENBANK (.gbff, .gb, .gbk) formats for all the subsequent analyses.

Table 2.1 Ninety-seven spore coat genes identified in *Bacillales* and their location in the genome of the model organism *B. subtilis* 168

Spore coat gene	**Locus Tag**	**References**
cgeA	BSU_19780	(Driks & Eichenberger, 2016)
cgeB	BSU_19790	(Driks & Eichenberger, 2016)
cgeC	BSU_19770	(Driks & Eichenberger, 2016)
cgeD	BSU_19760	(Driks & Eichenberger, 2016)
cgeE	BSU_19750	(Driks & Eichenberger, 2016)
cotA	BSU_06300	(Driks & Eichenberger, 2016) (Galperin et al., 2012)
cotB	BSU_36050	(Driks & Eichenberger, 2016) (Galperin et al., 2012)
cotC	BSU_17700	(Driks & Eichenberger, 2016) (Galperin et al., 2012)

Table 2.1 Ninety-seven spore coat genes identified in *Bacillales* and their location in the genome of the model organism *B. subtilis* 168. (Continued)

Spore coat gene	**Locus Tag**	**References**
cotD	BSU_22200	(Driks & Eichenberger, 2016) (Galperin et al., 2012)
cotE	BSU_17030	(Driks & Eichenberger, 2016) (Galperin et al., 2012)
cotF	BSU_40530	(Driks & Eichenberger, 2016) (Galperin et al., 2012) (Tirumalai et al., 2013)
cotG	BSU_36070	(Driks & Eichenberger, 2016)
cotH	BSU_36060	(Driks & Eichenberger, 2016) (Galperin et al., 2012)
cotI	BSU_30920	(Driks & Eichenberger, 2016) (Galperin et al., 2012)
cotJA	BSU_06890	(Driks & Eichenberger, 2016) (Galperin et al., 2012) (Tirumalai et al., 2013)
cotJB	BSU_06900	(Driks & Eichenberger, 2016) (Galperin et al., 2012) (Tirumalai et al., 2013)
cotJC	BSU_06910	(Driks & Eichenberger, 2016) (Galperin et al., 2012) (Tirumalai et al., 2013)
cotK/sspO	BSU_17990	(Watabe, 2013)
cotM	BSU_17970	(Driks & Eichenberger, 2016) (Galperin et al., 2012)
cotN/tasA	BSU_24620	(Galperin et al., 2012)

Table 2.1 Ninety-seven spore coat genes identified in *Bacillales* and their location in the genome of the model organism *B. subtilis* 168. (Continued)

Spore coat gene	**Locus Tag**	**References**
cotO	BSU_11730	(Driks & Eichenberger, 2016) (Galperin et al., 2012)
cotP	BSU_05550	(Driks & Eichenberger, 2016) (Galperin et al., 2012)
cotQ	BSU_34520	(Driks & Eichenberger, 2016)
cotR	BSU_34530	(Driks & Eichenberger, 2016)
cotS	BSU_30900	(Driks & Eichenberger, 2016) (Galperin et al., 2012) (Watabe, 2013)
cotSA	BSU_30910	(Driks & Eichenberger, 2016) (Galperin et al., 2012) (Watabe, 2013)
cotT	BSU_12090	(Driks & Eichenberger, 2016)
cotU	BSU_17670	(Driks & Eichenberger, 2016) (Galperin et al., 2012)
cotV	BSU_11780	(Driks & Eichenberger, 2016) (Watabe, 2013) (Imamura, Kuwana, Takamatsu, & Watabe, 2011)
cotW	BSU_11770	(Driks & Eichenberger, 2016) (Watabe, 2013) (Imamura et al., 2011)
cotX	BSU_11760	(Driks & Eichenberger, 2016)
cotY	BSU_11750	(Driks & Eichenberger, 2016) (Galperin et al., 2012) (Watabe, 2013)

Table 2.1 Ninety-seven spore coat genes identified in *Bacillales* and their location in the genome of the model organism *B. subtilis* 168. (Continued)

Spore coat gene	**Locus Tag**	**References**
cotZ	BSU_11740	(Driks & Eichenberger, 2016) (Galperin et al., 2012) (Watabe, 2013)
cwlJ	BSU_02600	(Driks & Eichenberger, 2016)
gerPA	BSU_10720	(Driks & Eichenberger, 2016)
gerPB	BSU_10710	(Driks & Eichenberger, 2016)
gerPC	BSU_10700	(Driks & Eichenberger, 2016)
gerPD	BSU_10690	(Driks & Eichenberger, 2016)
gerPE	BSU_10680	(Driks & Eichenberger, 2016)
gerPF	BSU_10670	(Driks & Eichenberger, 2016)
gerQ	BSU_37920	(Driks & Eichenberger, 2016)
gerT	BSU_19490	(Driks & Eichenberger, 2016)
lipC	BSU_04110	(Driks & Eichenberger, 2016) (Galperin et al., 2012)
oxdD	BSU_18670	(Driks & Eichenberger, 2016)
safA	BSU_27840	(Driks & Eichenberger, 2016) (Galperin et al., 2012) (Watabe, 2013)
spoIVA	BSU_22800	(Driks & Eichenberger, 2016) (Galperin et al., 2012)
spoVID	BSU_28110	(Driks & Eichenberger, 2016) (Galperin et al., 2012)
spoVM	BSU_15810	(Driks & Eichenberger, 2016) (Galperin et al., 2012)

Table 2.1 Ninety-seven spore coat genes identified in *Bacillales* and their location in the genome of the model organism *B. subtilis* 168. (Continued)

Spore coat gene	**Locus Tag**	**References**
spsB	BSU_37900	(Driks & Eichenberger, 2016) (Cangiano et al., 2014)
spsI	BSU_37810	(Driks & Eichenberger, 2016)
sscA	BSU_09958	(Driks & Eichenberger, 2016)
tgl	BSU_31270	(Driks & Eichenberger, 2016) (Galperin et al., 2012)
yaaH	BSU_00160	(Driks & Eichenberger, 2016) (Galperin et al., 2012) (Watabe, 2013) (Kodama et al., 1999)
ydgA	BSU_05560	(Driks & Eichenberger, 2016) (Galperin et al., 2012)
ydgB	BSU_05570	(Driks & Eichenberger, 2016) (Galperin et al., 2012)
ydhD	BSU_05710	(Driks & Eichenberger, 2016) (Galperin et al., 2012) (Watabe, 2013) (Kodama et al., 1999)
yhaX	BSU_09830	(Driks & Eichenberger, 2016) (Galperin et al., 2012)
yhbA/queG	BSU_08910	(Galperin et al., 2012)
yhbB	BSU_08920	(Driks & Eichenberger, 2016) (Galperin et al., 2012)
yhcN	BSU_09150	(Galperin et al., 2012)
yhcQ	BSU_09180	(Galperin et al., 2012)

Table 2.1 Ninety-seven spore coat genes identified in *Bacillales* and their location in the genome of the model organism *B. subtilis* 168. (Continued)

Spore coat gene	**Locus Tag**	**References**
yheC	BSU_09780	(Driks & Eichenberger, 2016)
yheD	BSU_09770	(Driks & Eichenberger, 2016)
yhjQ	BSU_10600	(Driks & Eichenberger, 2016) (Galperin et al., 2012)
yhjR	BSU_10610	(Driks & Eichenberger, 2016) (Galperin et al., 2012)
yisY	BSU_10900	(Driks & Eichenberger, 2016) (Galperin et al., 2012) (Tirumalai et al., 2013)
yjcB	BSU_11800	(Watabe, 2013)
yjqC	BSU_12490	(Driks & Eichenberger, 2016) (Galperin et al., 2012)
yjzB	BSU_11320	(Driks & Eichenberger, 2016)
yknT	BSU_14250	(Driks & Eichenberger, 2016) (Galperin et al., 2012)
ykvP	BSU_13780	(Driks & Eichenberger, 2016)
ykvQ	BSU_13790	(Driks & Eichenberger, 2016)
ykzQ	BSU_13789	(Driks & Eichenberger, 2016)
ylbD	BSU_14970	(Galperin et al., 2012)
ylbO/gerR	BSU_15090	(Watabe, 2013)
ymaG	BSU_17310	(Driks & Eichenberger, 2016) (Galperin et al., 2012)
yncD	BSU_17640	(Driks & Eichenberger, 2016) (Galperin et al., 2012)
ypeP	BSU_21970	(Galperin et al., 2012)

Table 2.1 Ninety-seven spore coat genes identified in *Bacillales* and their location in the genome of the model organism *B. subtilis* 168. (Continued)

Spore coat gene	**Locus Tag**	**References**
yppG	BSU_22250	(Driks & Eichenberger, 2016) (Galperin et al., 2012)
ypzA	BSU_21950	(Galperin et al., 2012)
yraD	BSU_26990	(Driks & Eichenberger, 2016) (Galperin et al., 2012) (Tirumalai et al., 2013)
yraF	BSU_26960	(Driks & Eichenberger, 2016) (Galperin et al., 2012) (Tirumalai et al., 2013)
yraG	BSU_26950	(Tirumalai et al., 2013)
yrbB/coxA	BSU_27830	(Watabe, 2013)
yrbC	BSU_27820	(Watabe, 2013)
ysnD	BSU_28320	(Driks & Eichenberger, 2016)
ysxE	BSU_28100	(Driks & Eichenberger, 2016) (Galperin et al., 2012)
ytdA	BSU_30850	(Driks & Eichenberger, 2016)
ytxO	BSU_30890	(Driks & Eichenberger, 2016)
yusN	BSU_32860	(Driks & Eichenberger, 2016) (Galperin et al., 2012)
yutH	BSU_32270	(Driks & Eichenberger, 2016) (Galperin et al., 2012)
yuzC	BSU_31730	(Driks & Eichenberger, 2016)
ywqH	BSU_36210	(Tirumalai et al., 2013)
ywrJ	BSU_36040	(Driks & Eichenberger, 2016) (Galperin et al., 2012)

Table 2.1 Ninety-seven spore coat genes identified in *Bacillales* and their location in the genome of the model organism *B. subtilis* 168. (Continued)

Spore coat gene	**Locus Tag**	**References**
		(Driks & Eichenberger, 2016)
yxeE	BSU_39580	(Galperin et al., 2012)
		(Watabe, 2013)
yybI	BSU_40630	(Driks & Eichenberger, 2016)
		(Driks & Eichenberger, 2016)
yeek	BSU_06850	(Galperin et al., 2012)
		(Watabe, 2013)

Table 2.2 Genomes of *Bacillales* retrieved from NCBIs FTP server

Genome	**Reference***
Bacillus	
Bacillus subtilis 168	NC_000964
Bacillus altitudinis W3	NZ_CP011150
Bacillus amyloliquefaciens DSM7	NC_014551
Bacillus anthracis Ames Ancestor	NC_007530
Bacillus aryabhattai B8W22	NZ_JYOO01000001
Bacillus atrophaeus SRCM101359	NZ_CP021500
Bacillus beveridgei MLTeJB	NZ_CP012502
Bacillus bombysepticus Wang	NZ_CP007512
Bacillus cellulosilyticus DSM2522	NC_014829
Bacillus cereus ATCC14579	NC_004722
Bacillus clausii KSM-K16	NC_006582
Bacillus coagulans DSM1 ATCC7050	NZ_ATUM01000001
Bacillus cohnii DSM6307	NZ_CP018866
Bacillus cytotoxicus NVH391-98	NC_009674

Note. *NCBI accession number

Table 2.2 Genomes of *Bacillales* retrieved from NCBIs FTP server (Continued)

Genome	**Reference***
Bacillus endophyticus DSM13796	NZ_FOXX01000035
Bacillus flexus KLBMP4941	NZ_CP016790
Bacillus gibsonii FJAT-10019	NZ_CP017070
Bacillus glycinifermentans BGLY	NZ_LT603683
Bacillus halodurans C-125	NC_002570
Bacillus horikoshii 20a	NZ_CP020880
Bacillus infantis NRRL B-14911	NC_022524
Bacillus krulwichiae AM31D	NZ_CP020814
Bacillus lehensis G1	NZ_CP003923
Bacillus licheniformis ATCC14580	NC_006270
Bacillus megaterium NBRC15308 ATCC14581	NZ_CP009920
Bacillus methanolicus MGA3	NZ_CP007739
Bacillus muralis G25-68	NZ_CP017080
Bacillus mycoides ATCC6462	NZ_CP009692
Bacillus oceanisediminis 2691	NZ_CP015506
Bacillus paralicheniformis ATCC9945a	NC_021362
Bacillus pseudofirmus OF4	NC_013791
Bacillus pumilus SH-B9	NZ_CP011007
Bacillus simplex SH-B26	NZ_CP011008
Bacillus smithii DSM4216	NZ_CP012024
Bacillus sonorensis SRCM101395	NZ_CP021920
Bacillus thuringiensis serovar konkukian 97-27	NC_005957
Bacillus toyonensis BCT-7112	NC_022781
Bacillus vallismortis NBIF-001	NZ_CP020893
Bacillus velezensis FZB42	NC_009725

Note. *NCBI accession number

Table 2.2 Genomes of *Bacillales* retrieved from NCBIs FTP server (Continued)

Genome	**Reference***
Bacillus weihaiensis Alg07	NZ_CP016020
Non-bacillus	
Alicyclobacillus acidocaldarius DSM446	NC_013205
Alkalibacillus haloalkaliphilus C5	NZ_AKIF01000001
Amphibacillus marinus CGMCC1-10434	NZ_FODJ01000028
Amphibacillus sediminis NBRC103570	NZ_BCQW01000001
Amphibacillus xylanus NBRC15112	NC_018704
Aneurinibacillus soli CB4	NZ_AP017312
Aneurinibacillus sp. XH2	NZ_CP014140
Anoxybacillus amylolyticus DSM15939	NZ_CP015438
Anoxybacillus gonensis G2	NZ_CP012152
Anoxybacillus tepidamans PS2	NZ_JHVN01000001
Brevibacillus brevis NBRC100599	NC_012491
Brevibacillus formosus NF2	NZ_CP018145
Brevibacillus laterosporus LMG1544	NZ_CP007806
Exiguobacterium antarcticum B7	NC_018665
Geobacillus genomosp JF8	NC_022080
Geobacillus kaustophilus HTA426	NC_006510
Geobacillus sp. C56-T3	NC_014206
Geobacillus stearothermophilus 10	NZ_CP008934
Geobacillus subterraneus KCTC3922	NZ_CP014342
Geobacillus thermoleovorans KCTC3570	NZ_CP014335
Gracilibacillus halophilus YIM-C55-5	NZ_APML01000001
Gracilibacillus kekensis CGMCC1.10681	NZ_FRCZ01000013
Gracilibacillus lacisalsi DSM19029	NZ_KB898662
Gracilibacillus massiliensis	NZ_CZRP01000001

Note. *NCBI accession number

Table 2.2 Genomes of *Bacillales* retrieved from NCBIs FTP server (Continued)

Genome	**Reference***
Gracilibacillus orientalis CGMCC1.4250	NZ_FOTR01000026
Gracilibacillus timonensis P2481	NZ_FLKH01000001
Gracilibacillus ureilyticus CGMCC1-7727	NZ_FOGL01000035
Halalkalibacillus halophilus DSM18494	NZ_KE383978
Halobacillus halophilus DSM2266	NC_017668
Halobacillus mangrovi KTB131	NZ_CP020772
Halolactibacillus alkaliphilus CGMCC1-6843	NZ_FOWN01000063
Halolactibacillus halophilu DSM17073	NZ_FOXC01000095
Halolactibacillus miurensis DSM17074	NZ_FPAI01000063
Halolactibacillus sp. JCM19043	NZ_BAXD01000001.1
Jeotgalibacillus malaysiensis D5	NZ_CP009416
Jeotgalicoccus marinus DSM19772	NZ_KE384460
Jeotgalicoccus saudimassiliensis 13MG44	NZ_CCSE01000001
Kyrpidia tusciae DSM2912	NC_014098
Lentibacillus amyloliquefaciens LAM0015	NZ_CP013862
Lysinibacillus fusiformis RB21	NZ_CP010820
Lysinibacillus sphaericus 2362	NZ_CP015224
Lysinibacillus sphaericus III37	NZ_CP014856
Lysinibacillus sphaericus LMG22257	NZ_CP017560
Lysinibacillus varians Gy32	NZ_CP006837
Oceanobacillus iheyensis HTE831	NC_004193
Ornithinibacillus halophilus IBRCM10683	NZ_FQVW01000109
Paenibacillus beijingensis DSM24997	NZ_CP011058
Paenibacillus borealis DSM13188	NZ_CP009285
Paenibacillus bovis BD3526	NZ_CP013023
Paenibacillus donghaensis KCTC13049	NZ_CP021780

Note. *NCBI accession number

Table 2.2 Genomes of *Bacillales* retrieved from NCBIs FTP server (Continued)

Genome	**Reference***
Paenibacillus durus DSM1735	NZ_CP009288
Paenibacillus larvae ATCC9545	NZ_CP019687
Paenibacillus mucilaginosus 3016	NC_016935
Paenibacillus mucilaginosus KNP414	NC_015690
Paenibacillus naphthalenovorans 32OY	NZ_CP013652
Paenibacillus odorifer DSM15391	NZ_CP009428
Paenibacillus polymyxa E681	NC_014483
Paenibacillus polymyxa M1	NC_017542
Paenibacillus polymyxa SC2	NC_014622
Paenibacillus sp. JDR-2	NC_012914
Paenibacillus sp. Y412MC10	NC_013406
Paenibacillus stellifer DSM14472	NZ_CP009286
Paenibacillus swuensis DY6	NZ_CP011388
Paenibacillus terrae HPL-003	NC_016641
Paenibacillus yonginensis DCY84	NZ_CP014167
Paenisporosarcina indica	NZ_MPTA01000001
Paenisporosarcina quisquiliarum	NZ_FOBQ01000020
Paenisporosarcina sp. HGH0030	NZ_KE150421
Paraliobacillus ryukyuensis Marseille-P3391	NZ_FVZO01000017
Paraliobacillus sp. PM-2	NZ_CTEI01000001
Paucisalibacillus globulus DSM1884	NZ_AXVK01000001
Piscibacillus halophilus DSM21633	NZ_FOES01000082
Pontibacillus halophilus JSM076056 DSM19796	NZ_AULI01000001
Pontibacillus marinus BH030004 DSM16465	NZ_AULJ01000001
Saccharibacillus kuerlensis DSM22868	NZ_KB899277
Saccharibacillus sacchari DSM19268	NZ_KK073875

Note. *NCBI accession number

Table 2.2 Genomes of *Bacillales* retrieved from NCBIs FTP server (Continued)

Genome	**Reference***
Salimicrobium jeotgali MJ3	NZ_CP011361
Salsuginibacillus kocurii DSM18087	NZ_KB898623
Sporolactobacillus nakayamae ATCC700379	NZ_FOOY01000021
Sporolactobacillus vineae DSM21990	NZ_KB899042
Sporosarcina psychrophila DSM6497	NZ_CP014616
Sporosarcina sp. P33	NZ_CP015027
Sporosarcina sp. P37	NZ_CP015349
Sporosarcina ureae P17a	NZ_CP015109
Sporosarcina ureae P8	NZ_CP015207
Tenuibacillus multivorans CGMCC1-344	NZ_FNIG01000018
Terribacillus halophilus T-h1	NZ_LT727815
Thalassobacillus cyri CCM7597	NZ_FNQR01000037
Thalassobacillus devorans MSP14	NZ_KI543236
Thermicanus aegyptius DSM12793	NZ_KI783301
Thermobacillus composti KWC4	NC_019897
Tuberibacillus calidus DSM17572	NZ_KE387193
Tumebacillus algifaecis	NZ_CP022657
Ureibacillus thermosphaericus A1	NZ_AP018335
Virgibacillus halodenitrificans PDBF2	NZ_CP017962
Virgibacillus necropolis LMG19488	NZ_CP022437
Virgibacillus phasianinus LM2416	NZ_CP022315
Virgibacillus sp. 6R	NZ_CP017762
Virgibacillus sp. SK37	NZ_CP007161
Vulcanibacillus modesticaldus BR	NZ_MIJF01000001

Note. *NCBI accession number

Command-line BLAST+ v2.7.1 was downloaded from the National Institutes of Health server (`https://www.nih.gov/`) for Windows 10 Home with a 64-bit operating system (Altschul, Gish, Miller, Myers, & Lipman, 1990; Altschul et al., 1997; Camacho et al., 2009) and used to search of spore coat protein homologs in *Bacillus* and *Bacillales* genomes. For this, we created genome databases for all the genomes of Bacillales using the FASTA files. All the spore coat genes were translated to aminoacids in the correct reading frame using the server Sequence Manipulation Suite (`https://www.bioinformatics.org/sms2/translate.html`) (Stothard, 2000) and were included on a single FASTA file. The file was then blasted against each genome database of the Bacillales order using local BLASTp with default settings, and the output was saved as a TXT file with the option -outfmt "7 bitscore evalue qlen length sseq" to keep a record of the Bit score value and E-value. As suggested by Pearson (2013), an E-value less than 0.001 and a bit score higher than 40 were used to validate a result since the protein databases contained less than 7000 entries (Pearson, 2013).

Clustering analysis of spore coat proteins was performed using the software package Many-against-Many sequence searching (MMseqs2) (Steinegger & Söding, 2017, 2018) by Dr. J. A Castillo to group spore coat proteins homologs with a 70 and 90% of coverage in *Bacillales* genome. Additionally, the KEGG Orthology database (Kanehisa, Sato, Kawashima, Furumichi, & Tanabe, 2016; Kanehisa, Sato, & Morishima, 2016) was used by Dr. J. A. Castillo to search for spore coat genes orthologs across *Bacillales* genomes of Table 2.2.

Genes with positive hits for the three methods (BLASTp, Clustering, KEGG Orthology) were recorded as highly significant and accepted to appear in the subject genomes. On the other hand, genes with hits for one or two methods were accepted as secondarily significant. A consensus heat map was created to summarize the results provided by the three methods (see Figure 3.1 below).

2.2 Selection pressure and statistical analyses

Based on the results of the presence of spore coat proteins on the genome databases shown in Figure 3.1 (see below), we searched for the possible presence of each spore coat

gene on the genomes of the genus *Bacillus*. We used the GENBANK annotated genomes to ensure that each spore coat gene sequence is in its correct reading frame. For this, we discarded spore coat genes if the description of the gene on the annotated genome contained the word "*frameshifted*". We created gene datasets that contained all the found spore coat genes sequences for each genome. Furthermore, phylogenetic analyses were performed by Dr. J. A. Castillo using core genome amino acid sequences to infer monophyletic groups within *Bacillus*.

We aligned the spore coat genes sequences using the TranslatorX server (`http://translatorx.co.uk/`) (Abascal, Zardoya, & Telford, 2010) with MAFFT aligner and default settings. The aligned sequences were saved as FASTA files.

We applied the summary statistic Tajima's D to detect selection pressure of the spore coat genes within the different *Bacillus* groups. For each spore coat gene, we employed the DNASP v6.12 software (Rozas et al., 2017) with the nucleotide substitutions considered as the total number of mutations for the Tajima's D summary statistic. Since DNASP requires a minimum of four gene sequences to calculate Tajima's D, spore coat gene datasets with less than four sequences were not taken into account. A p-value less than 0.05 was considered to validate a result as significant.

For the following analyses of this section, all the gene datasets did not contain any stop codons as required by the methods. We used the webserver DataMonkey (`http://test.datamonkey.org/`), which implements the method "Branch-Site Unrestricted Statistical Test for Episodic Diversification" (BUSTED) that is useful to detect positive selection on at least one branch of the phylogeny at a gene level (Murrell et al., 2015), and the method 'Mixed Effects Model of Evolution' (MEME) that detects episodic positive or diversifying selection at individual sites in genes (Murrell et al., 2012). In BUSTED, we selected all the branches of the phylogeny for the analysis (foreground branches).

Moreover, we employed CODEML that is part of the Phylogenetic Analysis by Maximum Likelihood (PAML) package to calculate the ratio (ω) of non-synonymous (dN) to synonymous (dS) changes across spore coat genes sequences (Yang, 1997, 2007). To provide the phylogeny required by CODEML, we used the webserver PhyML (`http://www.atgc-montpellier.fr/`

phyml/) (Guindon et al., 2010) with the Subtree Pruning and Regrafting (SPR) option and default settings to reconstruct the phylogeny for each spore coat gene. Then, the aligned gene sequences as FASTA files and phylogenetic trees as TREE files were used in CODEML. For this analysis, sites and branch models were used with default settings and 'codons' as the sequence type. In sites model, we tested each gene sequence for the following nested models 'M1 nearly neutral' ($\omega < 1$; $\omega = 1$) (Nielsen & Yang, 1998; Yang, Wong, & Nielsen, 2005), 'M2 positive selection' ($\omega < 1$; $\omega = 1$; $\omega > 1$) (Nielsen & Yang, 1998; Yang et al., 2005) and 'M7 β distribution' ($\omega < 1$; $\omega = 1$) (Yang, Nielsen, Goldman, & Pedersen, 2000), 'M8 β distribution + positive selection' ($\omega < 1$; $\omega = 1$; $\omega > 1$) (Yang et al., 2000). Then, we performed a 'Likelihood Ratio Test' (LRT) to select the model that best fits the given data.

2.3 Horizontal gene transfer events

To search for HGT events in spore spore coat genes, we employed the software Notung v2.9 that reconciles a gene tree with a species tree to infer duplication-transfer-loss (DTL) event models with a parsimony-based optimization criterion (Darby, Stolzer, Ropp, Barker, & Durand, 2017; Durand, Halldórsson, & Vernot, 2006; Stolzer et al., 2012; Vernot, Stolzer, Goldman, & Durand, 2008). First, we uploaded the gene tree and the species tree. Then, we selected the "Prefix of the gene label (i.e. SPECIESGENE)" option to reconcile the trees.

To perform DTL event models, Notung requires rooted trees. For this, we employed the Bayesian Evolutionary Analysis Sampling Tree (BEAST) v1.8.4 package (Suchard et al., 2018) to reconstruct the phylogeny of each spore coat gene and its species. For this reconstruction, we employed the best-fit model of nucleotide substitution for our datasets. The models were inferred using the webserver Smart Model Selection in PhyML (SMS) (http://www.atgc-montpellier.fr/sms/) (Lefort, Longueville, & Gascuel, 2017) to select the best selection model using a likelihood-based criterion (AIC) for nucleotide spore coat genes and amino acid core genome sequences.

The following parameters were used in BEAUti (part of the BEAST package) for the phylogenetic reconstruction: substitution model suggested by SMS, strict molecular clock, Co-

alescent Bayesian Skyline as the tree prior, 10 000 000 iterations as the length of chains for Markov Chain Monte Carlo (MCMC), and 1000 as echo state. We employed Tracer v1.7.1 (Rambaut, Drummond, Xie, Baele, & Suchard, 2018) to assess the Effective Sample Size (ESS) values of the MCMC trace files produced by BEAST, and to confirm that the analysis reached the convergence. Furthermore, TreeAnnotator v1.8.4 was employed to generate a single tree as a TREE file that summarizes the information of sampled trees produced by BEAST.

In order to reduce false positives, we scanned the genomes of possible candidates of HGT events for traces of integrative, conjugative, and mobile elements, based on the results provided by Notung. For this, we downloaded a region of the genome of approximately ten genes upward and downward from the spore coat gene subjected to HGT as nucleotides in FASTA files from the NCBI's FTP server.

Then, we used the detection tool "WU-BLAST2 search" of the web server ICEberg 2.0 (`http://db-mml.sjtu.edu.cn/ICEberg/`), which is a database containing information about bacterial integrative and conjugative elements (ICEs), as well as integrative and mobilizable elements (IMEs) and cis-mobilizable elements (CIMEs) (Liu et al., 2019)(Liu et al., 2019). Furthermore, we employed the Genomic Island Prediction Software v1.1.2 (GIPSy) to detect genomic islands (GEIs), which are further classified into four categories: pathogenicity islands (PAIs), metabolic islands (MIs), resistance islands (RIs), and symbiotic islands (SIs), which may contain genes encoding virulence, metabolism, antibiotic resistance, and symbiosis-related functions respectively (Barcellos, Menna, Batista, & Hungria, 2007; Dobrindt et al., 2000; Krizova & Nemec, 2010; Soares et al., 2016; Tumapa et al., 2008). GEIs tend to carry novel genes that do not have homologs in other species compared to the rest of the genome (Hsiao et al., 2005; Juhas et al., 2009). Thus, we applied GIPSy to detect spore coat genes that have been acquired by HGT and are located on GEIs for each *Bacillus* group. Hits with an E-value less than 0.001 and a Bit score higher than 40 were considered as valid (Pearson, 2013).

CHAPTER 3. RESULTS

3.1 Spore coat proteins diversity across *Bacillales* and consensus heat map

In order to understand the diversity of spore coat genes on *Bacillales*, we carried out three distinct methods to identify the possible existence of coat genes homologs on different genomes. The three methods were used to reduce false positives. In the first approach, local databases were created using command-line BLAST+ with one hundred forty-one genomes of *Bacillales* and ninety-seven spore coat genes of *B. subtilis* 168, the model organism for sporulation studies (Driks & Eichenberger, 2016).

BLAST+ provided statistical information of similarity ('bit score' and 'e-value') for each alignment (Camacho et al., 2009) between the spore coat protein sequences (query) and the *Bacillales* genomes sequences (subject). Given that E-value depends on database size, the alignment score may be less significant in larger databases compared to smaller databases with the same score (Pearson, 2013). Bit score and E-value are more sensitive to infer homology than percent identity (Pearson, 2013). In protein databases with less than 7000 entries a bit score of 40 and E-value < 0.001 is considered significant (Pearson, 2013), as the case for this work. Furthermore, we employed a translated-DNA:protein alignment because this type of alignment can detect homology in sequences that diverged more than 2.5 billion years ago, as the case of bacteria (Pearson, 2013). The results of BLASTp+ are shown in Figure SX as the number of positive hits of each spore coat protein on each genome.

To provide further evidence of the presence of spore coat proteins on *Bacillales* the results of Clustering and KEGG Orthology were taken into account (Castillo, J. A. unpublished results, 2020). The software MMseqs2 was used to cluster similar spore coat protein sequences with a coverage of 70 to 90%, whereas KEGG Orthology tool was used to find spore coat genes orthologs across different *Bacillales* genomes.

Figures 3.1 and 3.2 represents the presence/absence relation of spore coat proteins homologs across *Bacillales*. As expected, *B. subtillis* 168 contains the ninety-seven spore coat

proteins, this is confirmed by BLASTp+, Clustering, and KEGG Orthology. *Bacillus gibsonii* FJAT 10019, *Bacillus atrophaeus* SRCM101359, *Bacillus vallismortis* NBIF 001, and *Bacillus velezensis* FZB42 contain a higher number of spore coat proteins compared to the rest of *Bacillales* because they belong to the *Bacillus subtilis* group, according to the NCBI taxonomy (Benson, Karsch-Mizrachi, Lipman, Ostell, & Sayers, 2009; Sayers et al., 2009) and our phylogenetic analyses of the core genome (Castillo, J. A. unpublished results, 2020).

Moreover, spore coat proteins such as CotE, CotJA, CotJB, CotJC, CwlJ GerQ, SpoIVA, SpoVID, SpoVM, YhbA/QueG, YhbB, and YchN appear to be mostly ubiquitous among the *Bacillales* genomes analyzed in this work. Other spore coat proteins, such as GerPA, GerPB, GerPC, GerPD, GerPE, and GerPF seem to be present on *Alkalibacillus haloalkaliphilus* C5, *Amphibacillus marinus* CGMCC1-10434, *Amphibacillus sediminis* NBRC103570, genomes of the genus *Geobacillus* and *Gracibacillus*, *Halalkalibacillus halophilus* DSM18494, *Halobacillus halophilus* DSM2266, and *Halobacillus mangrovi* KTB131. Likewise, the spore coat protein CgeB have many homologs among the genomes of the genus *Paenibacillus*. Overall, genomes of the genus *Bacillus* contain some spore coat proteins homologs not present on other *Bacillales*, such as CgeD, CotA, CotD, CotI, CotN, CotR, CotS, CotSA, LipC, SafA, Tgl, YaaH, YdhD, YhcQ, YheC, YheD, YhjR, YisY, YjqC, YkvP, YkvQ, YkzQ, YlbD, YlbO/GerR, YncD, YpeP, YrbB/CoxA, YrbC, YsxE, YtdA, YusN, YutH, YuzC, and YwqH. However, those spore coat proteins homologs are only detected by one method, see Figure X.

Spore coat proteins, such as CgeA, CotC, CotT, CotU, YdgA, YdgB, YeeK, YjcB, YjzB, YknT, YmaG, YsnD, and YtxO are poorly present on genomes of *Bacillales* other than *B. subtilis* 168 and *B. gibsonii* FJAT 10019. The genomes of *Bacillus beveridgei* MLTeJB, *Exiguobacterium antarcticum* B7, *Jeotgalicoccus marinus* DSM19772 and *Jeotgalicoccus saudimassiliensis* 13MG44 contain few secondarily significant results of spore coat proteins homologs.

3.2 Selection pressure

In order to understand selection pressure that acts on spore coat genes, we employed the classical approaches Tajima's D test and dN/dS (ω) ratio and two new approaches (BUSTED,

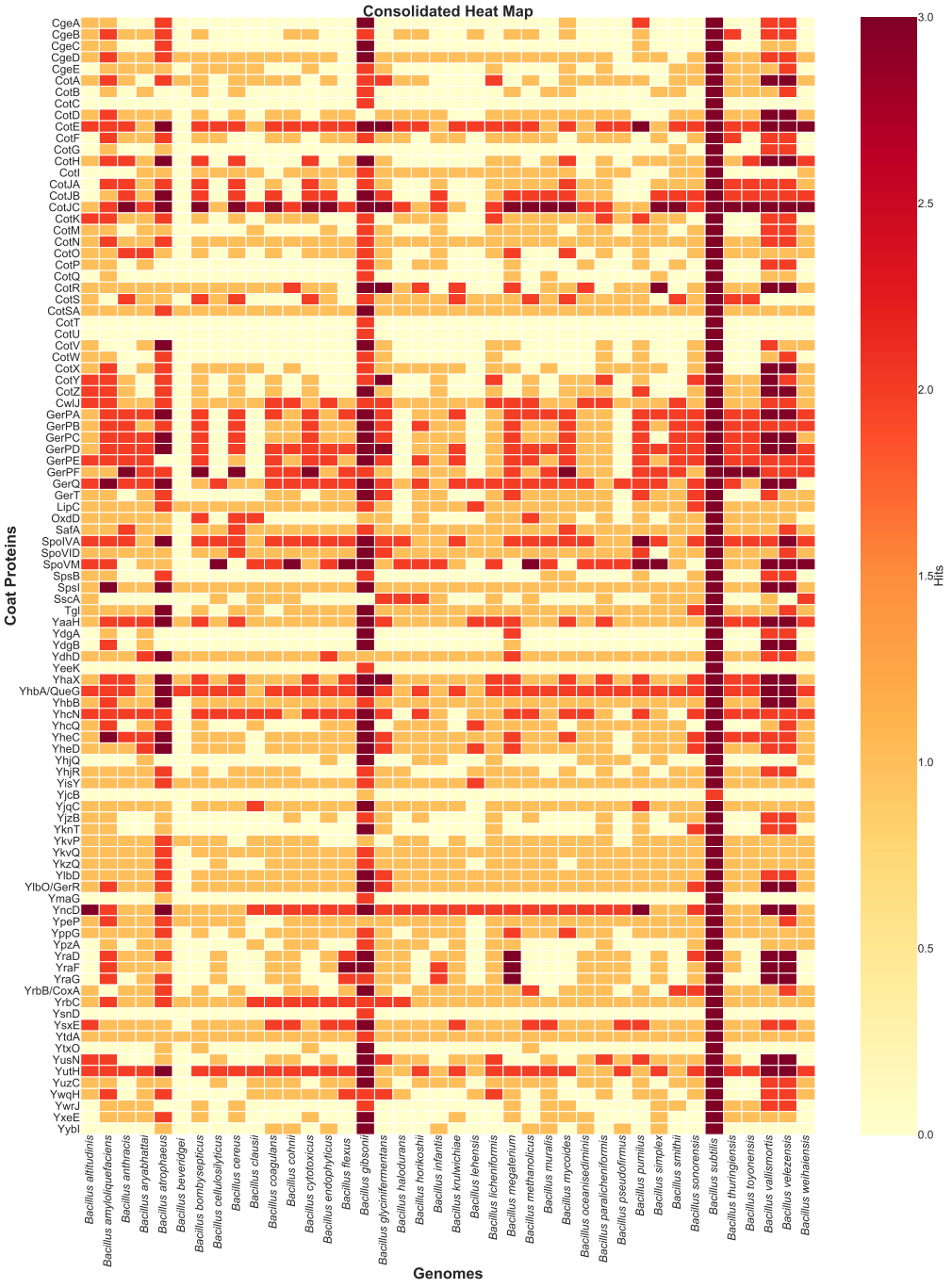


Figure 3.1. Consolidated heat map of ninety-seven spore coat proteins homologs over forty genomes of *Bacillus* based on three methods: BLAST+p, Clustering, and KEGG Orthology. Primarily significant results (dark red) have been confirmed by the three methods, whereas secondarily significant results (orange and yellow) have been confirmed by either one or two methods.

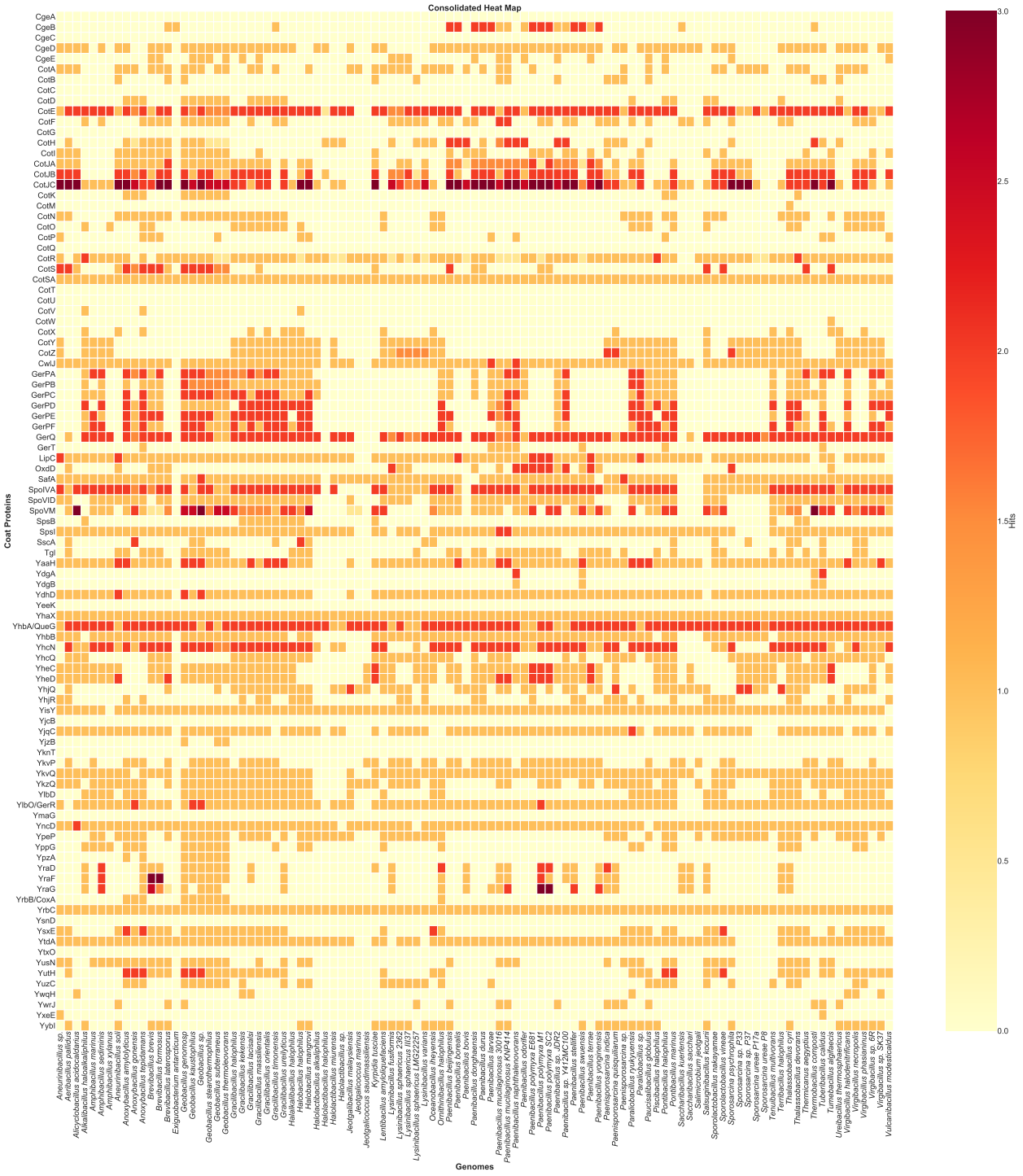


Figure 3.2. Consolidated heat map of ninety-seven spore coat proteins homologs over one hundred and one genomes of *Non-bacillus* based on three methods: BLAST+p, Clustering, and KEGG Orthology. Primarily significant results (dark red) have been confirmed by the three methods, whereas secondarily significant results (orange and yellow) have been confirmed by either one or two methods.

MEME) that use sophisticated algorithms for detecting episodic positive selection in all or a subset of branches on a phylogeny. For this, we created a spore coat genes sequences datasets for each *Bacillus* group, based on the results of the consensus heat map (Figure 3.1). Unfortunately, we were not able to retrieve all the spore coat gene sequences reported in Figure 3.1 due to the lack of information and annotation errors of spore coat genes on genomes available on NCBI, see Table 3. Moreover, we used the phylogenetic analyses of core genome performed by Dr. J. A. Castillo to classify groups within the *Bacillus* genus.

Table 3.1 Coat genes across *Bacillus* groups retrieved from genome databases, based on presence/absence results of Figure 3.1

Coat gene
Cereus group
cotE
cotH
cotJA
cotJB
cotJC
cotK/sspO
cotS
cotY
cwlJ
gerPA
gerPB
gerPC
gerPD
gerPE
gerPF
gerQ
spoIVA

Table 3.1 Coat genes across *Bacillus* groups retrieved from genome databases, based on presence/absence results of Figure 3.1. (Continued)

Coat gene
spoVID
spoVM
yhbA/queG
yhcN
yheC
yheD
ysxE
yutH/cotNH
Coagulans group
cotE
gerPE
gerQ
spoVM
yhbA/queG
ysxE
yutH/cotNH
Halodurans group
gerQ
yhbA/queG
Megaterium group
cotE
gerPE
gerQ
yhbA/queG
yhcN

Table 3.1 Coat genes across *Bacillus* groups retrieved from genome databases, based on presence/absence results of Figure 3.1. (Continued)

Coat gene
ysxE
yutH/cotNH
Methanolicus group
cotE
cotJA
cotJB
gerPE
gerQ
spoIVA
spoVID
yhbA/queG
yhcN
ysxE
yutH/cotNH
Pumilus group
cotE
gerPB
gerPC
gerPE
gerQ
gerT
yhbA/queG
yhcN
ysxE
Subtilis group

Table 3.1 Coat genes across *Bacillus* groups retrieved from genome databases, based on presence/absence results of Figure 3.1. (Continued)

Coat gene
cotA
cotE
cotJA
cotJB
cotJC
cotY
cotZ
gerPB
gerPE
gerPD
gerPE
gerQ
gerT
safA
spoIVA
spoVM
yhbA/queG
yhcN
yheC
yrbB/coxA
ysxE
yutH/cotNH

Tajima's D takes into account the number of segregating sites and the average number of nucleotide differences under a neutral mutation model and a population at mutation-drift equilibrium (Tajima, 1989). The null hypothesis for Tajima's D states that no selection is

acting at a locus and that the population has not experienced recent events of growth or contraction (Schmidt & Pool, 2002; Tajima, 1989). Furthermore, Tajima's D test values are sensitive to selection or demographic events (Campo et al., 2013; Schmidt & Pool, 2002). For instance, a positive value may reflect balancing selection or demographic events of population reduction, population subdivision or a recent bottleneck (Castillo & Agathos, 2019; Schmidt & Pool, 2002). On the other side, a negative value may reflect purifying selection or events of population expansion, a distant bottleneck or migration (Campo et al., 2013; Schmidt & Pool, 2002).

BUSTED is an unrestricted branch-site random effect model that is capable of detecting positive selection on a subset of branches on a phylogeny at a subset of sites within a gene (Murrell et al., 2015). BUSTED allows ω to vary from branch to branch (Kosakovsky Pond et al., 2011; Murrell et al., 2015), and can be employed on an entire phylogeny (without previous knowledge of which branches are under positive selection; all branches are treated as foreground branches) or on a pre-specified subset of branches (foreground and background branches) (Murrell et al., 2015). The use of a stochastic selection of the random effects framework to detect evidence of positive selection has an advantage over classical test which average ω across branches, codons or both (Murrell et al., 2015). It is important to mention that a significant result for positive selection does not imply that the gene evolved under positive selection along the entire foreground branch but that at least one site, at least some of the time, has experienced positive selection (Murrell et al., 2015).

MEME is based on branch-site random effects phylogenetic models with a mixed-effects maximum likelihood approach to test whether individual sites under a proportion of branches have evolved under episodic and pervasive positive selection (Murrell et al., 2012). MEME allows ω to vary from site to site and from branch to branch at a site (Murrell et al., 2012).

CodeML employs several codon substitution models to calculate the ratio ω between dN non-synonymous and dS synonymous substitution rates that have acted on protein-coding sequences (Yang, 1997, 2007). Values of $\omega < 1$, $=1$, and >1 represent purifying selection (non-synonymous rates is lower than synonymous rates), neutral evolution (similar non-synonymous and synonymous rates), and positive selection (non-synonymous rates is higher than synony-

mous rates), respectively (Anisimova & Kosiol, 2009; Yang, 1997, 2007). Codon models are useful to detect protein-coding sequences evolving under purifying or positive selection (Delport, Scheffler, & Seoighe, 2009).

Significant results (p-value < 0.05) of spore coat genes displaying either strong negative purifying or positive selection on different *Bacillus* groups are reported in Table 3.2. On the Cereus group, *cotH* has ten sites, *cotK/sspO* and *ysxE* have one site, *yheC* has three sites, *yheD* has five sites, and *yutH-cotNH* has thirteen sites under positive selection according to MEME. Furthermore, *cwlJ* and *gerQ* are positively selected along the entire gene sequence according to BUSTED and have two and one sites under positive selection, respectively detected by MEME.

On the other hand, the Coagulans group present four genes under negative purifying selection (*cotE, gerPE, gerQ, spoVM, and yhbA-queG*) according to the summary statistic Tajima's D. On the contrary, *ysxE* has two sites under positive selection according to MEME. *gerQ* is negatively selected according to Tajima's D whereas MEME detects one positively selected site. This site can be positively selected because it is the active or interaction site of the GerQ spore coat protein.

The Halodurans group, show the spore coat genes *gerQ* and *yhbA/queG* under positive selection with five and three sites, respectively according to MEME. The Megaterium group, contains three significant negatively selected spore coat genes *gerQ, yhcN, yutH/cotNH* according to Tajima's D and one spore coat gene *ysxE* with two sites evolving under positive selection. Nonetheless, *gerPE* appear to be significant for both Tajima's D and MEME, which is contradictory.

On the Methanolicus group there are four spore coat genes (*cotE, gerPE, ysxE, yutH/cotNH*) with one site evolving under positive selection according to MEME and three spore coat genes (*cotJA, cotJB, spoIVA, yhcN*) under negative purifying selection according to Tajima's D. On one hand, *spoVID* seems to be under a positive selection along the entire gene sequences according to BUSTED and with three positively selected sites according to MEME. On the other hand, Tajima's D gives a significant negative value for *spoVID* indicating negative purifying selection. This could be a false positive detected by Tajima's D. However, we cannot draw any conclusions for *spoVID* due to the contradiction of the methods.

In the Pumilus group, *cotE, gerPC, gerPE, gerQ, yhbA/queG, yhcN,* and *ysxE* are under positive selection with two, four, six, one, four, five and ten sites respectively. On the Subtilis group *cotE, gerPE, gerQ, gerT, safA, yhbA/queG, yheC*, and *ysxE* are positively selected in one or two sites, according to MEME. Moreover, *gerQ* is under positive selection along its entire gene sequence according to BUSTED. Overall, ω ratios calculated branch and site models in CodeML are not able to detect positive or negative selection acting on the spore coat genes of Table 4 since all the values are indicating neutral selection.

Table 3.2 Five summary statistics (Tajima's D, BUSTED, MEME, dN/dS (branch and site models) showing strong negative purifying selection, neutral, and positive selection across different *Bacillus* groups

Coat gene	**Sumary statistics**				
	Tajima's D	**BUSTED**$^{\gamma}$	**MEME**$^{\delta}$	**dN/dS (Branch model)**	**dN/dS (Site models)**
Cereus group					
cotH	-0.69763	1	10*	0.07182	M1: Nearly neutral 0.1028 M7: β distribution 0.0962
cotK/sspO	0.28136	0.262	1*	0.09271	M1: Nearly neutral 0.1123 M7: β distribution 0.1034

γp-value provided by BUSTED. A p-value < 0.05 indicates evidence of positive selection of the gene

δNumber of significant sites sites under positive selection by MEME

*Significant at a p-value < 0.05

ϵdN/dS values could not be computed in CodeML due to small branch size

Table 3.2 Five summary statistics (Tajima's D, BUSTED, MEME, dN/dS (branch and site models) showing strong negative purifying selection, neutral, and positive selection across different *Bacillus* groups (Continued)

Coat gene	**Sumary statistics**				
	Tajima's D	**BUSTED**$^{\gamma}$	**MEME**$^{\delta}$	**dN/dS (Branch model)**	**dN/dS (Site models)**
cwlJ	0.09189	0*	2*	0.05187	M1: Nearly neutral 0.0794 M7: β distribution 0.0571
gerQ	-0.88202	0.007*	1*	0.07437	M1: Nearly neutral 0.1527 M7: β distribution 0.0795

γp-value provided by BUSTED. A p-value < 0.05 indicates evidence of positive selection of the gene

δNumber of significant sites sites under positive selection by MEME

*Significant at a p-value < 0.05

ϵdN/dS values could not be computed in CodeML due to small branch size

Table 3.2 Five summary statistics (Tajima's D, BUSTED, MEME, dN/dS (branch and site models) showing strong negative purifying selection, neutral, and positive selection across different *Bacillus* groups (Continued)

Coat gene	**Sumary statistics**				
	Tajima's D	**BUSTED$^{\gamma}$**	**MEME$^{\delta}$**	**dN/dS (Branch model)**	**dN/dS (Site models)**
yheC	-0.47554	1	3*	0.09094	M1: Nearly neutral 0.0984 M7: β distribution 0.0957
yheD	-0.43883	1	5*	0.08993	M1: Nearly neutral 0.0971 M7: β distribution 0.0941

γp-value provided by BUSTED. A p-value < 0.05 indicates evidence of positive selection of the gene

δNumber of significant sites sites under positive selection by MEME

*Significant at a p-value < 0.05

ϵdN/dS values could not be computed in CodeML due to small branch size

Table 3.2 Five summary statistics (Tajima's D, BUSTED, MEME, dN/dS (branch and site models) showing strong negative purifying selection, neutral, and positive selection across different *Bacillus* groups (Continued)

Coat gene	**Sumary statistics**				
	Tajima's D	**BUSTED$^{\gamma}$**	**MEME$^{\delta}$**	**dN/dS (Branch model)**	**dN/dS (Site models)**
ysxE	-0.97925	1	1*	0.09371	M1: Nearly neutral 0.1166 M7: β distribution 0.0995
yutH/cotNH	-0.54963	0.857	13*	0.11605	M1: Nearly neutral 0.1698 M7: β distribution 0.1318
Coaguolans group					

γp-value provided by BUSTED. A p-value < 0.05 indicates evidence of positive selection of the gene

δNumber of significant sites sites under positive selection by MEME

*Significant at a p-value < 0.05

ϵdN/dS values could not be computed in CodeML due to small branch size

Table 3.2 Five summary statistics (Tajima's D, BUSTED, MEME, dN/dS (branch and site models) showing strong negative purifying selection, neutral, and positive selection across different *Bacillus* groups (Continued)

Coat gene	**Sumary statistics**				
	Tajima's D	**BUSTED$^{\gamma}$**	**MEME$^{\delta}$**	**dN/dS (Branch model)**	**dN/dS (Site models)**
cotE	-1.37084*	1	0	0.12021	M1: Nearly neutral 0.2828 M7: β distribution 0.1708
gerPE	-1.34236*	0.963	0	0.14292	M1: Nearly neutral 0.2615 M7: β distribution 0.1778

γp-value provided by BUSTED. A p-value < 0.05 indicates evidence of positive selection of the gene

δNumber of significant sites sites under positive selection by MEME

*Significant at a p-value < 0.05

ϵdN/dS values could not be computed in CodeML due to small branch size

Table 3.2 Five summary statistics (Tajima's D, BUSTED, MEME, dN/dS (branch and site models) showing strong negative purifying selection, neutral, and positive selection across different *Bacillus* groups (Continued)

Coat gene	**Sumary statistics**				
	Tajima's D	**BUSTED$^{\gamma}$**	**MEME$^{\delta}$**	**dN/dS (Branch model)**	**dN/dS (Site models)**
gerQ	-1.21128*	0.328	1*	0.08630	M1: Nearly neutral 0.4039 M7: β distribution 0.1833
spoVM	-1.12118*	1	0	NA$^{\epsilon}$	NA$^{\epsilon}$
yhbA/queG	-1.32171*	0.868	0	0.06442	M1: Nearly neutral 0.2047 M7: β distribution 0.1083

γp-value provided by BUSTED. A p-value < 0.05 indicates evidence of positive selection of the gene

δNumber of significant sites sites under positive selection by MEME

*Significant at a p-value < 0.05

ϵdN/dS values could not be computed in CodeML due to small branch size

Table 3.2 Five summary statistics (Tajima's D, BUSTED, MEME, dN/dS (branch and site models) showing strong negative purifying selection, neutral, and positive selection across different *Bacillus* groups (Continued)

Coat gene	**Sumary statistics**				
	Tajima's D	**BUSTED**$^{\gamma}$	**MEME**$^{\delta}$	**dN/dS (Branch model)**	**dN/dS (Site models)**
ysxE	-1.23355	0.188	2*	0.02038	M1: Nearly neutral 0.1265 M7: β distribution 0.0675
Halodurans group					
gerQ	-0.56024	0.332	5*	0.10751	M1: Nearly neutral 0.3004 M7: β distribution 0.1627

γp-value provided by BUSTED. A p-value < 0.05 indicates evidence of positive selection of the gene

δNumber of significant sites sites under positive selection by MEME

*Significant at a p-value < 0.05

ϵdN/dS values could not be computed in CodeML due to small branch size

Table 3.2 Five summary statistics (Tajima's D, BUSTED, MEME, dN/dS (branch and site models) showing strong negative purifying selection, neutral, and positive selection across different *Bacillus* groups (Continued)

Coat gene	**Sumary statistics**				
	Tajima's D	**BUSTED$^{\gamma}$**	**MEME$^{\delta}$**	**dN/dS (Branch model)**	**dN/dS (Site models)**
yhbA/queG	-1.16799	1	3*	0.06531	M1: Nearly neutral 0.1854 M7: β distribution 0.0853
Megaterium group					
gerPE	-1.06119*	0.94	1*	0.09876	M1: Nearly neutral 0.5707 M7: β distribution 0.1188

γp-value provided by BUSTED. A p-value < 0.05 indicates evidence of positive selection of the gene

δNumber of significant sites sites under positive selection by MEME

*Significant at a p-value < 0.05

ϵdN/dS values could not be computed in CodeML due to small branch size

Table 3.2 Five summary statistics (Tajima's D, BUSTED, MEME, dN/dS (branch and site models) showing strong negative purifying selection, neutral, and positive selection across different *Bacillus* groups (Continued)

Coat gene	**Sumary statistics**				
	Tajima's D	**BUSTED$^{\gamma}$**	**MEME$^{\delta}$**	**dN/dS (Branch model)**	**dN/dS (Site models)**
gerQ	-0.90984*	1	0	0.07461	M1: Nearly neutral 0.2540 M7: β distribution 0.1193
yhcN	-1.02560*	1	0	0.11648	M1: Nearly neutral 0.1590 M7: β distribution 0.1368

γp-value provided by BUSTED. A p-value < 0.05 indicates evidence of positive selection of the gene

δNumber of significant sites sites under positive selection by MEME

*Significant at a p-value < 0.05

ϵdN/dS values could not be computed in CodeML due to small branch size

Table 3.2 Five summary statistics (Tajima's D, BUSTED, MEME, dN/dS (branch and site models) showing strong negative purifying selection, neutral, and positive selection across different *Bacillus* groups (Continued)

Coat gene	**Sumary statistics**				
	Tajima's D	**BUSTED$^{\gamma}$**	**MEME$^{\delta}$**	**dN/dS (Branch model)**	**dN/dS (Site models)**
ysxE	-0.74173	1	2*	0.09351	M1: Nearly neutral 0.1689 M7: β distribution 0.1026
yutH/cotNH	-0.99407*	1	0	0.04157	M1: Nearly neutral 0.1355 M7: β distribution 0.0509
Methanolicus group					

γp-value provided by BUSTED. A p-value < 0.05 indicates evidence of positive selection of the gene

δNumber of significant sites sites under positive selection by MEME

*Significant at a p-value < 0.05

ϵdN/dS values could not be computed in CodeML due to small branch size

Table 3.2 Five summary statistics (Tajima's D, BUSTED, MEME, dN/dS (branch and site models) showing strong negative purifying selection, neutral, and positive selection across different *Bacillus* groups (Continued)

Coat gene	**Sumary statistics**				
	Tajima's D	**BUSTED$^{\gamma}$**	**MEME$^{\delta}$**	**dN/dS (Branch model)**	**dN/dS (Site models)**
cotE	-1.03037	0.909	1*	0.10079	M1: Nearly neutral 0.1793 M7: β distribution 0.1162
cotJA	-1.02802*	0.544	0	0.15234	M1: Nearly neutral 0.3805 M7: β distribution 0.2220

γp-value provided by BUSTED. A p-value < 0.05 indicates evidence of positive selection of the gene

δNumber of significant sites sites under positive selection by MEME

*Significant at a p-value < 0.05

ϵdN/dS values could not be computed in CodeML due to small branch size

Table 3.2 Five summary statistics (Tajima's D, BUSTED, MEME, dN/dS (branch and site models) showing strong negative purifying selection, neutral, and positive selection across different *Bacillus* groups (Continued)

Coat gene	**Sumary statistics**				
	Tajima's D	**BUSTED$^{\gamma}$**	**MEME$^{\delta}$**	**dN/dS (Branch model)**	**dN/dS (Site models)**
cotJB	-1.36531*	0.769	0	0.07999	M1: Nearly neutral 0.1289 M7: β distribution 0.0960
gerPE	-1.2526	0.988	1*	0.11502	M1: Nearly neutral 0.2912 M7: β distribution 0.1685

γp-value provided by BUSTED. A p-value < 0.05 indicates evidence of positive selection of the gene

δNumber of significant sites sites under positive selection by MEME

*Significant at a p-value < 0.05

ϵdN/dS values could not be computed in CodeML due to small branch size

Table 3.2 Five summary statistics (Tajima's D, BUSTED, MEME, dN/dS (branch and site models) showing strong negative purifying selection, neutral, and positive selection across different *Bacillus* groups (Continued)

Coat gene	**Sumary statistics**				
	Tajima's D	**BUSTED$^{\gamma}$**	**MEME$^{\delta}$**	**dN/dS (Branch model)**	**dN/dS (Site models)**
spoIVA	-1.17678*	1	0	0.02073	M1: Nearly neutral 0.0320 M7: β distribution 0.0230
spoVID	-1.33519*	0.009*	3*	0.13830	M1: Nearly neutral 0.3608 M7: β distribution 0.2175

γp-value provided by BUSTED. A p-value < 0.05 indicates evidence of positive selection of the gene

δNumber of significant sites sites under positive selection by MEME

*Significant at a p-value < 0.05

ϵdN/dS values could not be computed in CodeML due to small branch size

Table 3.2 Five summary statistics (Tajima's D, BUSTED, MEME, dN/dS (branch and site models) showing strong negative purifying selection, neutral, and positive selection across different *Bacillus* groups (Continued)

Coat gene	**Sumary statistics**				
	Tajima's D	**BUSTED$^{\gamma}$**	**MEME$^{\delta}$**	**dN/dS (Branch model)**	**dN/dS (Site models)**
yhcN	-1.26681*	1	0	0.17294	M1: Nearly neutral 0.3185 M7: β distribution 0.2264
ysxE	-1.10637	0.229	1*	0.09149	M1: Nearly neutral 0.1956 M7: β distribution 0.1123

γp-value provided by BUSTED. A p-value < 0.05 indicates evidence of positive selection of the gene

δNumber of significant sites sites under positive selection by MEME

*Significant at a p-value < 0.05

ϵdN/dS values could not be computed in CodeML due to small branch size

Table 3.2 Five summary statistics (Tajima's D, BUSTED, MEME, dN/dS (branch and site models) showing strong negative purifying selection, neutral, and positive selection across different *Bacillus* groups (Continued)

Coat gene	**Sumary statistics**				
	Tajima's D	**BUSTED**$^{\gamma}$	**MEME**$^{\delta}$	**dN/dS (Branch model)**	**dN/dS (Site models)**
yutH/cotNH	-1.14348	1	1*	0.13182	M1: Nearly neutral 0.3628 M7: β distribution 0.1879
			Pumilus group		
cotE	-0.36331	1	2*	NA$^{\epsilon}$	NA$^{\epsilon}$

γp-value provided by BUSTED. A p-value < 0.05 indicates evidence of positive selection of the gene

δNumber of significant sites sites under positive selection by MEME

*Significant at a p-value < 0.05

ϵdN/dS values could not be computed in CodeML due to small branch size

Table 3.2 Five summary statistics (Tajima's D, BUSTED, MEME, dN/dS (branch and site models) showing strong negative purifying selection, neutral, and positive selection across different *Bacillus* groups (Continued)

Coat gene	**Sumary statistics**				
	Tajima's D	**BUSTED$^{\gamma}$**	**MEME$^{\delta}$**	**dN/dS (Branch model)**	**dN/dS (Site models)**
gerPC	-0.34568	1	4*	0.05307	M1: Nearly neutral 0.0564 M7: β distribution 0.0551
gerPE	-0.6854	0.075	6*	0.07784	M1: Nearly neutral 0.0804 M7: β distribution 0.0812

γp-value provided by BUSTED. A p-value < 0.05 indicates evidence of positive selection of the gene

δNumber of significant sites sites under positive selection by MEME

*Significant at a p-value < 0.05

ϵdN/dS values could not be computed in CodeML due to small branch size

Table 3.2 Five summary statistics (Tajima's D, BUSTED, MEME, dN/dS (branch and site models) showing strong negative purifying selection, neutral, and positive selection across different *Bacillus* groups (Continued)

Coat gene	**Sumary statistics**				
	Tajima's D	**BUSTED$^{\gamma}$**	**MEME$^{\delta}$**	**dN/dS (Branch model)**	**dN/dS (Site models)**
gerQ	-0.71509	0.711	1*	0.03936	M1: Nearly neutral 0.0379 M7: β distribution 0.0419
yhbA/queG	-0.62315	1	4*	0.03740	M1: Nearly neutral 0.0428 M7: β distribution 0.0415

γp-value provided by BUSTED. A p-value < 0.05 indicates evidence of positive selection of the gene

δNumber of significant sites sites under positive selection by MEME

*Significant at a p-value < 0.05

ϵdN/dS values could not be computed in CodeML due to small branch size

Table 3.2 Five summary statistics (Tajima's D, BUSTED, MEME, dN/dS (branch and site models) showing strong negative purifying selection, neutral, and positive selection across different *Bacillus* groups (Continued)

Coat gene	**Sumary statistics**				
	Tajima's D	**BUSTED**$^{\gamma}$	**MEME**$^{\delta}$	**dN/dS (Branch model)**	**dN/dS (Site models)**
yhcN	-0.47643	1	5*	0.08450	M1: Nearly neutral 0.0858 M7: β distribution 0.0873
ysxE	-0.43205	0.499	10*	0.05727	M1: Nearly neutral 0.0679 M7: β distribution 0.0604
Subtilis group					

γp-value provided by BUSTED. A p-value < 0.05 indicates evidence of positive selection of the gene

δNumber of significant sites sites under positive selection by MEME

*Significant at a p-value < 0.05

ϵdN/dS values could not be computed in CodeML due to small branch size

Table 3.2 Five summary statistics (Tajima's D, BUSTED, MEME, dN/dS (branch and site models) showing strong negative purifying selection, neutral, and positive selection across different *Bacillus* groups (Continued)

Coat gene	**Sumary statistics**				
	Tajima's D	**BUSTED**$^{\gamma}$	**MEME**$^{\delta}$	**dN/dS (Branch model)**	**dN/dS (Site models)**
cotE	0.10466	0.261	1*	NA$^{\epsilon}$	NA$^{\epsilon}$
gerPE	-0.3031	1	2*	0.16190	M1: Nearly neutral 0.2413 M7: β distribution 0.1826
gerQ	0.09547	0.001*	2*	NA$^{\epsilon}$	NA$^{\epsilon}$

γp-value provided by BUSTED. A p-value < 0.05 indicates evidence of positive selection of the gene

δNumber of significant sites sites under positive selection by MEME

*Significant at a p-value < 0.05

ϵdN/dS values could not be computed in CodeML due to small branch size

Table 3.2 Five summary statistics (Tajima's D, BUSTED, MEME, dN/dS (branch and site models) showing strong negative purifying selection, neutral, and positive selection across different *Bacillus* groups (Continued)

Coat gene	**Sumary statistics**				
	Tajima's D	**BUSTED$^{\gamma}$**	**MEME$^{\delta}$**	**dN/dS (Branch model)**	**dN/dS (Site models)**
gerT	-0.44083	0.053	1*	0.12826	M1: Nearly neutral 0.2203 M7: β distribution 0.1609
safA	1.55523	0.394	1*	0.11156	M1: Nearly neutral 0.1423 M7: β distribution 0.1359
yhbA/queG	-0.18729	0.687	1*	NA$^{\epsilon}$	NA$^{\epsilon}$

γp-value provided by BUSTED. A p-value < 0.05 indicates evidence of positive selection of the gene

δNumber of significant sites sites under positive selection by MEME

*Significant at a p-value < 0.05

ϵdN/dS values could not be computed in CodeML due to small branch size

Table 3.2 Five summary statistics (Tajima's D, BUSTED, MEME, dN/dS (branch and site models) showing strong negative purifying selection, neutral, and positive selection across different *Bacillus* groups (Continued)

Coat gene	**Sumary statistics**				
	Tajima's D	**BUSTED$^{\gamma}$**	**MEME$^{\delta}$**	**dN/dS (Branch model)**	**dN/dS (Site models)**
yheC	-0.66537	0.649	1*	0.07152	M1: Nearly neutral 0.1145 M7: β distribution 0.0776
ysxE	-0.20345	0.072	1*	0.13384	M1: Nearly neutral 0.2159 M7: β distribution 0.1503

γp-value provided by BUSTED. A p-value < 0.05 indicates evidence of positive selection of the gene

δNumber of significant sites sites under positive selection by MEME

*Significant at a p-value < 0.05

ϵdN/dS values could not be computed in CodeML due to small branch size

3.3 Horizontal gene transfer

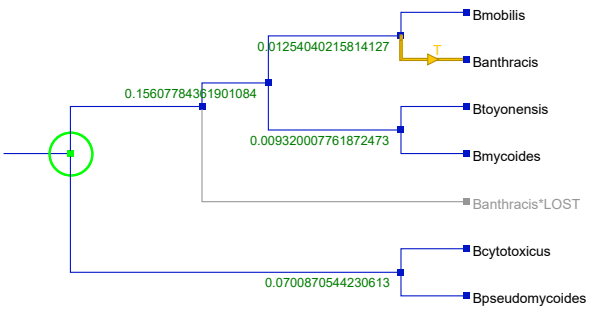
Studies from comparative genomics have demonstrated two broadly classes of genes: core genes (widely distributed across different lineages and encode central function in the cell) and accessory genes (encode taxa-specific functions and are loosely distributed across lineages) (Andersson, 2009). Genes that have been transferred between closely related species might be selectively neutral in the recipient organism (Gogarten & Townsend, 2005; Soucy, Huang, & Gogarten, 2015). Nonetheless, a transferred gene must provide a beneficial advantage to itself or the recipient to survive in the recipient lineage over generations (Soucy et al., 2015). HGT events can occur in prokaryotes and eukaryotes (Andersson, 2009; Zhaxybayeva & Doolittle, 2011). However, HGT events and why this happens are far better understood in prokaryotes (Zhaxybayeva & Doolittle, 2011). HGT does not follow the typical pattern of vertical descent and derives in phylogenetic incongruences, which are helpful to detect HGT events (Zhaxybayeva & Doolittle, 2011). Besides, traces of the mechanism of transfer, such as independently replicating plasmids or mobile genetic elements may confirm HGT events in prokaryotes (Zhaxybayeva & Doolittle, 2011).

Notung employs a duplication-transfer-loss (DTL) event model that minimizes transfers with duplication and losses (i.e. DTL score) and infers all most parsimonious event histories. Furthermore, Notung analyzes all event histories for temporal feasibility (Darby et al., 2017; Durand et al., 2006; Stolzer et al., 2012; Vernot et al., 2008). In Notung, transfers are represented by a yellow node and an edge while duplications are represented by a red circle on a node. Green circles on nodes represents the existence of multiple optimal solutions.

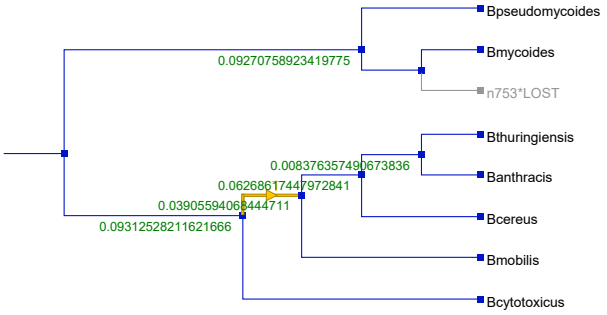
In prokaryotes, key mediators of HGT events are conjugative plasmids, integrated prophages, integrative transposons, GEIs (genomic islands are blocks of DNA with signature of genetic mobile elements), and other unclassified elements (Bellanger, Payot, Leblond-Bourget, & Guédon, 2014; Burrus, Pavlovic, Decaris, & Guédon, 2002; Hacker & Carniel, 2001). Together, these elements contribute to the prokaryotic horizontal gene pool (HGP) (Hacker & Carniel, 2001; Osborn & Böltner, 2002). Several authors suggest that GEIs help in microbial evolution and adaptation and are strongly selected for adaptive and auxiliary functions (Dobrindt et al., 2000; Hacker & Carniel, 2001; Hacker & Kaper, 2000; Juhas et al., 2009).

Spore coat genes that displayed evidence of HGT as reconciled phylogenetic trees are shown in Figures 3.3, 3.4, 3.5, 3.6, 3.7, and 3.8 for the Cereus, Coagulans, Halodurans, Methanolicus, Pumilus, and Subtilis groups, respectively. Only one optimal solution is shown below.

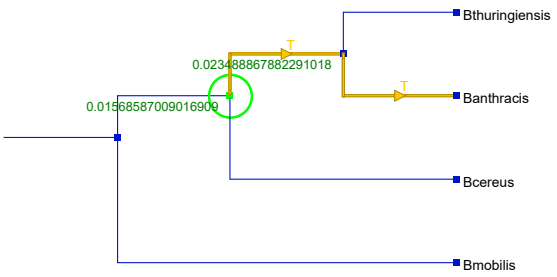
The Cereus group has sixteen spore coat genes that have undergone HGT events (Figure 3.3), according to Notung. For instance, *cotE* (a) has probably been acquired from *B. mobilis* to *B. anthracis* or vice versa. *cotJA* (b) has probably been acquired from *B. cytotoxicus* to the paraphyletic group comprised by *B. anthracis, B. cereus, B. mobilis,* and *B. thuringiensis*. *cotJC* (c) present two subsequent HGT events, from *B. cereus* to *B. anthracis* and *B. thuringiensis*. *cotK-sspO* (d) has probably been laterally transferred from *B. anthracis* to *B. bombysepticus*. *cotS* (e) and *gerPA* (f) have been transferred from *B. thuringiensis* to *B. cytotoxicus* and *B. anthracis*, respectively. *gerPB* (g) suffered two probable subsequent HGT events from *B. toyonensis* and *B. mobilis* to *B. thuringiensis* and *B. anthracis*. *gerPC* (h) presents three probable HGT events from *B. toyonensis* to the group of *B. mobilis, B. anthracis*, and *B. thuringiensis* and from *B. cytotoxicus* to *B. pseudomycoides*. *gerPE* (i) and *gerQ* (k) have three probable HGT events, from *B. cytotoxicus* and from *B. pseudomycoides*, respectively. *gerPF* (j) has been probably transferred from B. mobilis to the common ancestor of *B. thuringiensis* and *B. anthracis*. *spoIVA* (l) has probably been transferred from *B. pseudomycoides* to *B. cytotoxicus*. *yhbA-queG* (m) has suffered one lateral transfer from *B. anthracis* to *B. mobilis* and one duplication event (red node). *yhcN* (n), *ysxE* (p) and *yutH-cotNH* (q) has probably been transferred from the common ancestor of *B. anthracis* and *B. mobilis* to *B. toyonensis* and from *B. mobilis* to *B. anthracis*. *yheD* has been transferred from *B. thuringiensis* to *B. anthracis*.



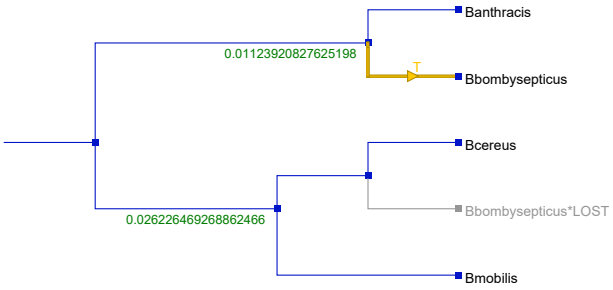
(a) *cotE*



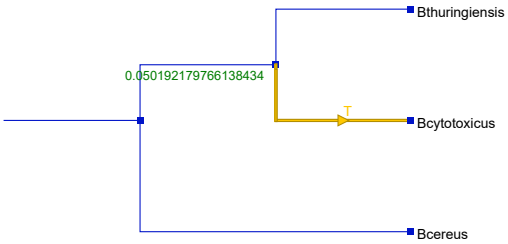
(b) *cotJA*



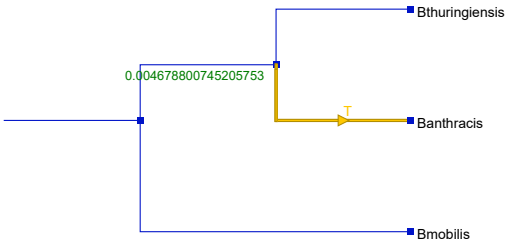
(c) *cotJC*



(d) *cotK-sspO*

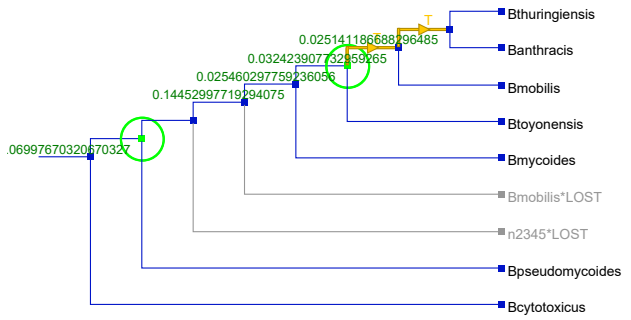


(e) *cotS*

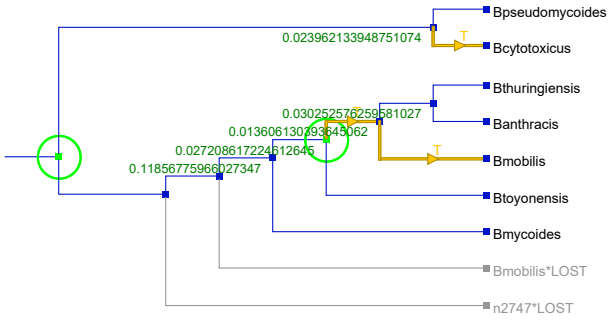


(f) *gerPA*

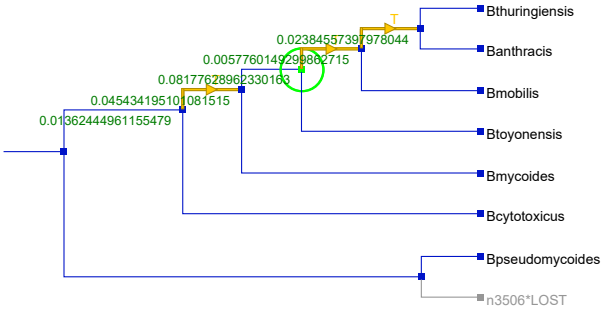
Figure 3.3. Reconciled phylogenetic trees between gene and species trees displaying HGT events of coat genes across different genomes of the Cereus group, according to Notung DTL models. HGT events are represented with a yellow node and an edge for the recipient species.



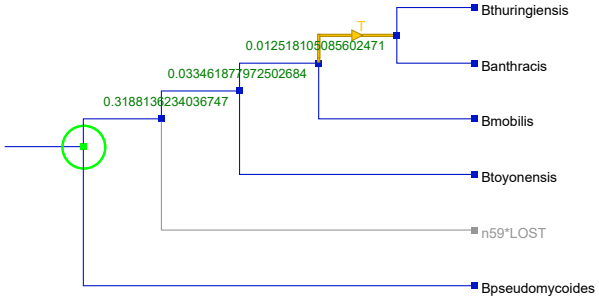
(g) *gerPB*



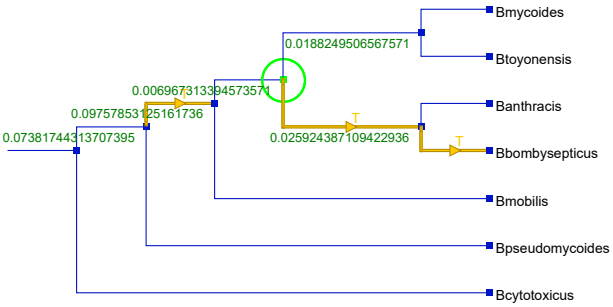
(h) *gerPC*



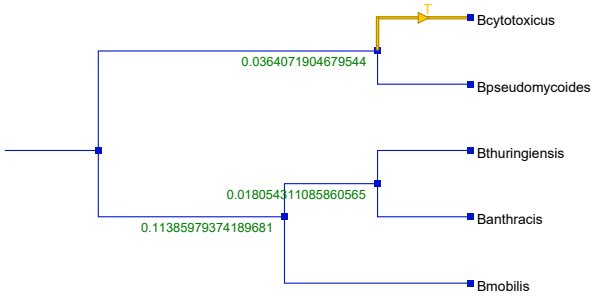
(i) *gerPE*



(j) *gerPF*

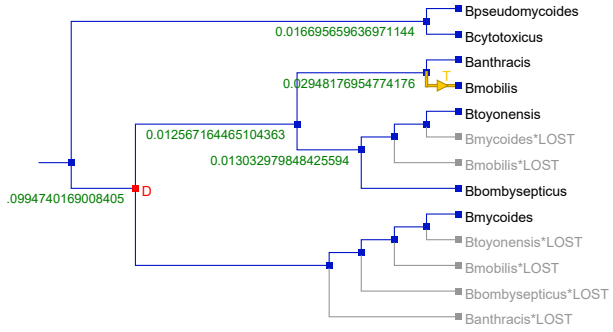


(k) *gerQ*

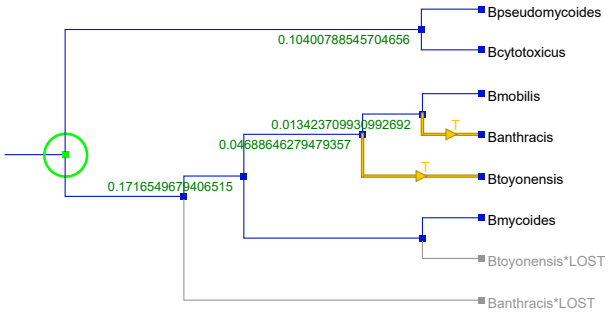


(l) *spoIVA*

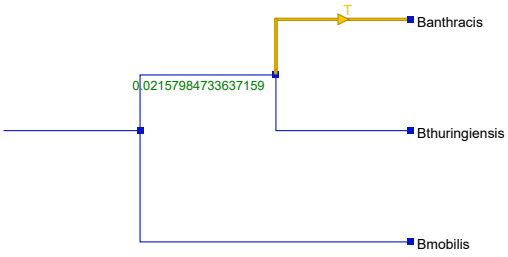
Figure 3.3. Reconciled phylogenetic trees between gene and species trees displaying HGT events of coat genes across different genomes of the Cereus group, according to Notung DTL models. HGT events are represented with a yellow node and an edge for the recipient species. (*Continued*)



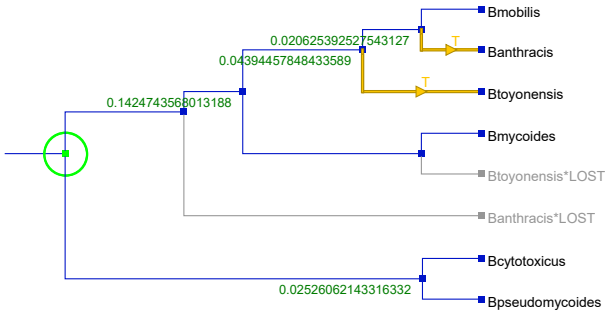
(m) *yhbA-queG*



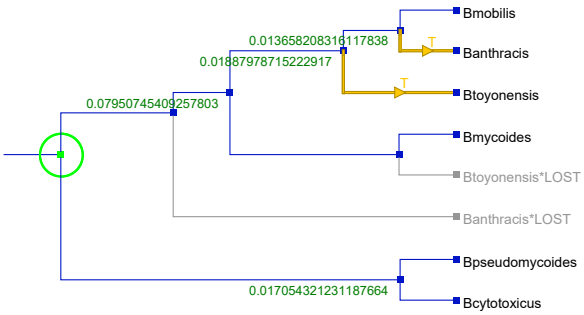
(n) *yhcN*



(o) *yheD*



(p) *ysxE*



(q) *yutH-cotNH*

Figure 3.3. Reconciled phylogenetic trees between gene and species trees displaying HGT events of coat genes across different genomes of the Cereus group, according to Notung DTL models. HGT events are represented with a yellow node and an edge for the recipient species. (*Continued*)

The Coagulans group has one probable HGT event for *gerQ* from *B. thermoamylovorans* to *B. coagulans*, see Figure 3.4. The Halodurans group presents two probable HGT events for *yhbA-queG*, see Figure 3.5. The first one, from the common ancestor of *B. pseudofirmus, B. halodurans*, and *B. krulwichiae* to *B. cellulosilyticus*. The second, from *B. krulwichiae* to *B. halodurans*. The Megaterium group does not present HGT events.

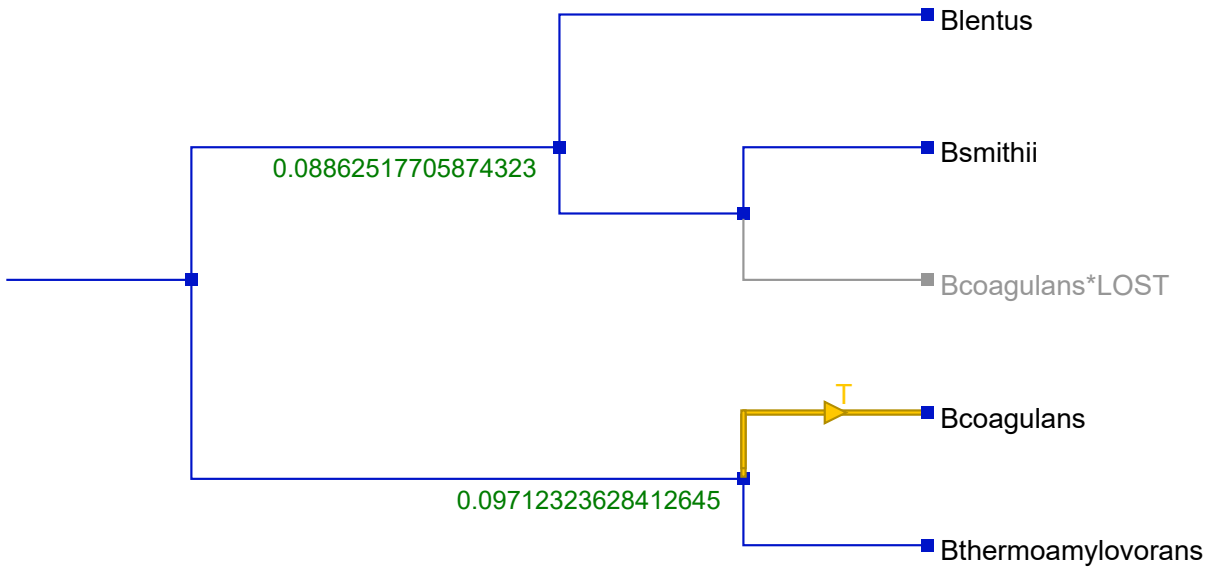


Figure 3.4. Reconciled phylogenetic tree between gene and species trees displaying HGT events of *gerQ* across different genomes of the Coagulans group, according to Notung DTL models. HGT events are represented with a yellow node and an edge for the recipient species.

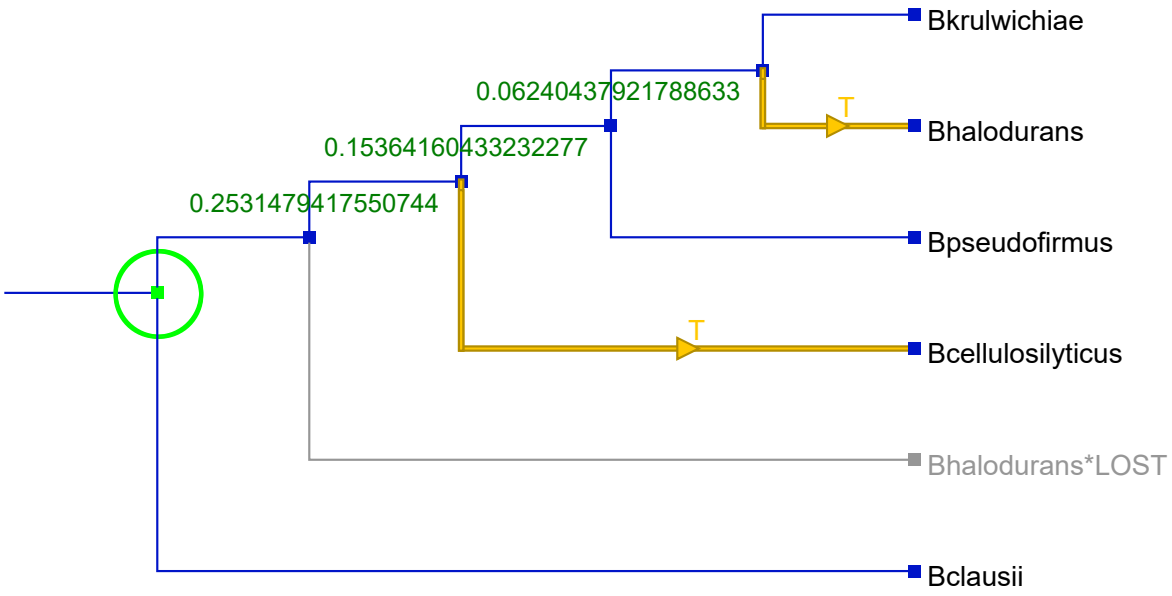
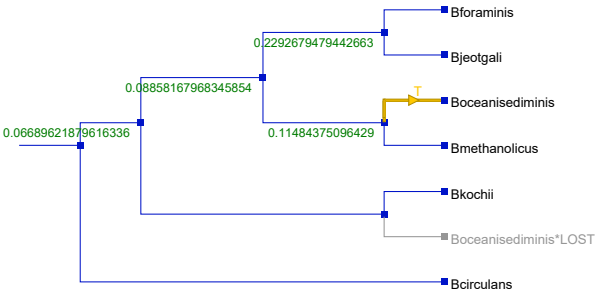
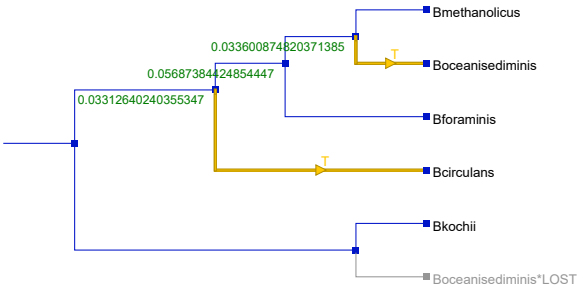


Figure 3.5. Reconciled phylogenetic tree between gene and species trees displaying HGT events of *yhbA-queG* across different genomes of the Halodurans group, according to Notung DTL models. HGT events are represented with a yellow node and an edge for the recipient species.

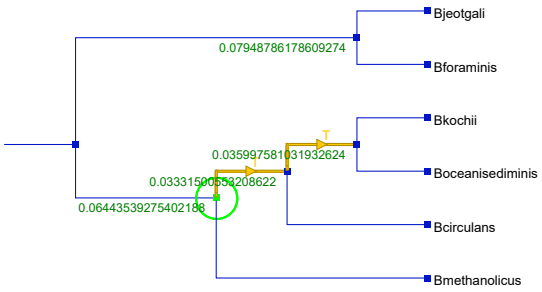
The Methanolicus group has four spore coat genes that have undergone HGT events, see Figure 3.6. *gerPE* (a), from *B. methanolicus* to *B. oceanisediminis*. *yhcN* (b) presents two probable HGT events, from the common ancestor of *B. foraminis, B. oceanisediminis*, and *B. methanolicus* to *B. circulans* and from *B. methanolicus* to *B. oceanisediminis*. *ysxE* (c) has been probably transferred to *B. methanolicus* to the common ancestor of *B. circulans, B. oceanisediminis*, and *B. jeotgali*.



(a) *gerPE*



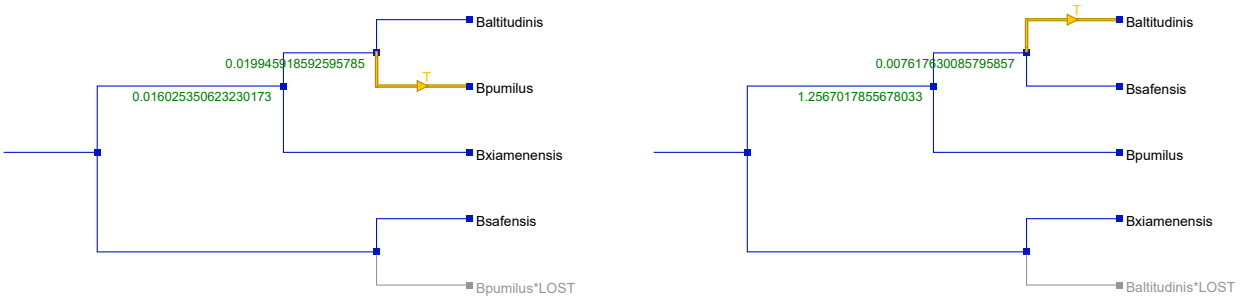
(b) *yhcN*



(c) *ysxE*

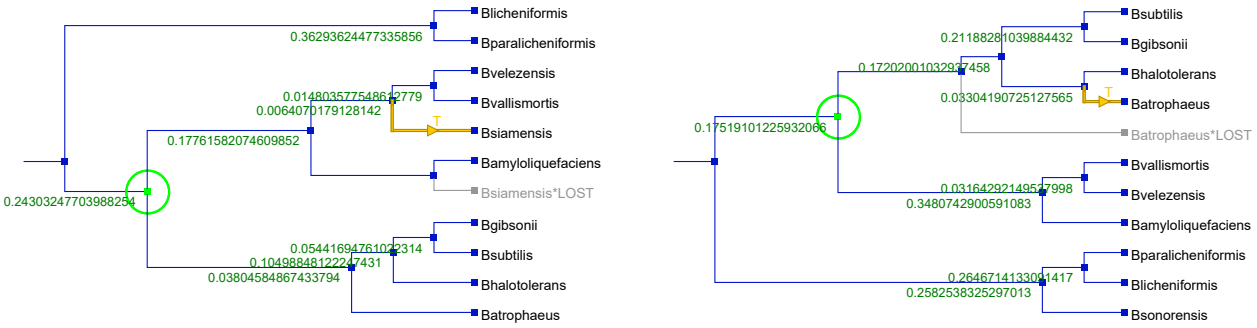
Figure 3.6. Reconciled phylogenetic trees between gene and species trees displaying HGT events of coat genes across different genomes of the Methanolicus group, according to Notung DTL models. HGT events are represented with a yellow node and an edge for the recipient species.

Figure 3.7, shows two spore coat genes that have undergone HGT in the Pumilus group. *gerPC* (a) has been transferred from *B. altitudinis* to *B. pumilus* and *yhcN* (b) from *B. safensis* to *B. altitudinis*. The Subtilis group has two spore coat genes under HGT events, see Figure 3.8. *gerPC* (a) has been probably transferred from the common ancestor of *B. velezensis* and *B. vallismortis* to *B. siamensis*. *yhcN* (b) has probably been transferred from *B. halotolerans* to *B. atrophaeus*. All these HGT events are further confirmed by ICEs using WU-BLAST2 of the web server ICEBerg, see Table 5. Nevertheless, the GIPSy software shows that spore coat genes are not present on GEIs.



(a) *gerPC* (b) *yhcN*

Figure 3.7. Reconciled phylogenetic trees between gene and species trees displaying HGT events of coat genes across different genomes of the Pumilus group, according to Notung DTL models. HGT events are represented with a yellow node and an edge for the recipient species.



(a) *gerPC* (b) *yhcN*

Figure 3.8. Reconciled phylogenetic trees between gene and species trees displaying HGT events of coat genes across different genomes of the Subtilis group, according to Notung DTL models. HGT events are represented with a yellow node and an edge for the recipient species.

Table 3.3 Coat genes that have undergone HGT events and its associated scanned genome regions with WU-BLAST2 for ICEs element

Species under HGT	**Coat gene**	**Genome location$^{\lambda}$**	**Region scanned$^{\mu}$**	**ICEberg WU-BLAST2***
		Cereus group		
B. anthracis	*cotE*	3580099-3580641	3566099-3591641	Clostridium perfringens plasmid pJIR2774 insertion sequence tISCpe8 transposase (tnp) and lincosamide nucleotidyltransferase genes (FJ589781)
B. anthracis	*cotJA*	821562-821777	812562-829500	Enterococcus faecium transfer chromosomal element, MM5-F9a (HE963029)
B. cereus		806934-807149	797934-814900	
B. mobilis		858664-858879	849664-866879	
B. thuringiensis		819912-820127	810912-827127	

$^{\lambda}$Location of coat gene in the genome of the species under HGT

$^{\mu}$Scanned region, approximately 10 genes upstream and downstream from the coat gene. Used for WU-BLAST2 in ICEberg

*Hits significant at p-value < 0.001

Table 3.3 Coat genes that have undergone HGT events and its associated scanned genome regions with WU-BLAST2 for ICEs element (Continued)

Species under HGT	**Coat gene**	**Genome location$^{\lambda}$**	**Region scanned$^{\mu}$**	**ICEberg WU-BLAST2***
B. anthracis	*cotJC*	820708-821277	809708-826277	Enterococcus faecium MM5-F9a transfer chromosomal element, (HE963029) Streptococcus agalactiae 2584 DDE transposase UPF0236 family (KC492042)

$^{\lambda}$Location of coat gene in the genome of the species under HGT

$^{\mu}$Scanned region, approximately 10 genes upstream and downstream from the coat gene. Used for WU-BLAST2 in ICEberg

*Hits significant at p-value < 0.001

Table 3.3 Coat genes that have undergone HGT events and its associated scanned genome regions with WU-BLAST2 for ICEs element (Continued)

Species under HGT	**Coat gene**	**Genome location**$^{\lambda}$	**Region scanned**$^{\mu}$	**ICEberg WU-BLAST2***
B. thuringiensis		819058-819627	808400-824400	Enterococcus faecium MM5-F9a transfer chromosomal element, (HE963029) Streptococcus agalactiae 2584 DDEmtransposase UPF0236 family (KC492042) Streptococcus pneumoniae DP1322 conjugative transposon Tn5253 (EU351020) Streptococcus pneumoniae SpnA213 partial Tn5253-like transposon (FM201786)
B. bombysepticus	*cotK-sspO*	3342593-3342742	333593-3350742	Streptococcus pneumoniae ATCC700669 (FM211187) region (922465..951058)

$^{\lambda}$Location of coat gene in the genome of the species under HGT

$^{\mu}$Scanned region, approximately 10 genes upstream and downstream from the coat gene. Used for WU-BLAST2 in ICEberg

*Hits significant at p-value < 0.001

Table 3.3 Coat genes that have undergone HGT events and its associated scanned genome regions with WU-BLAST2 for ICEs element (Continued)

Species under HGT	**Coat gene**	**Genome location$^{\lambda}$**	**Region scanned$^{\mu}$**	**ICEberg WU-BLAST2***
B. cytotoxicus	*cotS*	2182130-2183194	2173130-2193565	Vibrio alginolyticus HN437 transposon integrating conjugative element ICEValHN437 (KT072771)
B. anthracis	*gerPA*	1121477-1121698	1110809-1132132	Mesorhizobium loti R7A symbiosis island (AL672115; AL672114; AL672113; AL672112)
B. anthracis	*gerpPB*	1121256-1121462	1109954-1130625	Mesorhizobium loti R7A symbiosis island (AL672115; AL672114; AL672113; AL672112)
B. mobilis		1172297-1172503	1161632-1181787	
B. thuringiensis		1143593-1143745	1132413-1152913	
B. anthracis	*gerPC*	1120575-1121189	1109487-1129099	Bordetella petrii DSM12804 (AM902716), region(1350129..1493558)
B. cytotoxicus		983499-984119	970890-993293	
B. mobilis		1171614-1172228	1160870-1180261	
B. thuringiensis		1142857-1143471	1131811-1151387	
B. anthracis	*gerPE*	1119972-1120358	1108881-1125865	Lactococcus lactis KF147 (CP001834) region(2295682..2347036)
B. mobilis		1171011-1171397	1159870-1176907	NA

$^{\lambda}$Location of coat gene in the genome of the species under HGT

$^{\mu}$Scanned region, approximately 10 genes upstream and downstream from the coat gene. Used for WU-BLAST2 in ICEberg

*Hits significant at p-value < 0.001

Table 3.3 Coat genes that have undergone HGT events and its associated scanned genome regions with WU-BLAST2 for ICEs element (Continued)

Species under HGT	**Coat gene**	**Genome location$^{\lambda}$**	**Region scanned$^{\mu}$**	**ICEberg WU-BLAST2***
B. mycoides		4595797-4596183	4590293-4608219	NA
B. toyonensi		3718076-3718462	3706829-3723971	Lactococcus lactis KF147 (CP001834) region(2295682..2347036)
B. thuringiensis		1142254-1142640	1128584-1148154	Bacteroides fragilis 86-5443-2-2 conjugative transposon CTn86 (AY372755; AY375536)
B. anthracis	*gerPF*	1119714-1119929	1107741-1125252	Lactococcus lactis KF147 (CP001834) region(2295682..2347036)
B. thuringiensis		1141996-1142211	1127593-1147541	Bacteroides fragilis 86-5443-2-2 conjugative transposon CTn86 (AY372755; AY375536)
B. anthracis		5126358-5126795	5117024-5134933	Clostridium difficile 630
B. bombysepticus		5042319-5042747	5032821-5050877	(AM180355) region(428851..453332)
B. mobilis	*gerQ*	5355062-5355490	5344998-5363855	Clostridium difficile CII7
B. mycoides		596427-596855	588805-606331	Tn6194-like conjugative transposon
B. toyonensis		2518784-2519212	2509763-2527422	(HG475346)

$^{\lambda}$Location of coat gene in the genome of the species under HGT

$^{\mu}$Scanned region, approximately 10 genes upstream and downstream from the coat gene. Used for WU-BLAST2 in ICEberg

*Hits significant at p-value < 0.001

Table 3.3 Coat genes that have undergone HGT events and its associated scanned genome regions with WU-BLAST2 for ICEs element (Continued)

Species under HGT	**Coat gene**	**Genome location$^{\lambda}$**	**Region scanned$^{\mu}$**	**ICEberg WU-BLAST2***
B. cytotoxicus	*spoIVA*	1342603-1344081	1334855-1353902	Mesorhizobium loti MAFF303099, (BA000012) region(4643427..5255769) Finegoldia magna ATCC 29328 (AP008971), region(1002528..1069524) Clostridium difficile 630, (AM180355), region(2137789..2181291)
B. mobilis	*yhbA-queG*	550034-551176	543713-569062	Streptococcus equi 4047, (FM204883), region(1206317..1269371) Mycoplasma bovis PG45 MU clone A2, (CP002188), region(210737..248144)
B. anthracis	*yhcN*	2560329-2560925	2551657-2568862	Streptococcus equi 4047, (FM204883), region(1206317..1269371)
B. mobilis		2812161-2812757	2804034-2820837	NA
B. toyonensis		48613-49209	40021-57668	NA

$^{\lambda}$Location of coat gene in the genome of the species under HGT

$^{\mu}$Scanned region, approximately 10 genes upstream and downstream from the coat gene. Used for WU-BLAST2 in ICEberg

*Hits significant at p-value < 0.001

Table 3.3 Coat genes that have undergone HGT events and its associated scanned genome regions with WU-BLAST2 for ICEs element (Continued)

Species under HGT	**Coat gene**	**Genome location$^{\lambda}$**	**Region scanned$^{\mu}$**	**ICEberg WU-BLAST2***
B. anthracis	*yheD*	866934-868394	859264-877366	Mesorhizobium loti R7A symbiosis island, (AL672115; AL672114; AL672113; AL672112) Finegoldia magna ATCC 29328, (AP008971), region(1002528..1069524) Clostridium difficile 630, (AM180355), region(480597..520127) Proteus mirabilis HI4320, (AM942759), region(2793762..2886224) Streptococcus equi 4047, (FM204883), region(1206317..1269371)
B. anthracis		4266676-4267701	4256239-4276956	Vibrio alginolyticus HN437
B. mobilis	*ysxE*	4476574-4477599	4466142-4486906	transposon integrating conjugative
B. toyonensis		1660842-1661867	1650352-1671268	element ICEValHN437, (KT072771)

$^{\lambda}$Location of coat gene in the genome of the species under HGT

$^{\mu}$Scanned region, approximately 10 genes upstream and downstream from the coat gene. Used for WU-BLAST2 in ICEberg

*Hits significant at p-value < 0.001

Table 3.3 Coat genes that have undergone HGT events and its associated scanned genome regions with WU-BLAST2 for ICEs element (Continued)

Species under HGT	**Coat gene**	**Genome location$^{\lambda}$**	**Region scanned$^{\mu}$**	**ICEberg WU-BLAST2***
B. anthracis	*yutH-cotNH*	4711811-4712812	4703947-4719574	Streptococcus pneumoniae TIGR4 (AE00567), region(972255..1000373) Streptococcus pneumoniae PN1 transposon ICESpPN1, (KC488257) Finegoldia magna ATCC29328 (AP008971), region(1002528..1069524) Enterococcus faecium transfer chromosomal element, (HE963029)
B. mobilis		4923721-4924722	4917335-4934447	Streptococcus equi 4047, (FM204883), region(1206317..1269371) Streptococcus pneumoniae DP1322 conjugative transposon Tn5253, (EU351020)

$^{\lambda}$Location of coat gene in the genome of the species under HGT

$^{\mu}$Scanned region, approximately 10 genes upstream and downstream from the coat gene. Used for WU-BLAST2 in ICEberg

*Hits significant at p-value < 0.001

Table 3.3 Coat genes that have undergone HGT events and its associated scanned genome regions with WU-BLAST2 for ICEs element (Continued)

Species under HGT	**Coat gene**	**Genome location$^{\lambda}$**	**Region scanned$^{\mu}$**	**ICEberg WU-BLAST2***
B. toyonensis		2116161-2117162	2110191-2125720	Streptococcus pneumoniae TIGR4 (AE00567), region(972255..1000373) Streptococcus pneumoniae PN1 transposon ICESpPN1, (KC488257)
Coagulans group				
B. coagulans	*gerQ*	3082-3531	NZ_ATVM01000123	Cupriavidus metallidurans CH34 (CP000352), region(1585489..1682531) Clostridium perfringens plasmid pJIR4150, (LN835295), region(52023..84499)
Halodurans group				

$^{\lambda}$Location of coat gene in the genome of the species under HGT

$^{\mu}$Scanned region, approximately 10 genes upstream and downstream from the coat gene. Used for WU-BLAST2 in ICEberg

*Hits significant at p-value < 0.001

Table 3.3 Coat genes that have undergone HGT events and its associated scanned genome regions with WU-BLAST2 for ICEs element (Continued)

Species under HGT	**Coat gene**	**Genome location$^{\lambda}$**	**Region scanned$^{\mu}$**	**ICEberg WU-BLAST2***
B. cellulosilyticus	*yhbA-queG*	1529457-1530596	1518291-1543013	Bacteroides fragilis conjugative transposon CTn341, (AY515263) Streptococcus equi subsp. zooepidemicus H70, (FM204884), region(1399879..1447451) Enterococcus faecium D344R transposon Tn5386, (DQ321786)
B. halodurans		1100040-1101215	1088909-1112360	Streptococcus pneumoniae TIGR4 (AE005672), region(972255..1000373)
B. krulwichiae		709582-710721	698822-721478	NA
B. pseudofirmus		637661-638818	627135-650159	Mesorhizobium loti MAFF303099 (BA000012), region(4643427..5255769) Mesorhizobium loti R7A symbiosis island (AL672115; AL672114; AL672113; AL672112)

$^{\lambda}$Location of coat gene in the genome of the species under HGT

$^{\mu}$Scanned region, approximately 10 genes upstream and downstream from the coat gene. Used for WU-BLAST2 in ICEberg

*Hits significant at p-value < 0.001

Table 3.3 Coat genes that have undergone HGT events and its associated scanned genome regions with WU-BLAST2 for ICEs element (Continued)

Species under HGT	**Coat gene**	**Genome location$^{\lambda}$**	**Region scanned$^{\mu}$**	**ICEberg WU-BLAST2***
		Methanolicus group		
B. oceanisediminis	*gerPE*	1404524-1404922	1391304-1414990	Bordetella petrii DSM 12804 (AM902716), region(1350129..1493558, 2250540..2297731)
B. circulans	*yhcN*	621937-622545	613226-631551	Acidithiobacillus caldus ATCC 51756, (CP005986), region(1803307..1980282) Streptococcus intermedius transposon Tn916S tetracycline resistance protein (AY534326) Enterococcus casseliflavus 664.1H1 conjugative transposon Tn6000, (FN555436)
B. foraminis		324412-325119	314019-332292	Bacteroides fragilis HMW 615 transposon CTnHyb, (KJ816753) Bacteroides fragilis conjugative transposon CTn341, (AY515263)

$^{\lambda}$Location of coat gene in the genome of the species under HGT

$^{\mu}$Scanned region, approximately 10 genes upstream and downstream from the coat gene. Used for WU-BLAST2 in ICEberg

*Hits significant at p-value < 0.001

Table 3.3 Coat genes that have undergone HGT events and its associated scanned genome regions with WU-BLAST2 for ICEs element (Continued)

Species under HGT	**Coat gene**	**Genome location**$^{\lambda}$	**Region scanned**$^{\mu}$	**ICEberg WU-BLAST2***
B. methanolicus		1669879-1670409	1660167-1680217	Acidithiobacillus caldus SM-1 (NC_015850), region(115486..212210) Streptococcus agalactiae 2603V/R, (AE009948), region(1962397..2005294)
B. oceanisediminis		2562095-2562649	2550776-2572954	Vibrio alginolyticus A056 transposon ICEValA056-2, (KR231689) Streptococcus pneumoniae Tn5252 like transposon, (FM178797)
B. circulans	*ysxE*	2244158-2245210	2229203-2255470	Streptococcus pyogenes integrative conjugative element ICESp1108, (FR691054), region(1694..47149)

$^{\lambda}$Location of coat gene in the genome of the species under HGT

$^{\mu}$Scanned region, approximately 10 genes upstream and downstream from the coat gene. Used for WU-BLAST2 in ICEberg

*Hits significant at p-value < 0.001

Table 3.3 Coat genes that have undergone HGT events and its associated scanned genome regions with WU-BLAST2 for ICEs element (Continued)

Species under HGT	**Coat gene**	**Genome location$^{\lambda}$**	**Region scanned$^{\mu}$**	**ICEberg WU-BLAST2***
B. kochii		3322545-3323612	3312533-3336376	Streptococcus pneumoniae integrative and conjugative element ICESpn529IQ, (HG965092), region (1..59466) Streptococcus pneumoniae DP1322 conjugative transposon Tn5253, (EU351020)
B. oceanisediminis		4318702-4319760	4307085-4330368	Mesorhizobium loti R7A symbiosis island (AL672115; AL672114; AL672113; AL672112)
			Pumilus group	
B. pumilus	*gerPC*	1112115-1112729	1097904-1121324	Bordetella petrii DSM12804 (AM902716), region(1350129..1493558)
B. altitudinis	*yhcN*	904667-905173	897238-918346	Lactococcus lactis KF147 (CP001834), region(2295682..2347036)
			Subtilis group	

$^{\lambda}$Location of coat gene in the genome of the species under HGT

$^{\mu}$Scanned region, approximately 10 genes upstream and downstream from the coat gene. Used for WU-BLAST2 in ICEberg

*Hits significant at p-value < 0.001

Table 3.3 Coat genes that have undergone HGT events and its associated scanned genome regions with WU-BLAST2 for ICEs element (Continued)

Species under HGT	**Coat gene**	**Genome location**$^{\lambda}$	**Region scanned**$^{\mu}$	**ICEberg WU-BLAST2***
B. siamensis	*gerPC*	3190461-3191078	3182933-3205503	Bordetella petrii DSM12804 (AM902716), region(1350129..1493558)
B. atrophaeus	*yhcN*	1017478-1018062	1009284-1032112	Lactococcus lactis KF147 (CP001834), region(2295682..2347036)

$^{\lambda}$Location of coat gene in the genome of the species under HGT

$^{\mu}$Scanned region, approximately 10 genes upstream and downstream from the coat gene. Used for WU-BLAST2 in ICEberg

*Hits significant at p-value < 0.001

CHAPTER 4. DISCUSION

In this work, we reported the existence of several spore coat protein homologs across one hundred forty-one genomes of spore-forming species of the *Bacillales* order. Our results show that a significant majority of the ninety-seven spore coat proteins are highly conserved among the forty genomes of Bacillus genus (Figure 3.1). As reported by J. A. Castillo (personal communication, 2019) and NCBI taxonomy, *B. amyloliquefaciens, B. velezensis, B. atrophaeus, B. licheniformis, B. paralicheniformis, B. sonorensis, B. subtilis, B. vallismortis*, and *B. gibsonii* constitute a monophyletic group inside the *Bacillus* genus, known as the Subtilis group. Bacterial species closely related to *B. subtilis* have few phenotypic differences that distinguish them (Nakamura, Roberts, & Cohan, 1999; Roberts, Nakamura, & Cohan, 1994) but can be easily distinguished by fatty acid composition analysis, restriction digest analysis and DNA–DNA hybridization analysis (L.-T. Wang, Lee, Tai, & Kasai, 2007). *B. subtilis* grows in close association with decaying organic matter in the soil, plant root surfaces, and inside the gastrointestinal tract of terrestrial and marine organisms as probiotics (Earl, Losick, & Kolter, 2008; Siala, Hill, & Gray, 1974; Tam et al., 2006). However, highly resistant *B. subtilis* spores can be found in many aquatic and terrestrial environments because they can be carried away by wind across different environments (Earl et al., 2008). Hence, those environments may not reflect real ecological niches inhabited by vegetative cells (Earl et al., 2008).

Figure 3.1 shows that the Subtilis group possess the most conserved spore coat proteins compared to other Bacillus groups and non-Bacillus spore-forming species. Spore coat proteins, such as CotE, CotH, CotO, CotX, CotY, CotZ, SafA, SpoIVA, SpoVM, SpoVID (morphogenetic coat proteins), CotJA, CotJB, CotJC, LipC, YhaX, YheD, YjzB, YppG (basement layer), CotD, CotF, CwlJ, GerQ, OxdD, Tgl, YaaH, YhjR, YisY, YjqC, YsxE, YutH, YuzC, YxeE (inner layer), CotA, CotG, CotM, GerT, SpsB, SpsI, YknT, YlbD, YncD, YkzQ, YtdA (outer layer), CgeA, CgeB, CotV, CotW (crust), CotI, CotR, CotSA, GerPA, GerPB, GerPC, GerPD, GerPE, GerPF, YdhD, YhbB, YheC, YkvP. YkvQ, CotN, and YpzA (localization not determined) are

widely distributed among members of the Bacillus group. Thus, our results reveal a genomic signature of spore coat genes in the Subtilis group that confers spore resistance to different ecological niches, as previously demonstrated (Galperin, 2013; Galperin et al., 2012; Onyenwoke et al., 2004).

The Cereus group composed of *B. anthracis, B. bombysepticus, B. cereus, B. cytotoxicus, B. mycoides, B. thuringienses*, and *B. toyonensis*. *B. cereus* is an aerobic to facultative anaerobic opportunistic pathogen that can survive in a diversity of habitats, such as low nutrient soil and the intestinal flora of various animals (Kotiranta et al., 2000; Stenfors Arnesen et al., 2008). The highly adhesive endospore of *B. cereus* allows it to attach to different types of food causing food-borne illnesses (Dierick et al., 2005; Stenfors Arnesen et al., 2008). *B. anthracis* is the etiological agent of anthrax, and it is an aerobic to facultative anaerobic gram positive bacteria. The entire life cycle of *B. anthracis* takes place within a mammalian host compared to other members of the Cereus group that share the same ecological niche (Mock & Fouet, 2001). *B. thuringiensis* is an insect pathogen that produces insecticidal protoxins during sporulation to exploit nutrient-rich insect larvae (Aronson & Shai, 2001; Zothansanga, Senthilkumar, & Gurusubramanian, 2016). *B. mycoides* is an aerobic spore-forming species that tend to form rhizoidal-shape colonies (Nakamura, 1998). *B. cytotoxicus* is a thermotolerant species compared to *B. mycoides* (Bağcıoğlu, Fricker, Johler, & Ehling-Schulz, 2019; Stevens et al., 2019).

Morphogenetic spore coat proteins are highly conserved in the Cereus group, except SpoVM and CotX involved in the spore cortex formation, spore assembly, spore encasement, and crust assembly (Driks & Eichenberger, 2016). Since coat assmbly is a highly hierarchical process (Driks & Eichenberger, 2016), other morphogenetic proteins present with the same role, such as SpoIVA, CotY, CotZ may take the over the task. Thus, compensating the absence of SpoVM and CotX. Some spore coat proteins (YuzC, YybI, YeeK) that are part of the inner layer are absent. Moreover, several spore coat proteins present in the outer layer are absent despite the presence of the morphogenetic coat proteins SpoIVA and CotE.

B. simplex is a psychrophilic and soil bacterial species that can exert biocontrol activities in microbial pathogens (Abe, Koyama, Kawamura, & Koseki, 2019; Rosenberg et al., 2016;

Tatu, Clatici, & Cristea, 2016). *B. muralis* is a bacterial species closely related to *B. simplex* (J.-P. Wang et al., 2016). In both bacterial species that compose the Simplex group, several spore coat protein homologs of the crust, inner layer, and outer layer are absent. This is not surprising given the absence of the morphogenetic coat proteins CotO, CotY, CotZ that control those processes (Driks & Eichenberger, 2016).

B. pumilus is naturally found in the soil and the plant root systems of tobacco, pepper, cucumber, and tomato (Stepanov et al., 2016; Val-Calvo et al., 2019). Moreover, *B. pumilus* spores are highly resilient to radiation, desiccation, and hydrogen peroxide treatment (Stepanov et al., 2016). Likewise, spores of *B. altitudinis* are highly resistant to UV-light and H_2O_2 due to the presence of the spore coat proteins CotA and YjqC (Y. Zhang et al., 2016). Despite some spore coat proteins of the outer layer and inner layer are absent in the Pumilus group, protein homologs for all the morphogenetic coat proteins are present. Thus, a proper assembly of the spore coat is highly conserved in this group, which is beneficial for the higher spore resilience previously reported (Stepanov et al., 2016; Y. Zhang et al., 2016).

B. methanolicus is methylotropic and obligate aerobic endospore-forming bacteria that can be used for methanol-based biotechnologies (Arfman et al., 1992; Carnicer, Vieira, Brautaset, Portais, & Heux, 2016). *B. oceanisediminis* is an aerobic bacterium isolated from marine sediments that closely related to *B. infantis*. Its spores are resistant to environmental stressors, such as H_2O_2, UV-radiation, heat, and space vacuum (Lee et al., 2012; J. Zhang et al., 2010). These bacterial species compose the Methanolicus group. Our results indicate that the morphogenetic coat proteins CotO, CotH, CotX, and other spore coat proteins of the crust and outer layer are absent in this group.

B. coagulans is a facultative anaerobic, nonpathogenic, lactic acid-producing, and probiotic agent that is capable of form endospores to survive extreme environmental conditions, such as high temperature and acidity (Konuray & Erginkaya, 2018; Saw et al., 2019). *B. smithii* can ferment carbon sources into lactate, and it is a close relative to *B. coagulans* (Bosma et al., 2016). Both species correspond to the Coagulans group and have been isolated from a wide variety of environments, such as the human gut and marine sediments (Bosma et al., 2015; Suitso et al., 2010). The morphogenetic coat proteins CotH, CotX, CotO, CotY, and CotZ responsible

for the assembly of the outer layer and the crust are absent in this group. As expected, several spore coat proteins of the outer layer dependent on CotH and CotO are also absent. Whereas proteins dependent on CotX, CotY, and CotZ are completely absent in this group. This implies that the crust may be absent in this group. Further studies are needed to confirm this idea.

B. megaterium is an endophytic and aerobic spore-forming species with a volume of 60 μm^3 that can be found in the soil, sediments, and dried food (Vary et al., 2007). Spores of this species have antioxidant activity and can be used as a probiotic product (Mazzoli et al., 2019). *B. flexus* is an aerobic bacterium used for the production of polyhydroxyalkanoates and as an arsenic transformer (Jebeli et al., 2017; Wagle, Dixit, & Vakil, 2019). *B. endophyticus* is an endophytic and aerobic spore-forming bacteria closely-related to *B. megaterium* (Reva, Smirnov, Pettersson, & Priest, 2002). The Megaterium group lacks the morphogenetic coat proteins CotH, CotX, CotY, and CotZ. Thus, we can expect that several spore coat proteins of the crust, inner and outer layers are absent, as confirmed by our results.

B. halodurans is an alkaliphilic bacterium species used for the production of alkaline enzymes (Takami & Horikoshi, 1999; Takami et al., 2000). The Halodurans group composed by *B. halodurans, B. cellulosilyticus, B, clausii, B. krulwichiae, B. lehensis*, and *B. pseudofirmus* do not present the morphogenetic coat proteins responsible for the assembly of the outer coat and the crust. Hence, several spore coat protein homologs dependent on those morphogenetic proteins are also absent.

According to above described results, some Bacillus groups lack the morphogenetic coat proteins CotO and CotH. Several studies have reported that CotH and CotO are responsible for the assembly of the outer coat as minor players because they are CotE-controlled (Driks & Eichenberger, 2016; McKenney et al., 2013; McPherson et al., 2005; Zilhão et al., 1999). Although CotH and CotO mutants may have a disorganized outer coat, the major assembly step is carried out by CotE and CotE-dependent coat proteins (McPherson et al., 2005; Zilhão et al., 1999). Thus, our results confirm that assembly of the outer coat is highly diverse and among the *Bacillus* genus and reflect overlapping functions of spore coat proteins to adapt to specific morphologies in response to niche variation (Driks & Eichenberger, 2016; McKenney et al., 2013; McPherson et al., 2005; J. Zhang, Fitz-James, & Aronson, 1993; Zilhão et al., 1999).

The morphogenetic coat proteins CotX, CotY, and CotZ are collectively known as the insoluble fraction of the spore because they influence spore hydrophobicity and accessibility of germinants (Krajčíková, Lukáčová, Müllerová, Cutting, & Barák, 2009; J. Zhang et al., 1993). Moreover, they are responsible for crust assembly around the spore (Driks & Eichenberger, 2016; McKenney et al., 2010). CotX, CotY, and CotZ mutants have an incomplete outer coat but the resistance to heat or lysozyme is not affected (J. Zhang et al., 1993). Hence, the absence of these morphogenetic coat proteins and their dependent-proteins in various spore-forming species, reflect overlapping functions and a spore coat protein interaction network that is highly adapted to unique environmental conditions (Driks & Eichenberger, 2016; J. Zhang et al., 1993).

On the other hand, the morphogenetic coat proteins CotE, SpoIVA, SpoVM, SpoVID, and SafA are highly conserved among all spore-forming species. CotE controls the assembly of the outer coat layer and other coat proteins, designated as CotE-controlled proteins (Driks & Eichenberger, 2016; McKenney & Eichenberger, 2012). SafA has been found to interact with SpoVID in the early stages of coat assembly (Driks & Eichenberger, 2016; McKenney & Eichenberger, 2012; Ozin et al., 2000). Furthermore, previous studies report that SpoIVA and CotE, SpoVM, and SpoVID contribute to the formation of a spore coat scaffold during earlier stages of sporulation (Bauer et al., 1999; Driks & Eichenberger, 2016; McKenney & Eichenberger, 2012). Similarly, CotE-controlled proteins, such as CotSA and YaaH (Bauer et al., 1999; Driks & Eichenberger, 2016; McKenney & Eichenberger, 2012) are conserved in all spore-forming species analyzed in this study.

The SpoIVA-dependent proteins CotJA, CotJB, and CotJC are also ubiquitous among the one hundred forty-one spore-forming species analyzed in this study. These proteins are necessary for the assembly of the basement layer of the spore coat (Driks & Eichenberger, 2016; Henriques, Beall, Roland, & Moran, 1995; Seyler, Henriques, Ozin, & Moran, 1997). Spore coat proteins whose location is unknown but have a role in germination (allowing the passage of germinants) (Butzin et al., 2012; Driks & Eichenberger, 2016), such as the GerPA-GerPF proteins are well preserved in all spore-forming species addressed here. Another protein involved in germination and highly conserved is GerQ along with CwlJ (cell wall hydrolase). GerQ is

cross-linked in the inner layer of the spore coat and is necessary for the localization of CwlJ (Driks & Eichenberger, 2016; Monroe & Setlow, 2006; Ragkousi & Setlow, 2004). In species of the *Bacillus* genus, the spore coat protein Tgl responsible for the GerQ cross-linking (Driks & Eichenberger, 2016; Monroe & Setlow, 2006; Ragkousi & Setlow, 2004) is highly conserved. This is not true for non-bacillus spore-forming species.

Our results suggest that there is a well-conserved core of spore coat proteins, mainly composed of morphogenetic coat proteins and germinant coat proteins that are essential for spore assembly in early stages and germination. In addition, spore coat proteins that directly depend on these morphogenetic and germinant proteins are also preserved. Other minor spore coat proteins that may confer a unique spore morphology are not preserved since they are unique to a group of species that share a common ecological niche (Driks & Eichenberger, 2016; Galperin et al., 2012; McKenney et al., 2013; McPherson et al., 2005; Paredes-Sabja, Setlow, & Sarker, 2011; J. Zhang et al., 1993; Zilhão et al., 1999).

According to our natural selection results, the spore coat genes *cotH, cotK-sspO, cwlJ, gerQ, yheC, yheD, ysxE,*and *yutH-cotNH* are positively selected in the Cereus group. Only *cwlJ* and *gerQ* are positively selected along the entire gene sequence, whereas the remaining are positively selected in certain sites. Similarly, *spoVID* and *gerQ* are positively selected along the entire gene sequence in the Methanolicus and Subtilis group, respectively. On the other hand, *ysxE* has positively selected sites in the Coagulans, Megaterium, Methanolicus, Pumilus, and Subtilis groups. *yutH* has positively selected sites in the Methanolicus group and negatively selected sites in the Megaterium group.

Gene sequences encoding for morphogenetic coat proteins CotE, CotH, SafA, and SpoVID are positively selected in the Cereus, Pumilus, and Subtilis groups. These spore coat proteins are under positive selection either in specific sites or along the entire gene sequence, according to our analyses. Similarly, *gerQ, gerPC, gerPE, gerT*, and *yhcN* coat genes are positively selected. Considering the important roles in coat assembly, structure, and germination of these spore coat proteins, it is not surprising they provide significant survival advantages to the individuals carrying them (Butzin et al., 2012; Driks & Eichenberger, 2016; Ferguson, Camp, & Losick, 2007; Galperin, 2013; Galperin et al., 2012; Johnson & Moir, 2017; McKenney & Eichen-

berger, 2012; McPherson et al., 2005; Monroe & Setlow, 2006; Ozin et al., 2000; Ragkousi & Setlow, 2004; Zilhão et al., 1999).

Nevertheless, in the Coagulans, Megaterium, and Methanolicus groups the morphogenetic coat genes *cotE, spoVM*, and *spoIVA* seem to be under purifying selection along with genes that encode for the proteins GerPE, CotJA and CotJB that are dependent on the latter. It has been previously hypothesized that adaptation of spore-forming species to nutrient-rich niches may lead to the loss of their ability to form spores (Galperin, 2013). Consequently, we suggest a possible explanation for this unexpected result. Bacterial specimens of the aforementioned groups have been extensively used for industrial processes (Carnicer et al., 2016; Jebeli et al., 2017; Mazzoli et al., 2019; Saw et al., 2019; Wagle et al., 2019), which led them to a shift into a new nutrient-rich niche. Thus, spore formation is not advantageous in the new niche and genes associated with this process are favored by a purifying negative selection. However, more studies must confirm this explanation since the genomes used in this study do not represent the entire diversity of a group.

YheC and YheD are positively selected spore coat proteins that have an ATP binding domain and are controlled by the same operon (Driks & Eichenberger, 2016). YheD is located in the basement layer of the spore coat and is dependent on SpoIVA, whereas the localization of YheC is not determined (Driks & Eichenberger, 2016; Ooij, Eichenberger, & Losick, 2004). During the initial stages of sporulation, YheD forms two rings that encircle the forespore (Ooij et al., 2004). In the later stages of sporulation, the two rings disappear, and YheD is redistributed around the basement layer of the forespore (Driks & Eichenberger, 2016; Ooij et al., 2004). Hence, these spore coat proteins are important for the initial stages of sporulation in the Cereus and Subtilis groups.

YutH and YsxE are bacterial spore kinase proteins located in the inner layer and are SpoIVA and SafA dependent (Driks & Eichenberger, 2016; Ooij et al., 2004; Scheeff et al., 2010). YutH and YsxE provide protection against lysozyme, hypochlorite, and predation to the spore (Scheeff et al., 2010). Thus, these bacterial spore kinases are evolutionary important for the survival of the spore in different environments (Scheeff et al., 2010). Our selection pressure analyses revealed that these spore coat proteins are positively selected in specific sites. These

sites may be highly conserved motifs associated with likely enzymatic activity (Scheeff et al., 2010). In addition, positively selected sites in the coat genes analyzed may exert an important function in the final protein product as interaction/binding partners. However, more studies are needed to confirm this hypothesis.

According to our HGT results, the majority of spore coat proteins analyzed here are not subjected to HGT events. This supports the idea that the ability to form spores in *Firmicutes* (*Bacilli* and *Clostridia*) is an ancestral feature (Galperin, 2013; Onyenwoke et al., 2004; Ramos-Silva, Serrano, & Henriques, 2019). Nevertheless, it is important to highlight that some coat genes in different Bacillus groups have been subjected to HGT events at least once during its evolutionary history. This is further confirmed by the presence of IS sequences in genomes of the recipient species. Moreover, these HGT events occur more frequently at the tips of the reconciled gene-species phylogenetic trees, demonstrating a recent event. This is notably true for the morphogenetic coat proteins CotE and SpoIVA in the Cereus group. These are the only morphogenetic coat proteins under HGT. The remaining spore coat genes under HGT in different Bacillus groups have an overlapping function with other coat proteins already present in the recipient genome. This contributes to the idea of sporulation as an ancestral character in *Firmicutes*.

In addition, the spore coat genes transferred by HGT are positively selected and confer an advantage in the sporulation and germination processes for the recipient species (Butzin et al., 2012; Driks & Eichenberger, 2016; Ferguson et al., 2007; Galperin, 2013; Galperin et al., 2012; Johnson & Moir, 2017; McKenney & Eichenberger, 2012; McPherson et al., 2005; Monroe & Setlow, 2006; Ozin et al., 2000; Ragkousi & Setlow, 2004; Zilhão et al., 1999), which may lead to a fixation of these genes in the genome (Soucy et al., 2015). Bacterial species that contain spore coat genes under HGT events may reflect a complex evolutionary history adapted to lineage-specific environmental conditions (Galperin et al., 2012; Ramos-Silva et al., 2019). This idea must be further explored by future studies on the evolutionary dynamics of these species.

CHAPTER 5. CONCLUSIONS

The presence of spores in cultures confirms the bacterial organism as a spore-former species. However, the absence of visible spores in a culture is not sufficient to preclude the possibility of spore formation under different culture conditions (Galperin, 2013). Our presence/absence results may provide a novel method to classify a bacterial species as a spore-former or asporogeneous. Moreover, our results suggest that there is a well-conserved genomic signature of spore coat proteins, composed mainly of morphogenetic coat proteins. This genomic signature reflects overlapping functions and a spore coat protein interaction network that is highly adapted to unique environmental conditions Driks and Eichenberger (2016); Galperin et al. (2012); Ramos-Silva et al. (2019); J. Zhang et al. (1993).

Spore coat genes that are positively selected provide important roles in coat assembly, structure, and germination (Butzin et al., 2012; Driks & Eichenberger, 2016; Ferguson et al., 2007; Galperin, 2013; Galperin et al., 2012; Johnson & Moir, 2017; McKenney & Eichenberger, 2012; McPherson et al., 2005; Monroe & Setlow, 2006; Ozin et al., 2000; Ragkousi & Setlow, 2004; Zilhão et al., 1999). Thus, providing significant survival advantages to the individuals carrying them. In addition, sites of coat genes that are positively selected may be highly conserved motifs associated with likely enzymatic activity (Scheeff et al., 2010) or may exert an important function in the final protein product as interaction/binding partners. Coat genes, such as *cotE, spoVM, spoIVA gerPE, cotJA,* and *cotJB* seem to be under purifying selection in the Coagulans, Megaterium, and Methanolicus groups. This could reflect a shift into a new nutrient-rich niche, which may no longer favor genes associated with sporulation. It is important to mention that the genomes analyzed in this work do not reflect the diversity of the species or group.

The majority of spore coat proteins analyzed here are not subjected to HGT events. Thus, supporting the idea of sporulation as an ancestral feature in *Firmicutes* (Galperin, 2013; Onyenwoke et al., 2004; Ramos-Silva et al., 2019). Exceptions are the morphogenetic coat

proteins CotE and SpoIVA in the Cereus group. Since these morphogenetic coat proteins are positively selected and confer an advantage in the sporulation and germination processes (Butzin et al., 2012; Driks & Eichenberger, 2016; Ferguson et al., 2007; Galperin, 2013; Galperin et al., 2012; Johnson & Moir, 2017; McKenney & Eichenberger, 2012; McPherson et al., 2005; Monroe & Setlow, 2006; Ozin et al., 2000; Ragkousi & Setlow, 2004; Zilhão et al., 1999), they are likely to be fixated in the recipient species (Soucy et al., 2015). These HGT events may reflect a complex evolutionary history.

CHAPTER 6. RECOMMENDATIONS AND FUTURE WORK

Protein coevolution analyses may help to explain why there are positively selected sites on some spore coat genes. Moreover, increasing the number of genomes for the Coagulans, Megaterium, and Methanolicus group may provide a better insight on the selection pressures acting upon the coat genes *cotE, spoVM, spoIVA, gerPE, cotJA*, *cotJB*.

LIST OF REFERENCES

LIST OF REFERENCES

Abascal, F., Zardoya, R., & Telford, M. J. (2010, July). TranslatorX: multiple alignment of nucleotide sequences guided by amino acid translations. *Nucleic Acids Research*, *38*(suppl_2), W7–W13. Retrieved 2019-07-25, from https://academic.oup.com/nar/article/38/suppl_2/W7/1094709 doi: 10.1093/nar/gkq291

Abe, H., Koyama, K., Kawamura, S., & Koseki, S. (2019, September). Stochastic modeling of variability in survival behavior of Bacillus simplex spore population during isothermal inactivation at the single cell level using a Monte Carlo simulation. *Food Microbiology*, *82*, 436–444. doi: 10.1016/j.fm.2019.03.005

Abecasis, A. B., Serrano, M., Alves, R., Quintais, L., Pereira-Leal, J. B., & Henriques, A. O. (2013, May). A Genomic Signature and the Identification of New Sporulation Genes. *Journal of Bacteriology*, *195*(9), 2101–2115. Retrieved 2019-10-08, from https://jb.asm.org/content/195/9/2101 doi: 10.1128/JB.02110-12

Abhyankar, W., Pandey, R., Ter Beek, A., Brul, S., de Koning, L. J., & de Koster, C. G. (2015, February). Reinforcement of Bacillus subtilis spores by cross-linking of outer coat proteins during maturation. *Food Microbiology*, *45*(Pt A), 54–62. doi: 10.1016/j.fm.2014.03.007

Altschul, S. F., Gish, W., Miller, W., Myers, E. W., & Lipman, D. J. (1990, October). Basic local alignment search tool. *Journal of Molecular Biology*, *215*(3), 403–410. Retrieved 2019-08-05, from http://www.sciencedirect.com/science/article/pii/S0022283605803602 doi: 10.1016/S0022-2836(05)80360-2

Altschul, S. F., Madden, T. L., Schäffer, A. A., Zhang, J., Zhang, Z., Miller, W., & Lipman, D. J. (1997, September). Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Research*, *25*(17), 3389–3402. Retrieved 2019-08-05, from https://academic.oup.com/nar/article/25/17/3389/1061651 doi: 10.1093/nar/25.17.3389

Andersson, J. O. (2009). Horizontal gene transfer between microbial eukaryotes. *Methods in Molecular Biology (Clifton, N.J.)*, *532*, 473–487. doi: 10.1007/978-1-60327-853-9_27

Anisimova, M., & Kosiol, C. (2009, February). Investigating Protein-Coding Sequence Evolution with Probabilistic Codon Substitution Models. *Molecular Biology and Evolution*, *26*(2), 255–271. Retrieved 2019-08-13, from https://academic.oup.com/mbe/article/26/2/255/1029542 doi: 10.1093/molbev/msn232

Arfman, N., Dijkhuizen, L., Kirchhof, G., Ludwig, W., Schleifer, K. H., Bulygina, E. S., ... White, D. (1992, July). Bacillus methanolicus sp. nov., a new species of thermotolerant, methanol-utilizing, endospore-forming bacteria. *International Journal of Systematic Bacteriology*, *42*(3), 439–445. doi: 10.1099/00207713-42-3-439

Aronson, A. I., & Shai, Y. (2001, February). Why Bacillus thuringiensis insecticidal toxins are so effective: unique features of their mode of action. *FEMS Microbiology Letters*, *195*(1), 1–8. Retrieved 2019-10-03, from https://academic.oup.com/femsle/article/195/1/1/521054 doi: 10.1111/j.1574-6968.2001.tb10489.x

Barcellos, F. G., Menna, P., Batista, J. S. d. S., & Hungria, M. (2007, April). Evidence of Horizontal Transfer of Symbiotic Genes from a Bradyrhizobium japonicum Inoculant Strain to Indigenous Diazotrophs Sinorhizobium (Ensifer) fredii and Bradyrhizobium elkanii in a Brazilian Savannah Soil. *Applied and Environmental Microbiology*, *73*(8), 2635–2643. Retrieved 2019-08-05, from https://aem.asm.org/content/73/8/2635 doi: 10.1128/AEM.01823-06

Barák, I., Ricca, E., & Cutting, S. M. (2005, January). From fundamental studies of sporulation to applied spore research. *Molecular Microbiology*, *55*(2), 330–338. doi: 10.1111/j.1365-2958.2004.04445.x

Bauer, T., Little, S., Stöver, A. G., & Driks, A. (1999, November). Functional Regions of the Bacillus subtilis Spore Coat Morphogenetic Protein CotE. *Journal of Bacteriology*, *181*(22), 7043–7051. Retrieved 2019-10-05, from https://jb.asm.org/content/181/22/7043

Bağcıoğlu, M., Fricker, M., Johler, S., & Ehling-Schulz, M. (2019). Detection and Identification of Bacillus cereus, Bacillus cytotoxicus, Bacillus thuringiensis, Bacillus mycoides and Bacillus weihenstephanensis via Machine Learning Based FTIR Spectroscopy. *Frontiers in Microbiology*, *10*, 902. doi: 10.3389/fmicb.2019.00902

Beladjal, L., Gheysens, T., Clegg, J. S., Amar, M., & Mertens, J. (2018, September). Life from the ashes: survival of dry bacterial spores after very high temperature exposure. *Extremophiles: Life Under Extreme Conditions*, *22*(5), 751–759. doi: 10.1007/s00792-018-1035-6

Bellanger, X., Payot, S., Leblond-Bourget, N., & Guédon, G. (2014, July). Conjugative and mobilizable genomic islands in bacteria: evolution and diversity. *FEMS Microbiology Reviews*, *38*(4), 720–760. Retrieved 2019-08-15, from https://academic.oup.com/femsre/article/38/4/720/758158 doi: 10.1111/1574-6976.12058

Benson, D. A., Karsch-Mizrachi, I., Lipman, D. J., Ostell, J., & Sayers, E. W. (2009, January). GenBank. *Nucleic Acids Research*, *37*(suppl_1), D26–D31. Retrieved 2019-08-07, from https://academic.oup.com/nar/article/37/suppl_1/D26/1007536 doi: 10.1093/nar/gkn723

Bhattacharyya, P. N., & Jha, D. K. (2012, April). Plant growth-promoting rhizobacteria (PGPR): emergence in agriculture. *World Journal of Microbiology and Biotechnology*, *28*(4), 1327–1350. Retrieved 2019-10-08, from https://doi.org/10.1007/s11274-011-0979-9 doi: 10.1007/s11274-011-0979-9

Bosma, E. F., Koehorst, J. J., van Hijum, S. A. F. T., Renckens, B., Vriesendorp, B., van de Weijer, A. H. P., ... van Kranenburg, R. (2016, August). Complete genome sequence of thermophilic Bacillus smithii type strain DSM 4216t. *Standards in Genomic Sciences*, *11*(1), 52. Retrieved 2019-10-04, from https://doi.org/10.1186/s40793-016-0172-8 doi: 10.1186/s40793-016-0172-8

Bosma, E. F., van de Weijer, A. H. P., Daas, M. J. A., van der Oost, J., de Vos, W. M., & van Kranenburg, R. (2015, March). Isolation and screening of thermophilic bacilli

from compost for electrotransformation and fermentation: characterization of Bacillus smithii ET 138 as a new biocatalyst. *Applied and Environmental Microbiology*, *81*(5), 1874–1883. doi: 10.1128/AEM.03640-14

Bozue, J. A., Welkos, S., & Cote, C. K. (2015, October). The Bacillus anthracis Exosporium: What's the Big "Hairy" Deal? *Microbiology Spectrum*, *3*(5). Retrieved 2019-10-26, from https://www.asmscience.org/content/journal/microbiolspec/10.1128/microbiolspec.TBS-0021-2015 doi: 10.1128/microbiolspec.TBS-0021-2015

Burrus, V., Pavlovic, G., Decaris, B., & Guédon, G. (2002). Conjugative transposons: the tip of the iceberg. *Molecular Microbiology*, *46*(3), 601–610. Retrieved 2019-08-15, from https://onlinelibrary.wiley.com/doi/abs/10.1046/j.1365-2958.2002.03191.x doi: 10.1046/j.1365-2958.2002.03191.x

Butzin, X. Y., Troiano, A. J., Coleman, W. H., Griffiths, K. K., Doona, C. J., Feeherry, F. E., ... Setlow, P. (2012, November). Analysis of the Effects of a gerP Mutation on the Germination of Spores of Bacillus subtilis. *Journal of Bacteriology*, *194*(21), 5749–5758. Retrieved 2019-10-05, from https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3486119/ doi: 10.1128/JB.01276-12

Camacho, C., Coulouris, G., Avagyan, V., Ma, N., Papadopoulos, J., Bealer, K., & Madden, T. L. (2009, December). BLAST+: architecture and applications. *BMC Bioinformatics*, *10*(1), 421. Retrieved 2019-08-05, from https://doi.org/10.1186/1471-2105-10-421 doi: 10.1186/1471-2105-10-421

Campo, D., Lehmann, K., Fjeldsted, C., Souaiaia, T., Kao, J., & Nuzhdin, S. V. (2013). Whole-genome sequencing of two North American Drosophila melanogaster populations reveals genetic differentiation and positive selection. *Molecular Ecology*, *22*(20), 5084–5097. Retrieved 2019-08-08, from https://www.onlinelibrary.wiley.com/doi/abs/10.1111/mec.12468 doi: 10.1111/mec.12468

Cangiano, G., Sirec, T., Panarella, C., Isticato, R., Baccigalupi, L., Felice, M. D., & Ricca, E. (2014, December). The sps Gene Products Affect the Germination, Hydrophobicity, and Protein Adsorption of Bacillus subtilis Spores. *Applied and Environmental Microbiology*, *80*(23), 7293–7302. Retrieved 2019-08-06, from https://aem.asm.org/content/80/23/7293 doi: 10.1128/AEM.02893-14

Carnicer, M., Vieira, G., Brautaset, T., Portais, J.-C., & Heux, S. (2016, June). Quantitative metabolomics of the thermophilic methylotroph Bacillus methanolicus. *Microbial Cell Factories*, *15*(1), 92. Retrieved 2019-10-04, from https://doi.org/10.1186/s12934-016-0483-x doi: 10.1186/s12934-016-0483-x

Castillo, J. A., & Agathos, S. N. (2019, June). A genome-wide scan for genes under balancing selection in the plant pathogen Ralstonia solanacearum. *BMC Evolutionary Biology*, *19*(1), 123. Retrieved 2019-08-22, from https://doi.org/10.1186/s12862-019-1456-6 doi: 10.1186/s12862-019-1456-6

Cutting, S. M., Hong, H. A., Baccigalupi, L., & Ricca, E. (2009). Oral vaccine delivery by recombinant spore probiotics. *International Reviews of Immunology*, *28*(6), 487–505. doi: 10.3109/08830180903215605

Darby, C. A., Stolzer, M., Ropp, P. J., Barker, D., & Durand, D. (2017, March). Xenolog classification. *Bioinformatics*, *33*(5), 640–649. Retrieved 2019-08-05, from https://academic.oup.com/bioinformatics/article/33/5/640/2725487 doi: 10.1093/bioinformatics/btw686

Delport, W., Scheffler, K., & Seoighe, C. (2009, January). Models of coding sequence evolution. *Briefings in Bioinformatics*, *10*(1), 97–109. Retrieved 2019-08-12, from https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2638624/ doi: 10.1093/bib/bbn049

Dierick, K., Coillie, E. V., Swiecicka, I., Meyfroidt, G., Devlieger, H., Meulemans, A., ... Mahillon, J. (2005, August). Fatal Family Outbreak of Bacillus cereus-Associated Food Poisoning. *Journal of Clinical Microbiology*, *43*(8), 4277–4279. Retrieved 2019-10-03, from https://jcm.asm.org/content/43/8/4277 doi: 10.1128/JCM.43.8.4277-4279.2005

Dobrindt, U., Janke, B., Piechaczek, K., Nagy, G., Ziebuhr, W., Fischer, G., ... Hacker, J. (2000, October). Toxin genes on pathogenicity islands: impact for microbial evolution. *International Journal of Medical Microbiology*, *290*(4), 307–311. Retrieved 2019-08-05, from http://www.sciencedirect.com/science/article/pii/S1438422100800284 doi: 10.1016/S1438-4221(00)80028-4

Driks, A., & Eichenberger, P. (2016, March). The Spore Coat. *Microbiology Spectrum*, *4*(2). Retrieved 2019-07-14, from http://www.asmscience.org/content/journal/microbiolspec/10.1128/microbiolspec.TBS-0023-2016 doi: 10.1128/microbiolspec.TBS-0023-2016

Durand, D., Halldórsson, B. V., & Vernot, B. (2006, March). A Hybrid Micro–Macroevolutionary Approach to Gene Tree Reconstruction. *Journal of Computational Biology*, *13*(2), 320–335. Retrieved 2019-08-05, from https://www.liebertpub.com/doi/abs/10.1089/cmb.2006.13.320 doi: 10.1089/cmb.2006.13.320

Earl, A. M., Losick, R., & Kolter, R. (2008, June). Ecology and genomics of Bacillus subtilis. *Trends in Microbiology*, *16*(6), 269–275. Retrieved 2019-09-18, from http://www.sciencedirect.com/science/article/pii/S0966842X08000887 doi: 10.1016/j.tim.2008.03.004

Ferguson, C. C., Camp, A. H., & Losick, R. (2007, November). gerT, a Newly Discovered Germination Gene under the Control of the Sporulation Transcription Factor K in Bacillus subtilis. *Journal of Bacteriology*, *189*(21), 7681. Retrieved 2019-10-06, from https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2168713/ doi: 10.1128/JB.01053-07

Galperin, M. Y. (2013, December). Genome Diversity of Spore-Forming Firmicutes. *Microbiology spectrum*, *1*(2), TBS–0015–2012. Retrieved 2019-09-21, from https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4306282/ doi: 10.1128/microbiolspectrum.TBS-0015-2012

Galperin, M. Y., Mekhedov, S. L., Puigbo, P., Smirnov, S., Wolf, Y. I., & Rigden, D. J. (2012, November). Genomic determinants of sporulation in Bacilli and Clostridia: towards the minimal set of sporulation-specific genes. *Environmental Microbiology*, *14*(11), 2870–2890. Retrieved 2019-07-14, from https://onlinelibrary.wiley.com/doi/full/10.1111/j.1462-2920.2012.02841.x doi: 10.1111/j.1462-2920.2012.02841.x

Glick, B. R. (2012). *Plant Growth-Promoting Bacteria: Mechanisms and Applications* [Research article]. Retrieved 2019-10-08, from https://www.hindawi.com/journals/scientifica/2012/963401/ doi: 10.6064/2012/963401

Gogarten, J. P., & Townsend, J. P. (2005, September). Horizontal gene transfer, genome innovation and evolution. *Nature Reviews. Microbiology*, *3*(9), 679–687. doi: 10.1038/nrmicro1204

Guindon, S., Dufayard, J.-F., Lefort, V., Anisimova, M., Hordijk, W., & Gascuel, O. (2010, May). New Algorithms and Methods to Estimate Maximum-Likelihood Phylogenies: Assessing the Performance of PhyML 3.0. *Systematic Biology*, *59*(3), 307–321. Retrieved 2019-08-01, from https://academic.oup.com/sysbio/article/59/3/307/1702850 doi: 10.1093/sysbio/syq010

Hacker, J., & Carniel, E. (2001, May). Ecological fitness, genomic islands and bacterial pathogenicity. *EMBO Reports*, *2*(5), 376–381. Retrieved 2019-08-16, from https://www.ncbi.nlm.nih.gov/pmc/articles/PMC1083891/ doi: 10.1093/embo-reports/kve097

Hacker, J., & Kaper, J. B. (2000). Pathogenicity islands and the evolution of microbes. *Annual Review of Microbiology*, *54*, 641–679. doi: 10.1146/annurev.micro.54.1.641

Henriques, A. O., Beall, B. W., Roland, K., & Moran, C. P. (1995, June). Characterization of cotJ, a sigma E-controlled operon affecting the polypeptide composition of the coat of Bacillus subtilis spores. *Journal of Bacteriology*, *177*(12), 3394–3406. Retrieved 2019-10-05, from https://www.ncbi.nlm.nih.gov/pmc/articles/PMC177041/

Henriques, A. O., & Moran, C. P. (2007). Structure, assembly, and function of the spore surface layers. *Annual Review of Microbiology*, *61*, 555–588. doi: 10.1146/annurev.micro.61.080706.093224

Hsiao, W. W. L., Ung, K., Aeschliman, D., Bryan, J., Finlay, B. B., & Brinkman, F. S. L. (2005, November). Evidence of a Large Novel Gene Pool Associated with Prokaryotic Genomic Islands. *PLOS Genetics*, *1*(5), e62. Retrieved 2019-08-16, from https://journals.plos.org/plosgenetics/article?id=10.1371/journal.pgen.0010062 doi: 10.1371/journal.pgen.0010062

Imamura, D., Kuwana, R., Takamatsu, H., & Watabe, K. (2011, August). Proteins Involved in Formation of the Outermost Layer of Bacillus subtilis Spores. *Journal of Bacteriology*, *193*(16), 4075–4080. Retrieved 2020-01-06, from https://jb.asm.org/content/193/16/4075 doi: 10.1128/JB.05310-11

Jebeli, M. A., Maleki, A., Amoozegar, M. A., Kalantar, E., Izanloo, H., & Gharibi, F. (2017, February). Bacillus flexus strain As-12, a new arsenic transformer bacterium isolated from contaminated water resources. *Chemosphere*, *169*, 636–641. doi: 10.1016/j.chemosphere.2016.11.129

Johnson, C. L., & Moir, A. (2017, April). Proteins YlaJ and YhcN contribute to the efficiency of spore germination in Bacillus subtilis. *FEMS Microbiology Letters*, *364*(7). Retrieved 2019-10-06, from https://academic.oup.com/femsle/article/364/7/fnx047/3045907 doi: 10.1093/femsle/fnx047

Juhas, M., van der Meer, J. R., Gaillard, M., Harding, R. M., Hood, D. W., & Crook, D. W. (2009, March). Genomic islands: tools of bacterial horizontal gene transfer and evolution. *FEMS Microbiology Reviews*, *33*(2), 376–393. Retrieved 2019-08-16, from https://academic.oup.com/femsre/article/33/2/376/589749 doi: 10.1111/j.1574-6976.2008.00136.x

Kanehisa, M., Sato, Y., Kawashima, M., Furumichi, M., & Tanabe, M. (2016, January). KEGG as a reference resource for gene and protein annotation. *Nucleic Acids Research*, *44*(D1), D457–462. doi: 10.1093/nar/gkv1070

Kanehisa, M., Sato, Y., & Morishima, K. (2016, February). BlastKOALA and GhostKOALA: KEGG Tools for Functional Characterization of Genome and Metagenome Sequences. *Journal of Molecular Biology*, *428*(4), 726–731. doi: 10.1016/j.jmb.2015.11.006

Klobutcher, L. A., Ragkousi, K., & Setlow, P. (2006, January). The Bacillus subtilis spore coat provides "eat resistance" during phagocytic predation by the protozoan Tetrahymena thermophila. *Proceedings of the National Academy of Sciences of the United States of America*, *103*(1), 165–170. doi: 10.1073/pnas.0507121102

Kodama, T., Takamatsu, H., Asai, K., Kobayashi, K., Ogasawara, N., & Watabe, K. (1999, August). The Bacillus subtilis yaaH Gene Is Transcribed by SigE RNA Polymerase during Sporulation, and Its Product Is Involved in Germination of Spores. *Journal of Bacteriology*, *181*(15), 4584–4591. Retrieved 2019-07-14, from `https://jb.asm.org/content/181/15/4584`

Konuray, G., & Erginkaya, Z. (2018, June). Potential Use of Bacillus coagulans in the Food Industry. *Foods*, *7*(6). Retrieved 2019-10-04, from `https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6025323/` doi: 10.3390/foods7060092

Kosakovsky Pond, S. L., Murrell, B., Fourment, M., Frost, S. D. W., Delport, W., & Scheffler, K. (2011, November). A Random Effects Branch-Site Model for Detecting Episodic Diversifying Selection. *Molecular Biology and Evolution*, *28*(11), 3033–3043. Retrieved 2019-08-12, from `https://academic.oup.com/mbe/article/28/11/3033/1044451` doi: 10.1093/molbev/msr125

Kotiranta, A., Lounatmaa, K., & Haapasalo, M. (2000, February). Epidemiology and pathogenesis of Bacillus cereus infections. *Microbes and Infection*, *2*(2), 189–198.

Krajčíková, D., Forgáč, V., Szabo, A., & Barák, I. (2017, November). Exploring the interaction network of the Bacillus subtilis outer coat and crust proteins. *Microbiological Research*, *204*, 72–80. doi: 10.1016/j.micres.2017.08.004

Krajčíková, D., Lukáčová, M., Müllerová, D., Cutting, S. M., & Barák, I. (2009, May). Searching for Protein-Protein Interactions within the Bacillus subtilis Spore Coat. *Journal of Bacteriology*, *191*(10), 3212–3219. Retrieved 2019-10-05, from `https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2687167/` doi: 10.1128/JB.01807-08

Krizova, L., & Nemec, A. (2010, September). A 63 kb genomic resistance island found in a multidrug-resistant Acinetobacter baumannii isolate of European clone I from 1977. *Journal of Antimicrobial Chemotherapy*, *65*(9), 1915–1918. Retrieved 2019-08-05, from `https://academic.oup.com/jac/article/65/9/1915/719721` doi: 10.1093/jac/dkq223

Lee, Y.-J., Lee, S.-J., Jeong, H., Kim, H. J., Ryu, N., Kim, B.-C., ... Lee, S. J. (2012, November). Draft genome sequence of Bacillus oceanisediminis 2691. *Journal of Bacteriology*, *194*(22), 6351–6352. doi: 10.1128/JB.01643-12

Lefort, V., Longueville, J.-E., & Gascuel, O. (2017, September). SMS: Smart Model Selection in PhyML. *Molecular Biology and Evolution*, *34*(9), 2422–2424. Retrieved 2019-08-05, from `https://academic.oup.com/mbe/article/34/9/2422/3788860` doi: 10.1093/molbev/msx149

Liu, M., Li, X., Xie, Y., Bi, D., Sun, J., Li, J., ... Ou, H.-Y. (2019, January). ICEberg 2.0: an updated database of bacterial integrative and conjugative elements. *Nucleic Acids Research*, *47*(D1), D660–D665. Retrieved 2019-08-05, from `https://academic.oup.com/nar/article/47/D1/D660/5165266` doi: 10.1093/nar/gky1123

Maayer, P. D., Aliyu, H., & Cowan, D. A. (2019, March). Reorganising the order Bacillales through phylogenomics. *Systematic and Applied Microbiology*, *42*(2), 178–189. Retrieved 2019-10-08, from http://www.sciencedirect.com/science/article/pii/S0723202018303291 doi: 10.1016/j.syapm.2018.10.007

Mazzoli, A., Donadio, G., Lanzilli, M., Saggese, A., Guarino, A. M., Rivetti, M., ... Isticato, R. (2019, August). Bacillus megaterium SF185 spores exert protective effects against oxidative stress in vivo and in vitro. *Scientific Reports*, *9*. Retrieved 2019-10-04, from https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6700169/ doi: 10.1038/s41598-019-48531-4

McKenney, P. T., Driks, A., & Eichenberger, P. (2013, January). The Bacillus subtilis endospore: assembly and functions of the multilayered coat. *Nature Reviews. Microbiology*, *11*(1), 33–44. doi: 10.1038/nrmicro2921

McKenney, P. T., Driks, A., Eskandarian, H. A., Grabowski, P., Guberman, J., Wang, K. H., ... Eichenberger, P. (2010, May). A Distance-Weighted Interaction Map Reveals a Previously Uncharacterized Layer of the B. subtilis Spore Coat. *Current biology : CB*, *20*(10), 934–938. Retrieved 2019-10-05, from https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2920530/ doi: 10.1016/j.cub.2010.03.060

McKenney, P. T., & Eichenberger, P. (2012, January). Dynamics of Spore Coat Morphogenesis in Bacillus subtilis. *Molecular Microbiology*, *83*(2), 245–260. Retrieved 2019-10-05, from https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3256263/ doi: 10.1111/j.1365-2958.2011.07936.x

McPherson, D. C., Kim, H., Hahn, M., Wang, R., Grabowski, P., Eichenberger, P., & Driks, A. (2005, December). Characterization of the Bacillus subtilis Spore Morphogenetic Coat Protein CotO. *Journal of Bacteriology*, *187*(24), 8278–8290. Retrieved 2019-10-05, from https://jb.asm.org/content/187/24/8278 doi: 10.1128/JB.187.24.8278-8290.2005

Mock, M., & Fouet, A. (2001). Anthrax. *Annual Review of Microbiology*, *55*, 647–671. doi: 10.1146/annurev.micro.55.1.647

Moir, A., & Cooper, G. (2015, November). Spore Germination. *Microbiology Spectrum*, *3*(6). Retrieved 2019-10-08, from https://www.asmscience.org/content/journal/microbiolspec/10.1128/microbiolspec.TBS-0014-2012 doi: 10.1128/microbiolspec.TBS-0014-2012

Molle, V., Fujita, M., Jensen, S. T., Eichenberger, P., González-Pastor, J. E., Liu, J. S., & Losick, R. (2003, December). The Spo0a regulon of Bacillus subtilis. *Molecular Microbiology*, *50*(5), 1683–1701. doi: 10.1046/j.1365-2958.2003.03818.x

Monroe, A., & Setlow, P. (2006, November). Localization of the Transglutaminase Cross-Linking Sites in the Bacillus subtilis Spore Coat Protein GerQ. *Journal of Bacteriology*, *188*(21), 7609–7616. Retrieved 2019-10-05, from https://www.ncbi.nlm.nih.gov/pmc/articles/PMC1636287/ doi: 10.1128/JB.01116-06

Murrell, B., Weaver, S., Smith, M. D., Wertheim, J. O., Murrell, S., Aylward, A., ... Kosakovsky Pond, S. L. (2015, May). Gene-Wide Identification of Episodic Selection. *Molecular Biology and Evolution*, *32*(5), 1365–1371. Retrieved 2019-08-01, from https://academic.oup.com/mbe/article/32/5/1365/1134918 doi: 10.1093/molbev/msv035

Murrell, B., Wertheim, J. O., Moola, S., Weighill, T., Scheffler, K., & Pond, S. L. K. (2012, July). Detecting Individual Sites Subject to Episodic Diversifying Selection. *PLOS Genetics*, *8*(7), e1002764. Retrieved 2019-08-01, from https://journals.plos.org/plosgenetics/article?id=10.1371/journal.pgen.1002764 doi: 10.1371/journal.pgen.1002764

Nakamura, L. K. (1998, July). Bacillus pseudomycoides sp. nov. *International Journal of Systematic Bacteriology*, *48 Pt 3*, 1031–1035. doi: 10.1099/00207713-48-3-1031

Nakamura, L. K., Roberts, M. S., & Cohan, F. M. (1999). Note: Relationship of Bacillus subtilis clades associated with strains 168 and W23: A proposal for Bacillus subtilis subsp. subtilis subsp. nov. and Bacillus subtilis subsp. spizizenii subsp. nov. *International Journal of Systematic and Evolutionary Microbiology,*, *49*(3), 1211–1215. Retrieved 2019-09-18, from https://www.microbiologyresearch.org/content/journal/ijsem/10.1099/00207713-49-3-1211 doi: 10.1099/00207713-49-3-1211

Nicholson, W. L., Munakata, N., Horneck, G., Melosh, H. J., & Setlow, P. (2000, September). Resistance of Bacillus endospores to extreme terrestrial and extraterrestrial environments. *Microbiology and molecular biology reviews: MMBR*, *64*(3), 548–572. doi: 10.1128/mmbr.64.3.548-572.2000

Nielsen, R., & Yang, Z. (1998, March). Likelihood Models for Detecting Positively Selected Amino Acid Sites and Applications to the HIV-1 Envelope Gene. *Genetics*, *148*(3), 929–936. Retrieved 2019-08-13, from https://www.genetics.org/content/148/3/929

Onyenwoke, R. U., Brill, J. A., Farahi, K., & Wiegel, J. (2004, October). Sporulation genes in members of the low G+C Gram-type-positive phylogenetic branch (Firmicutes). *Archives of Microbiology*, *182*(2-3), 182–192. doi: 10.1007/s00203-004-0696-y

Ooij, C. v., Eichenberger, P., & Losick, R. (2004, July). Dynamic Patterns of Subcellular Protein Localization during Spore Coat Morphogenesis in Bacillus subtilis. *Journal of Bacteriology*, *186*(14), 4441. Retrieved 2019-10-06, from https://www.ncbi.nlm.nih.gov/pmc/articles/PMC438564/ doi: 10.1128/JB.186.14.4441-4448.2004

Osborn, A. M., & Böltner, D. (2002, November). When phage, plasmids, and transposons collide: genomic islands, and conjugative- and mobilizable-transposons as a mosaic continuum. *Plasmid*, *48*(3), 202–212.

Ozin, A. J., Henriques, A. O., Yi, H., & Moran, C. P. (2000, April). Morphogenetic Proteins SpoVID and SafA Form a Complex during Assembly of the Bacillus subtilis Spore Coat. *Journal of Bacteriology*, *182*(7), 1828–1833. Retrieved 2019-10-05, from https://www.ncbi.nlm.nih.gov/pmc/articles/PMC101864/

Paredes-Sabja, D., Setlow, P., & Sarker, M. R. (2011, February). Germination of spores of Bacillales and Clostridiales species: mechanisms and proteins involved. *Trends in Microbiology*, *19*(2), 85–94. Retrieved 2019-10-05, from http://www.sciencedirect.com/science/article/pii/S0966842X10001927 doi: 10.1016/j.tim.2010.10.004

Paul, C., Filippidou, S., Jamil, I., Kooli, W., House, G. L., Estoppey, A., ... Junier, P. (2019). Bacterial spores, from ecology to biotechnology. *Advances in Applied Microbiology*, *106*, 79–111. doi: 10.1016/bs.aambs.2018.10.002

Pearson, W. R. (2013). An Introduction to Sequence Similarity ("Homology") Searching. *Current Protocols in Bioinformatics*, *42*(1), 3.1.1–3.1.8. Retrieved 2019-07-29, from https://currentprotocols.onlinelibrary.wiley.com/doi/abs/10.1002/0471250953.bi0301s42 doi: 10.1002/0471250953.bi0301s42

Qin, H., & Driks, A. (2013, November). Contrasting evolutionary patterns of spore coat proteins in two Bacillus species groups are linked to a difference in cellular structure. *BMC evolutionary biology*, *13*, 261. doi: 10.1186/1471-2148-13-261

Ragkousi, K., & Setlow, P. (2004, September). Transglutaminase-Mediated Cross-Linking of GerQ in the Coats of Bacillus subtilis Spores. *Journal of Bacteriology*, *186*(17), 5567–5575. Retrieved 2019-10-05, from `https://www.ncbi.nlm.nih.gov/pmc/articles/PMC516844/` doi: 10.1128/JB.186.17.5567-5575.2004

Rambaut, A., Drummond, A. J., Xie, D., Baele, G., & Suchard, M. A. (2018, September). Posterior Summarization in Bayesian Phylogenetics Using Tracer 1.7. *Systematic Biology*, *67*(5), 901–904. Retrieved 2019-08-05, from `https://academic.oup.com/sysbio/article/67/5/901/4989127` doi: 10.1093/sysbio/syy032

Ramos-Silva, P., Serrano, M., & Henriques, A. O. (2019, March). From root to tips: sporulation evolution and specialization in Bacillus subtilis and the intestinal pathogen Clostridioides difficile. *bioRxiv*, 473793. Retrieved 2019-10-06, from `https://www.biorxiv.org/content/10.1101/473793v1` doi: 10.1101/473793

Reva, O. N., Smirnov, V. V., Pettersson, B., & Priest, F. G. (2002, January). Bacillus endophyticus sp. nov., isolated from the inner tissues of cotton plants (Gossypium sp.). *International Journal of Systematic and Evolutionary Microbiology*, *52*(Pt 1), 101–107. doi: 10.1099/00207713-52-1-101

Roberts, M. S., Nakamura, L. K., & Cohan, F. M. (1994, April). Bacillus mojavensis sp. nov., distinguishable from Bacillus subtilis by sexual isolation, divergence in DNA sequence, and differences in fatty acid composition. *International Journal of Systematic Bacteriology*, *44*(2), 256–264. doi: 10.1099/00207713-44-2-256

Rosenberg, G., Steinberg, N., Oppenheimer-Shaanan, Y., Olender, T., Doron, S., Ben-Ari, J., ... Kolodkin-Gal, I. (2016, January). Not so simple, not so subtle: the interspecies competition between Bacillus simplex and Bacillus subtilis and its impact on the evolution of biofilms. *NPJ Biofilms and Microbiomes*, *2*, 15027. Retrieved 2019-10-04, from `https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5515258/` doi: 10.1038/npjbiofilms.2015.27

Rozas, J., Ferrer-Mata, A., Sánchez-DelBarrio, J. C., Guirao-Rico, S., Librado, P., Ramos-Onsins, S. E., & Sánchez-Gracia, A. (2017, December). DnaSP 6: DNA Sequence Polymorphism Analysis of Large Data Sets. *Molecular Biology and Evolution*, *34*(12), 3299–3302. Retrieved 2019-07-29, from `https://academic.oup.com/mbe/article/34/12/3299/4161815` doi: 10.1093/molbev/msx248

Saw, C.-Y., Chang, T.-J., Chen, P.-Y., Dai, F.-J., Lau, Y.-Q., Chen, T.-Y., & Chau, C.-F. (2019, August). Presence of Bacillus coagulans spores and vegetative cells in rat intestine and feces and their physiological effects. *Bioscience, Biotechnology, and Biochemistry*, 1–7. doi: 10.1080/09168451.2019.1651628

Sayers, E. W., Barrett, T., Benson, D. A., Bryant, S. H., Canese, K., Chetvernin, V., ... Ye, J. (2009, January). Database resources of the National Center for Biotechnology Information. *Nucleic Acids Research*, *37*(suppl_1), D5–D15. Retrieved 2019-08-07, from `https://academic.oup.com/nar/article/37/suppl_1/D5/1008346` doi: 10.1093/nar/gkn741

Scheeff, E. D., Axelrod, H. L., Miller, M. D., Chiu, H.-J., Deacon, A. M., Wilson, I. A., & Manning, G. (2010, May). Genomics, evolution, and crystal structure of a new family of bacterial spore kinases. *Proteins*, *78*(6), 1470. Retrieved 2019-10-06, from https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2860764/ doi: 10.1002/prot.22663

Schmidt, D., & Pool, J. (2002, January). The Effect of Population History on the Distribution of the Tajima's D Statistic.

Setlow, P. (2007, April). I will survive: DNA protection in bacterial spores. *Trends in Microbiology*, *15*(4), 172–180. doi: 10.1016/j.tim.2007.02.004

Setlow, P. (2014a, April). Germination of Spores of Bacillus Species: What We Know and Do Not Know. *Journal of Bacteriology*, *196*(7), 1297–1305. Retrieved 2019-10-08, from https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3993344/ doi: 10.1128/JB.01455-13

Setlow, P. (2014b, September). Spore Resistance Properties. *Microbiology Spectrum*, *2*(5). Retrieved 2019-10-08, from https://www.asmscience.org/content/journal/microbiolspec/10.1128/microbiolspec.TBS-0003-2012 doi: 10.1128/microbiolspec.TBS-0003-2012

Setlow, P., Wang, S., & Li, Y.-Q. (2017). Germination of Spores of the Orders Bacillales and Clostridiales. *Annual Review of Microbiology*, *71*, 459–477. doi: 10.1146/annurev-micro-090816-093558

Seyler, R. W., Henriques, A. O., Ozin, A. J., & Moran, C. P. (1997, September). Assembly and interactions of cotJ-encoded proteins, constituents of the inner layers of the Bacillus subtilis spore coat. *Molecular Microbiology*, *25*(5), 955–966. doi: 10.1111/j.1365-2958.1997.mmi532.x

Siala, A., Hill, I. R., & Gray, T. R. G. (1974). Populations of Spore-forming Bacteria in an Acid Forest Soil, with Special Reference to Bacillus subtilis. *Microbiology,*, *81*(1), 183–190. Retrieved 2019-09-18, from https://www.microbiologyresearch.org/content/journal/micro/10.1099/00221287-81-1-183 doi: 10.1099/00221287-81-1-183

Soares, S. C., Geyik, H., Ramos, R. T. J., de Sá, P. H. C. G., Barbosa, E. G. V., Baumbach, J., ... Azevedo, V. (2016, August). GIPSy: Genomic island prediction software. *Journal of Biotechnology*, *232*, 2–11. Retrieved 2019-08-05, from http://www.sciencedirect.com/science/article/pii/S0168165615301152 doi: 10.1016/j.jbiotec.2015.09.008

Soucy, S. M., Huang, J., & Gogarten, J. P. (2015, August). Horizontal gene transfer: building the web of life. *Nature Reviews. Genetics*, *16*(8), 472–482. doi: 10.1038/nrg3962

Steinegger, M., & Söding, J. (2017, October). MMseqs2 enables sensitive protein sequence searching for the analysis of massive data sets. *Nature Biotechnology*, *35*, 1026–1028. Retrieved 2019-08-22, from https://www.nature.com/articles/nbt.3988 doi: 10.1038/nbt.3988

Steinegger, M., & Söding, J. (2018, June). Clustering huge protein sequence sets in linear time. *Nature Communications*, *9*(1), 1–8. Retrieved 2019-08-22, from https://www.nature.com/articles/s41467-018-04964-5 doi: 10.1038/s41467-018-04964-5

Stenfors Arnesen, L. P., Fagerlund, A., & Granum, P. E. (2008, July). From soil to gut: Bacillus cereus and its food poisoning toxins. *FEMS Microbiology Reviews*, *32*(4), 579–606. Retrieved 2019-10-03, from https://academic.oup.com/femsre/article/32/4/579/1813157 doi: 10.1111/j.1574-6976.2008.00112.x

Stepanov, V. G., Tirumalai, M. R., Montazari, S., Checinska, A., Venkateswaran, K., & Fox, G. E. (2016, June). Bacillus pumilus SAFR-032 Genome Revisited: Sequence Update and Re-Annotation. *PLOS ONE*, *11*(6), e0157331. Retrieved 2019-10-04, from https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0157331 doi: 10.1371/journal.pone.0157331

Stevens, M. J. A., Tasara, T., Klumpp, J., Stephan, R., Ehling-Schulz, M., & Johler, S. (2019, February). Whole-genome-based phylogeny of Bacillus cytotoxicus reveals different clades within the species and provides clues on ecology and evolution. *Scientific Reports*, *9*(1), 1984. doi: 10.1038/s41598-018-36254-x

Stolzer, M., Lai, H., Xu, M., Sathaye, D., Vernot, B., & Durand, D. (2012, September). Inferring duplications, losses, transfers and incomplete lineage sorting with nonbinary species trees. *Bioinformatics*, *28*(18), i409–i415. Retrieved 2019-08-05, from https://academic.oup.com/bioinformatics/article/28/18/i409/246367 doi: 10.1093/bioinformatics/bts386

Stothard, P. (2000, June). The Sequence Manipulation Suite: JavaScript Programs for Analyzing and Formatting Protein and DNA Sequences. *BioTechniques*, *28*(6), 1102–1104. Retrieved 2019-07-20, from https://www.future-science.com/doi/10.2144/00286ir01 doi: 10.2144/00286ir01

Suchard, M. A., Lemey, P., Baele, G., Ayres, D. L., Drummond, A. J., & Rambaut, A. (2018, January). Bayesian phylogenetic and phylodynamic data integration using BEAST 1.10. *Virus Evolution*, *4*(1). Retrieved 2019-08-05, from https://academic.oup.com/ve/article/4/1/vey016/5035211 doi: 10.1093/ve/vey016

Suitso, I., Jõgi, E., Talpsep, E., Naaber, P., Lõivukene, K., Ots, M.-L., ... Nurk, A. (2010, March). Protective effect by Bacillus smithii TBMI12 spores of Salmonella serotype enteritidis in mice. *Beneficial Microbes*, *1*(1), 37–42. doi: 10.3920/BM2008.1001

Tajima, F. (1989, November). Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. *Genetics*, *123*(3), 585–595.

Takami, H., & Horikoshi, K. (1999). Reidentification of Facultatively Alkaliphilic Bacillus sp. C-125 to Bacillus halodurans. *Bioscience, Biotechnology, and Biochemistry*, *63*(5), 943–945. doi: 10.1271/bbb.63.943

Takami, H., Nakasone, K., Takaki, Y., Maeno, G., Sasaki, R., Masui, N., ... Horikoshi, K. (2000, November). Complete genome sequence of the alkaliphilic bacterium Bacillus halodurans and genomic sequence comparison with Bacillus subtilis. *Nucleic Acids Research*, *28*(21), 4317–4331. Retrieved 2019-10-04, from https://www.ncbi.nlm.nih.gov/pmc/articles/PMC113120/

Tam, N. K. M., Uyen, N. Q., Hong, H. A., Duc, L. H., Hoa, T. T., Serra, C. R., ... Cutting, S. M. (2006, April). The Intestinal Life Cycle of Bacillus subtilis and Close Relatives. *Journal of Bacteriology*, *188*(7), 2692–2700. Retrieved 2019-09-18, from https://www.ncbi.nlm.nih.gov/pmc/articles/PMC1428398/ doi: 10.1128/JB.188.7.2692-2700.2006

Tatu, A. L., Clatici, V., & Cristea, V. (2016, October). Isolation of Bacillus simplex strain from Demodex folliculorum and observations about Demodicosis spinulosa. *Clinical and Experimental Dermatology*, *41*(7), 818–820. doi: 10.1111/ced.12893

Tirumalai, M. R., Rastogi, R., Zamani, N., Williams, E. O., Allen, S., Diouf, F., ... Fox, G. E. (2013, June). Candidate Genes That May Be Responsible for the Unusual Resistances Exhibited by Bacillus pumilus SAFR-032 Spores. *PLOS ONE*, *8*(6), e66012. Retrieved 2019-07-14, from https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0066012 doi: 10.1371/journal.pone.0066012

Tumapa, S., Holden, M. T., Vesaratchavest, M., Wuthiekanun, V., Limmathurotsakul, D., Chierakul, W., ... Peacock, S. J. (2008, April). Burkholderia pseudomallei genome plasticity associated with genomic island variation. *BMC Genomics*, *9*(1), 190. Retrieved 2019-08-05, from https://doi.org/10.1186/1471-2164-9-190 doi: 10.1186/1471-2164-9-190

Val-Calvo, J., Miguel-Arribas, A., Gago-Córdoba, C., López-Pérez, A., Ramachandran, G., Singh, P. K., ... Meijer, W. J. J. (2019, April). Draft Genome Sequences of Sporulation-Impaired Bacillus pumilus Strain NRS576 and Its Native Plasmid p576. *Microbiology Resource Announcements*, *8*(16). doi: 10.1128/MRA.00089-19

Vary, P. S., Biedendieck, R., Fuerch, T., Meinhardt, F., Rohde, M., Deckwer, W.-D., & Jahn, D. (2007, October). Bacillus megaterium–from simple soil bacterium to industrial protein production host. *Applied Microbiology and Biotechnology*, *76*(5), 957–967. doi: 10.1007/s00253-007-1089-3

Vernot, B., Stolzer, M., Goldman, A., & Durand, D. (2008, September). Reconciliation with Non-Binary Species Trees. *Journal of Computational Biology*, *15*(8), 981–1006. Retrieved 2019-08-05, from https://www.liebertpub.com/doi/abs/10.1089/cmb.2008.0092 doi: 10.1089/cmb.2008.0092

Wagle, A. R., Dixit, Y. M., & Vakil, B. V. (2019, September). Scale Up Studies for Polyhydroxyalkanoate Production by a Bacillus flexus Strain with Industrial Potential. *Indian Journal of Microbiology*, *59*(3), 383–386. doi: 10.1007/s12088-019-00807-z

Waller, L. N., Fox, N., Fox, K. F., Fox, A., & Price, R. L. (2004, July). Ruthenium red staining for ultrastructural visualization of a glycoprotein layer surrounding the spore of Bacillus anthracis and Bacillus subtilis. *Journal of Microbiological Methods*, *58*(1), 23–30. doi: 10.1016/j.mimet.2004.02.012

Wang, J.-P., Liu, B., Liu, G.-H., Chen, Q., Pan, Z., Zheng, X.-F., & Chen, M. (2016, February). Draft Genome Sequence of Bacillus muralis LMG 20238t (DSM 16288), a Spore-Forming Bacterium Isolated from Deteriorated Mural Paintings. *Genome Announcements*, *4*(1). doi: 10.1128/genomeA.01691-15

Wang, L.-T., Lee, F.-L., Tai, C.-J., & Kasai, H. (2007). Comparison of gyrB gene sequences, 16s rRNA gene sequences and DNA–DNA hybridization in the Bacillus subtilis group. *International Journal of Systematic and Evolutionary Microbiology,*, *57*(8), 1846–1850. Retrieved 2019-09-18, from https://www.microbiologyresearch.org/content/journal/ijsem/10.1099/ijs.0.64685-0 doi: 10.1099/ijs.0.64685-0

Watabe, K. (2013). Overview of study on bacillus subtilis spores. *YAKUGAKU ZASSHI*, *133*(7), 783–797. doi: 10.1248/yakushi.13-00143

Wells-Bennik, M. H., Eijlander, R. T., den Besten, H. M., Berendsen, E. M., Warda, A. K., Krawczyk, A. O., ... Abee, T. (2016). Bacterial Spores in Food: Survival, Emergence, and Outgrowth. *Annual Review of Food Science and Technology*, *7*(1), 457–482. Retrieved 2019-10-08, from https://doi.org/10.1146/annurev-food-041715-033144 doi: 10.1146/annurev-food-041715-033144

Wu, C. H., Mulchandani, A., & Chen, W. (2008, April). Versatile microbial surface-display for environmental remediation and biofuels production. *Trends in Microbiology*, *16*(4), 181–188. doi: 10.1016/j.tim.2008.01.003

Yang, Z. (1997, October). PAML: a program package for phylogenetic analysis by maximum likelihood. *Bioinformatics*, *13*(5), 555–556. Retrieved 2019-08-01, from https://academic.oup.com/bioinformatics/article/13/5/555/420769 doi: 10.1093/bioinformatics/13.5.555

Yang, Z. (2007, August). PAML 4: Phylogenetic Analysis by Maximum Likelihood. *Molecular Biology and Evolution*, *24*(8), 1586–1591. Retrieved 2019-08-01, from https://academic.oup.com/mbe/article/24/8/1586/1103731 doi: 10.1093/molbev/msm088

Yang, Z., Nielsen, R., Goldman, N., & Pedersen, A.-M. K. (2000, May). Codon-Substitution Models for Heterogeneous Selection Pressure at Amino Acid Sites. *Genetics*, *155*(1), 431–449. Retrieved 2019-08-13, from https://www.genetics.org/content/155/1/431

Yang, Z., Wong, W. S. W., & Nielsen, R. (2005, April). Bayes Empirical Bayes Inference of Amino Acid Sites Under Positive Selection. *Molecular Biology and Evolution*, *22*(4), 1107–1118. Retrieved 2019-08-13, from https://academic.oup.com/mbe/article/22/4/1107/1083468 doi: 10.1093/molbev/msi097

Zhang, J., Fitz-James, P. C., & Aronson, A. I. (1993, June). Cloning and characterization of a cluster of genes encoding polypeptides present in the insoluble fraction of the spore coat of Bacillus subtilis. *Journal of Bacteriology*, *175*(12), 3757–3766. Retrieved 2019-10-05, from https://www.ncbi.nlm.nih.gov/pmc/articles/PMC204792/

Zhang, J., Wang, J., Fang, C., Song, F., Xin, Y., Qu, L., & Ding, K. (2010). Bacillus oceanisediminis sp. nov., isolated from marine sediment. *International Journal of Systematic and Evolutionary Microbiology,*, *60*(12), 2924–2929. Retrieved 2019-10-04, from https://www.microbiologyresearch.org/content/journal/ijsem/10.1099/ijs.0.019851-0 doi: 10.1099/ijs.0.019851-0

Zhang, Y., Li, X., Hao, Z., Xi, R., Cai, Y., & Liao, X. (2016, June). Hydrogen Peroxide-Resistant CotA and YjqC of Bacillus altitudinis Spores Are a Promising Biocatalyst for Catalyzing Reduction of Sinapic Acid and Sinapine in Rapeseed Meal. *PLOS ONE*, *11*(6), e0158351. Retrieved 2019-10-04, from https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0158351 doi: 10.1371/journal.pone.0158351

Zhaxybayeva, O., & Doolittle, W. F. (2011, April). Lateral gene transfer. *Current biology: CB*, *21*(7), R242–246. doi: 10.1016/j.cub.2011.01.045

Zhu, B., & Stülke, J. (2018, January). SubtiWiki in 2018: from genes and proteins to functional network annotation of the model organism Bacillus subtilis. *Nucleic Acids Research*, *46*(D1), D743–D748. Retrieved 2019-07-14, from https://academic.oup.com/nar/article/46/D1/D743/4372578 doi: 10.1093/nar/gkx908

Zilhão, R., Naclerio, G., Henriques, A. O., Baccigalupi, L., Moran, C. P., & Ricca, E. (1999, April). Assembly Requirements and Role of CotH during Spore Coat Formation in Bacillus subtilis. *Journal of Bacteriology*, *181*(8), 2631–2633. Retrieved 2019-10-05, from https://www.ncbi.nlm.nih.gov/pmc/articles/PMC93693/

Zothansanga, R., Senthilkumar, N., & Gurusubramanian, G. (2016). Diversity and Toxicity of Bacillus thuringiensis from Shifting Cultivation (Jhum) Habitat. *Biocontrol Science*, *21*(2), 99–111. doi: 10.4265/bio.21.99