



**UNIVERSIDAD AUTÓNOMA DEL
ESTADO DE MORELOS**

UNIVERSIDAD AUTÓNOMA DEL ESTADO DE MORELOS
INSTITUTO DE INVESTIGACIÓN EN CIENCIAS BÁSICAS Y APLICADAS
CENTRO DE INVESTIGACIÓN EN DINÁMICA CELULAR

**“Estudio de la conservación de las proteínas implicadas en el
compromiso para esporular y su distribución en bacterias.”**

TESIS

QUE PARA OBTENER EL GRADO DE

DOCTOR EN CIENCIAS

PRESENTA

M. en C. Lizeth Soto Avila.

DIRECTOR DE TESIS

Dra. Rosa María Gutiérrez Ríos.

CUERNAVACA, MORELOS

Marzo 2020.



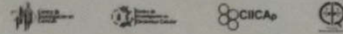
UNIVERSIDAD AUTÓNOMA DEL
ESTADO DE MORELOS



INSTITUTO DE INVESTIGACIÓN EN CIENCIAS BÁSICAS Y APLICADAS

Coordinación de Programas Educativos

Posgrado en Ciencias



DR. VICTOR BARBA LÓPEZ
COORDINADOR DEL POSGRADO EN CIENCIAS
PRESENTE

Atendiendo a la solicitud para emitir DICTAMEN sobre la revisión de la TESIS titulada “Estudio de la conservación de las proteínas implicadas en el compromiso para esporular y su distribución en bacterias” que presenta la alumna Lizeth Soto Avila (5620100808) para obtener el título de **Doctor en Ciencias**.

Nos permitimos informarle que nuestro voto es:

NOMBRE	DICTAMEN	FIRMA
Dr. Enrique Merino Pérez IBT-UNAM	Aprobado	
Dr. Armando Hernández Mendoza CIDC-UAEM	Aprobado	
Dr. Jorge Luis Folch Mallol CEIB-UAEM	Aprobado	
Dr. Ramón Gonzalez García-Conde CIDC-UAEM	Aprobado	
Dr. Ramón Alberto Batista García CIDC-UAEM	Aprobado	
Dra. Sonia Dávila Ramos CIDC-UAEM	Aprobado	
Dr. Rosa María Gutiérrez Ríos IBT-UNAM	Aprobado	

El presente trabajo se desarrollo en el departamento de Microbiologia Molecular del Instituto de Biotecnología, perteneciente a la Universidad Nacional Autónoma de México, bajo la dirección de la Dra. Rosa María Gutiérrez Ríos.

Para el desarrollo de este trabajo se contó con los recursos de la beca del Consejo Nacional de Ciencia y Tecnología (CONACYT) con el número de becario 248531.

Dedicatorias.

Esta tesis la dedico a toda mi familia, en especial a mi mamá a quien admiro mucho. Admiro tu fortaleza, tu paciencia y tu gran corazón. Agradezco tu apoyo en los momentos difíciles, como ahora, que tengo la responsabilidad de educar porque tengo la dicha de ser madre, siento que te amo más porque sé lo difícil que es.

A mi hermano Eduardo y familia.

A mi esposo Salvador, que es el amor de mi vida.

A mis hijos, Salvador, Sofia y Denisse con amor infinito.

Agradecimientos.

En primer lugar, quiero agradecer a mi tutora de tesis por todo su apoyo, dedicación y paciencia. Ha sido una gran bendición para mí aprender de la mejor, la más noble y admirable mujer. ¡Muchas gracias, Rosy!

Al Dr. Enrique Merino, por ser el mejor jefe de grupo. Por transmitirme su conocimiento y alegría.

A los miembros de mi comité tutorial. Al Dr. Armando Hernández, Dr. José Luis Puente y Dr. Ramón Alberto Batista por sus comentarios y sugerencias.

A todos los miembros del laboratorio a los que voy a extrañar mucho. En especial a María Luisa Tabche por su ayuda técnica en el laboratorio y al Ing. Ricardo Ciria por su enorme ayuda en todo lo de cómputo.

ÍNDICE

	ÍNDICE DE FIGURAS Y TABLAS.....	
I.	RESUMEN.....	7
II.	ABSTRACT.....	9
III.	INTRODUCCIÓN.....	11
	1. Modelo de estudio: <i>Bacillus subtilis</i>	11
	2. PROCESO DE ESPORULACIÓN EN <i>Bacillus subtilis</i>	12
	3. REGULACIÓN DEL PROCESO DE ESPORULACIÓN	14
	4. FORMACIÓN DEL SEPTO POLAR Y ACTIVACIÓN DE LOS FACTORES SIGMA F Y SIGMA E.....	16
	5. EL COMPROMISO PARA ESPORULAR.....	17
	6. PROTEÍNAS RELACIONADAS CON EL ENGULLIMIENTO.....	18
	6.2 EL COMPLEJO FORMADO POR LAS PROTEÍNAS SpoIIDMP.....	18
	6.3. FORMACIÓN DEL CANAL Q:AH	19
	6.3.2 Las proteínas del operon <i>spoIIIA</i>	20
	6.3.3 La proteína SpoIIQ.....	21
	7. LOS SISTEMAS DE SECRECIÓN EN BACTERIAS Y SU RELACIÓN CON EL CANAL DE ALIMENTACIÓN DE <i>B. subtilis</i>	21
IV.	ANTECEDENTES.....	23
V.	JUSTIFICACIÓN.....	27
VI.	PREGUNTA DE INVESTIGACIÓN.....	28
VII.	HIPÓTESIS.....	29
VIII	OBJETIVOS.....	20
	1. Objetivo General.....	29
	2. Objetivos Particulares.....	29
IX.	ESTRATEGIA METODOLÓGICA.....	30
X.	MÉTODOS.....	31
	1. Obtención de arquitecturas de las proteínas de estudio.....	31
	1.1 Identificación de proteínas implicadas en el compromiso para esporular....	31
	1.2 Evaluación de la composición de dominios.....	31
	2. Búsqueda de dominios homólogos en firmicutes.....	32
	3. Identificación de contextos genómicos conservados.....	32

	4. Análisis filogenético.....	33
XI.	RESULTADOS Y DISCUSIÓN.....	34
	Análisis de las proteínas que forman los complejos SpoIIIAA-H/spoIIQ y SpoIIDMP así como aquellas que favorecen su localización en organismos modelo.....	34
	Los componentes clave para completar engullimiento en <i>Bacillus</i> y en <i>Clostridia</i>	36
	Identificación de proteínas homologas al operon SpoIIIAA-AH en firmicutes.....	38
	Distribución del operon SpoIIIAA-AH, SpoIIQ, SpoIIB, GerM y SpoIIDMP en Firmicutes.....	47
	Conservación de las proteínas que forman el tubo de alimentación entre los organismos formadores de endosporas.....	50
	Distribución de los componentes del engullimiento en Firmicutes representativos.	66
	Proteínas que forman el canal de esporulación presentan contextos genómicos muy conservados.....	69
	Más allá de Firmicutes: <i>Ruegeria sp. NKCl-1</i> a Protoebateria que preserva Spo0A y la maquinaria del engullimiento.....	71
XII.	CONCLUSIONES.....	72
XIII.	PERSPECTIVAS.....	73
XIV.	REFERENCIAS.....	74

ÍNDICE DE FIGURAS.

Figura 1.	Representación esquemática del proceso de esporulación.....	11
Figura 2.	Representación esquemática de la cascada de regulación transcripcional, que se conoce para la generación de una spora madura en <i>Bacillus subtilis</i>	13
Figura 3.	Representación esquemática de la estructura del peptidoglicano en <i>Bacillus subtilis</i>	17
Figura 4.	Representación esquemática del complejo de secreción Q:AH en las membranas que rodean la prespora.....	18
Figura 5.	Representación esquemática de los diversos sistemas de secreción.....	19
Figura 6.	Esquema general de la estrategia metodológica.....	22
Figura 7.	Modelo esquemático de las proteínas requeridas para completar el engullimiento en <i>B. subtilis</i> y <i>C. difficile</i>	35
Figura 8.	Contexto genómico de las proteínas que forman en canal híbrido de esporulación.....	36
Figura 9.	Representación esquemática de las diferentes arquitecturas observadas por proteína usando la base de datos PfamA 3.1.....	38
Figura 10.	Representación que ejemplifica las arquitecturas propuestas para SpoIIIAA.....	43
Figura 11.	Especificidad de los motivos Walker A y Walker B encontrados en la clase Bacilli y clostridia.....	44
Figura 12.	Distribución de dominios homólogos a las proteínas del engulfasoma en los proteomas analizados.....	46
Figura 13.	Distribución por género de proteínas del engullimiento en Bacilli y Clostridia endoesporulantes.....	54
Figura 14.	Árbol filogenético basado en las secuencias del gen 16S rRNA y proteínas del engullimiento identificadas en 149 proteomas representativos.....	65

ÍNDICE DE TABLAS.

Tabla 1.	Proteínas seleccionadas para este estudio.	35
Tabla 2.	Resumen de las características usadas para seleccionar presumibles ortólogos (PO) de las proteínas involucradas en el engullimiento y el compromiso para esporular.	42
Tabla 3.	Clasificación de los Firmicutes. Resumen del proceso de curación..	50
Tabla 4.	Proteínas del operon <i>SpoIIIA</i> reanotadas.	56
Tabla 5.	Lista de K0's vecinos seleccionados, usando como referencia el K0 central.	69

I. RESUMEN

La esporulación en *Bacillus subtilis* (*B. subtilis*) es un sistema de desarrollo que implica la diferenciación de dos tipos celulares llamados prespora y célula madre. La limitación de nutrientes y la densidad celular inducen la acumulación de altas concentraciones del factor transcripcional Spo0A-P descrito como el regulador maestro de la esporulación. Spo0A-P controla la expresión del factor σ^F específico de la prespora y el factor σ^E específico de la célula madre. El programa de diferenciación de una espora se regula mediante la ultrasensibilidad de activación de factores sigma y circuitos de retroalimentación positiva. La progresión de la esporulación es reversible y ruidosa hasta el punto del compromiso, esto permite que las células inicien la esporulación pero pospongan por un mayor tiempo la decisión de compromiso. El punto en el que la bacteria se compromete a esporular se determina en la etapa II, donde se requiere la activación de los factores sigma SigF y SigE, para activar proteínas de señalización y reguladoras necesarias para promover las etapas subsecuentes. Después de la división asimétrica la célula madre y la prespora expresan un programa genético diferente, pero cada célula mantiene su propia membrana así como una comunicación constante entre ambas células. Se ha descrito que parte de la comunicación intercelular se realiza por medio del ensamble de un canal que cruza ambas membranas y que se ha propuesto sirve como un tubo de alimentación por el cual la célula madre nutre a la pre-espora. La formación del tubo de alimentación necesita al menos nueve proteínas en *Bacillus subtilis* para su formación, ocho proteínas SpoIIIA (SpoIIIAA-AH) bajo control de células madre y a la proteína SpoIIQ que está bajo el control de la célula hija. Otras proteínas descritas que interactúan con el canal son SpoIIDMP, GerM y SpoIIB, que en *Bacillus subtilis* también son cruciales para el engullimiento. Un estudio realizado mediante genómica comparativa en 72 genomas totalmente secuenciados sugieren que las proteínas codificadas en el operón *spoIIIA*, así como las proteínas del complejo SpoIIDMP, pueden ser considerados como firmas genómicas. En este trabajo, proponemos determinar la distribución y prevalencia de las proteínas que forman el tubo de alimentación, las proteínas del complejo SpoIIDMP, y las proteínas SpoIIB y GerM, que se propone interactúan con el canal y determinar si éstas son firmas genómicas de esporulación al evaluarlas en los Firmicutes actualmente secuenciados. Para comprobar esta hipótesis empleamos perfiles de proteínas con los que se construyeron arquitecturas de dominio que sirvieron para inspeccionar 4852 proteomas de bacterias completamente secuenciadas. Las arquitecturas identificadas fueron filtradas usando el contexto genómico que rodea cada secuencia identificada para definir ortólogos probables. Esto nos sirvió para evaluar a 953 Firmicutes, las cuales fueron sometidas a un proceso de curación manual para determinar su fenotipo

como esporulante o no esporulante. Los ortólogos probables encontrados en un compendio de especies de esporulación curadas manualmente se distribuyeron heterogéneamente entre las clases filogenéticas, pero se distribuyeron casi homogéneamente en géneros específicos. Los Firmicutes no esporulantes conservaron la arquitectura Spo0A y, en algunos casos, las arquitecturas del compromiso de esporular definidas como firmas genómicas, lo que sugiere que no una proteína sino la concurrencia de proteínas que definen una ruta de esporulación debería considerarse un mejor descriptor de una firma genómica. Finalmente, las proteínas en la ruta Spo involucradas en el compromiso de esporular se predijeron en *Ruegeria sp.* NKC1-1, una Alphaproteobacteria no esporulante. Sugerimos que una transferencia horizontal podría ser el origen de los genes Spo en esta especie si es clasificada correctamente.

II. ABSTRACT

Sporulation in *Bacillus subtilis* (*B. subtilis*) is a development system that involves the differentiation of two cell types called prespore and stem cell. Nutrient limitation, cell density, induced the accumulation of high concentrations of the master regulator Spo0A-P, transcription factor described being the master regulator of sporulation. The progression of sporulation is reversible and noisy to the point of commitment. Regulatory mechanisms as ultra-sensitivity and positive feedback modulate the differentiation program from a spore. Once the commitment decision is made, it has been proposed that an irreversible destination is generated. The point at which the bacterium commits to sporulate is determined in stage II, where the activation of the sigma factors SigF and SigE it is required, to activate signal and regulatory proteins, necessary to promote the following stages. Following asymmetric division, the mother cell engulfs the forespore, leaving it surrounded by two bilayer membranes. During the engulfment process, an essential channel apparatus is thought to cross both membranes to create a direct conduit between the mother cell and forespore. At least nine proteins in *Bacillus subtilis* are necessary for channel formation, including SpoIIQ under forespore control and the eight SpoIIIA proteins (SpoIIIAA to -AH) under mother cell control. Other proteins described to interact with the channel are SpoIIDMP, GerM, and SpoIIB, which in *Bacillus subtilis* are also crucial to the engulfment. Comparative genomic studies suggest that the *spoIIAA* operon, as well as SpoIIDMP complex, are considered genomic signatures. In this work, we propose to determine the distribution and prevalence of the proteins forming the feeding tube, the SpoIIDMP complex proteins, and the SpoIIB and GerM proteins, which are proposed to interact with the channel and determine if these are genomic sporulation signatures at evaluate them in Firmicutes currently sequenced. To test this hypothesis we use protein profiles with which domain architectures were built that served to inspect 4852 proteomes of fully sequenced bacteria. The identified architectures were filtered using the genomic context surrounding each hit to identify probable orthologs, which were mainly confined in 953 Firmicutes. Probable orthologs found in a compendium of manually curated sporulating species were heterogeneously distributed among phylogenetic classes but almost homogeneously distributed into specific genera. Nonsporulating Firmicutes conserved the Spo0A architecture and, in some cases, architectures of the commitment to sporulate defined as genomic signatures, suggesting that not one protein but the cooccurrence of proteins defining a sporulating pathway should be considered a better descriptor of a genomic signature. Finally, proteins in the Spo path involved in the commitment to sporulate were predicted in *Ruegeria sp.* NKC1-1, a nonsporulating

Alphaproteobacteria. We suggest that a horizontal transfer could be the origin of the Spo genes in this species if it is correctly classified.

III. INTRODUCCIÓN

1 MODELO DE ESTUDIO: *Bacillus subtilis*.

Bacillus subtilis es una bacteria del suelo que ha sido ampliamente usada tanto en la industria como en la investigación. En la industria se ha utilizado para la obtención de productos comerciales como la elaboración de probióticos y producción de enzimas usadas en la elaboración de alimentos, cosméticos y productos biológicos de uso agrícola. Por su parte, en investigación se ha utilizado como organismo modelo representativo del grupo de bacterias Gram-positivas. La gran mayoría de cepas de *Bacillus subtilis* utilizadas en la investigación derivan de la cepa de *B. subtilis* 168 una cepa auxótrofa de triptófano. Entre sus características destaca su capacidad para formar esporas, que implica un proceso de diferenciación celular. También se sabe que *Bacillus subtilis* 168 (*B. subtilis*) es una cepa competente natural, es decir que en ciertas condiciones incorpora ADN foráneo de manera natural (Anagnostopoulos y Spizizen, 1961), lo cual facilita la manipulación genética, además de que su genoma está totalmente secuenciado (Kunst *et al.*, 1997). Por las características antes mencionadas *B. subtilis* es un organismo modelo en el estudio de formación de endoesporas y diferenciación celular.

La formación de endoesporas es un ejemplo de un proceso de desarrollo, donde cada célula, la prespora y la célula madre siguen un destino celular distinto. La gran mayoría de las bacterias capaces de realizar este proceso, pertenecen a las clases Clostridia y Bacilli del filo firmicutes, sin embargo se han encontrado en otras clases de firmicutes, especialmente en la clase Negativicutes y en Heliobacterias que están más relacionadas a Clostridias (Tocheva *et al.*, 2011, Galperin *et al.*, 2013).

Las bacterias esporulantes siguen una secuencia de eventos que permite controlar cada etapa de la esporulación, desde la cascada de señalización que da inicio a la esporulación, la decisión del compromiso para esporular y hasta lograr la germinación cuando las condiciones mejoran. El control de este proceso es realizado por la combinación de diversos mecanismos de regulación, como la interacción de distintas proteínas formando módulos regulatorios, vías de señalización y regulación post-traducciona (Piggot y Losick 2002).

2 PROCESO DE ESPORULACIÓN EN *Bacillus subtilis*.

Las células de *B. subtilis* responden a condiciones ambientales adversas mediante la puesta en marcha de distintos destinos celulares. Se sabe que la concentración celular de Spo0A-P regula los procesos de canibalismo, producción de biofilm y la esporulación, siendo la formación de esporas la última estrategia de supervivencia. Las esporas son capaces de resistir y sobrevivir a situaciones extremas como temperaturas elevadas, desecación y radiaciones durante largos periodos de tiempo, gracias a la deshidratación del núcleo de la espora y la compactación del ADN cromosómico. Cuando las condiciones ambientales vuelven a ser favorables para el desarrollo de la bacteria, la espora puede germinar y reiniciar su crecimiento vegetativo (Higgins y Dworkin, 2012).

El proceso de esporulación inicia por la falta de nutrientes, aunque no es el único factor desencadenante del inicio del proceso, sino que la bacteria posee un sistema de detección de distintas señales ambientales, altamente complejo y sofisticado. Un aspecto importante es que no todas las células inician la esporulación al mismo tiempo, ni todas dan lugar a la formación de esporas. Se ha sugerido que la esporulación ocurre de una forma estocástica, en la que solamente una fracción de la población celular inicia este proceso aun en cultivos sincronizados (Chastanet *et al.*, 2010).

El programa que gobierna la esporulación ha sido ampliamente caracterizado a nivel de los genes involucrados en cada etapa del proceso. El proceso ha sido dividido en siete etapas o estados morfológicos (Figura 1), que pueden ser identificados mediante microscopia electrónica y son resumidos a continuación (Schultz *et al.*, 2009). No se reconoce como tal una etapa I, porque es cuando se lleva a cabo la duplicación del cromosoma y este proceso lo realizan tanto las células vegetativas como las que inician la esporulación. Posteriormente, la etapa II o estado morfológico II del proceso de esporulación se caracteriza por la división asimétrica mediante la formación de un septo polar. La molécula principal responsable de este proceso es la proteína FtsZ, homóloga a la tubulina en células eucariotas. Esta proteína polimeriza en forma de filamentos, dando lugar al anillo Z. La acumulación de FtsZ y SpoIIE hace que el anillo Z se sitúe en uno de los polos, dando lugar a la división asimétrica. La formación del septo da lugar a dos células distintas, una célula pequeña llamada pre-espora donde se activa la expresión del factor Sigma F y una célula grande llamada célula madre donde se activa la expresión del factor Sigma E. La activación de los factores específicos en cada célula son los que determinan su destino celular: la pre-espora se desarrolla en espora madura, mientras que la célula madre, que

nutre el desarrollo de la espora, se lisa al final de la esporulación para liberar la espora madura.

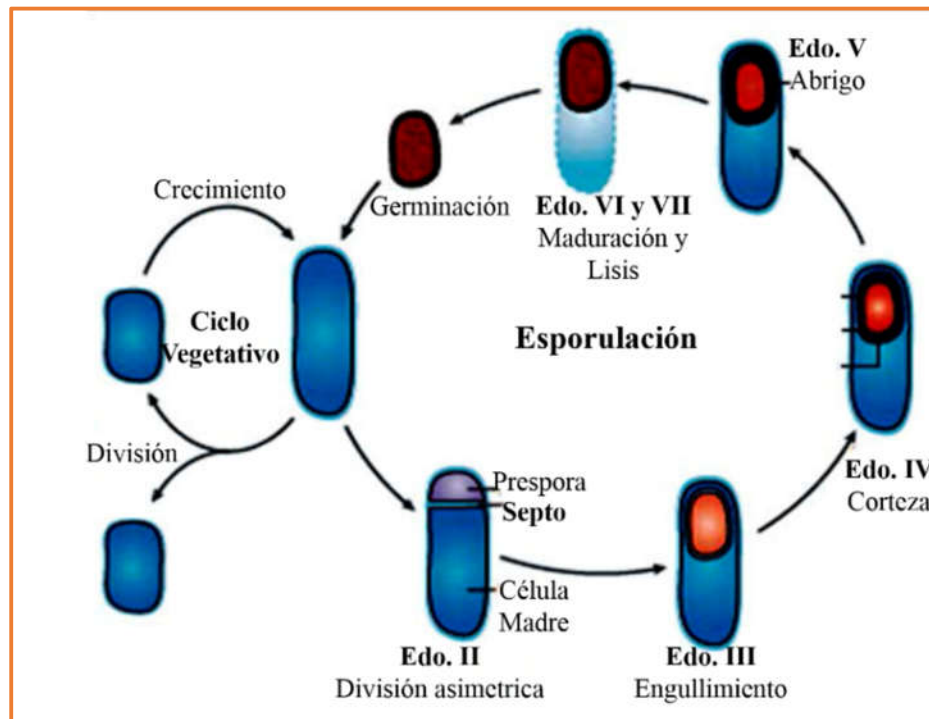


Figura 1. Representación esquemática del proceso de esporulación en *Bacillus subtilis*. El proceso ha sido dividido en siete estados morfológicos. Imagen tomada de Errington (2003).

Después de la división asimétrica solo un tercio del cromosoma está en la cámara pequeña y una DNA translocasa, SpoIIIE, localizada en el centro del septo empuja el resto del cromosoma (Wu *et al.*, 1994). Después en el estado morfológico III se lleva a cabo el engullimiento de la pre-espora por la célula madre. En esta etapa del proceso se requiere adelgazar el peptidoglicano del septo para darle flexibilidad a la membrana y facilitar su migración alrededor de la pre-espora, esta función es llevada a cabo por el complejo de proteínas SpoIIDMP que son activadas por σ^E . Una vez completado el engullimiento se activa el factor σ^G que permite que la síntesis de grandes cantidades de proteínas de bajo peso molecular que rodean el ADN, previniendo así que éste sufra daños.

A continuación se forman estructuras que van a proteger a la espora conocidas como corteza que se desarrolla en el estado IV y el abrigo en el estado V. El abrigo es un almacén de alrededor de 70 proteínas que envuelven a la prespora formando una estructura de múltiples capas. Finalmente en los estados VI y VII se da la maduración de espora y la

célula madre se lisa liberando a la espora madura, que puede permanecer latente largos periodos de tiempo, pero tiene el potencial de volver a germinar en cuanto las condiciones son favorables (de Hoon *et al.*, 2010).

3 REGULACIÓN DEL PROCESO DE ESPORULACIÓN

El regulador principal de la esporulación es el factor de transcripción Spo0A, cuya actividad es controlada por una cascada de fosforelevo mediada por las cinasas KinA, KinB, KinC, KinD y KinE en *B. subtilis*, que a través de las fosfatasa Spo0F y Spo0B, fosforilan a Spo0A. Spo0A-P activa la transcripción de genes específicos de la esporulación, tales como *spoIIA*, *spoIIE* y *spoIIG*, y causa el cambio en la transcripción de más de 500 genes, incluyendo al menos 120 genes blanco (Molle *et al.*, 2003).

La figura 2 representa la cascada de regulación transcripcional que se conoce para la generación de una espora en *B. subtilis*.

A nivel de la regulación genética, el cambio del crecimiento vegetativo al inicio de la esporulación involucra el cambio del factor σ^A por el factor σ^H , así como la acumulación del factor de transcripción Spo0A en su forma fosforilada (Spo0A~P). El regulador transcripcional Spo0A-P actúa como un regulador indirecto del gen *sigH*, por medio de la represión de AbrB. El factor σ^H dirige la transcripción del operon *spoIIAA-spoIIAB-sigF*, que codifican para el factor σ^F , el antisigma SpoIIAB y el anti-antisigma SpoIIAA. SpoIIAB mantiene inactivo al factor σ^F , hasta que la fosfatasa SpoIIE defosforila y activa a SpoIIAA permitiendo que la interacción de SpoIIAA con SpoIIAB libere al factor σ^F .

Antes de que se complete la formación del septo los factores σ^F y σ^E son sintetizados en su forma inactiva y están presentes en ambos compartimentos, sin embargo, únicamente σ^F estará activo en la preespora y σ^E en la célula madre. El factor σ^F es el primer regulador específico de compartimento en ser activado y dirige la transcripción de aproximadamente 50 genes (Wang, *et al.*, 2006), incluyendo *sigG*, *spoIIQ* y *spoIIR* que codifica para la proteína SpoIIR. SpoIIR es una proteína de señalización secretada desde la preespora, que activa la expresión de la proteasa SpoIIGA y permite el procesamiento de *pro-sigE* en la célula madre.

El factor σ^E regula la expresión de genes en el compartimento de la célula madre y controla genes que están involucrados en una gran variedad de funciones morfológicas y regulatorias. Algunas de las funciones de estos son: genes requeridos para el engullimiento, genes involucrados en la vía que lleva a la activación de σ^G en la preespora y σ^K en la célula madre, y los genes necesarios para la síntesis y ensamblado de estructuras que protegen la espora que son la corteza y el abrigo. El factor σ^E es activo por un periodo

extenso por ser un componente esencial en la esporulación y regula más de 250 genes ejerciendo un punto de control en el proceso (Eichenberger *et al.*, 2003). Como se observa en la figura 2, σ^E controla la expresión de dos complejos de proteínas, el primero formado por SpoIIDMP y el segundo complejo es el canal de esporulación, el cual se transcribe mediante el operon *spoIIIA* y el gen *spoIIQ*. SpoIIIAH interactúa con SpoIIQ que se localiza en la membrana de la prespora y en conjunto son los responsables de la activación de σ^G .

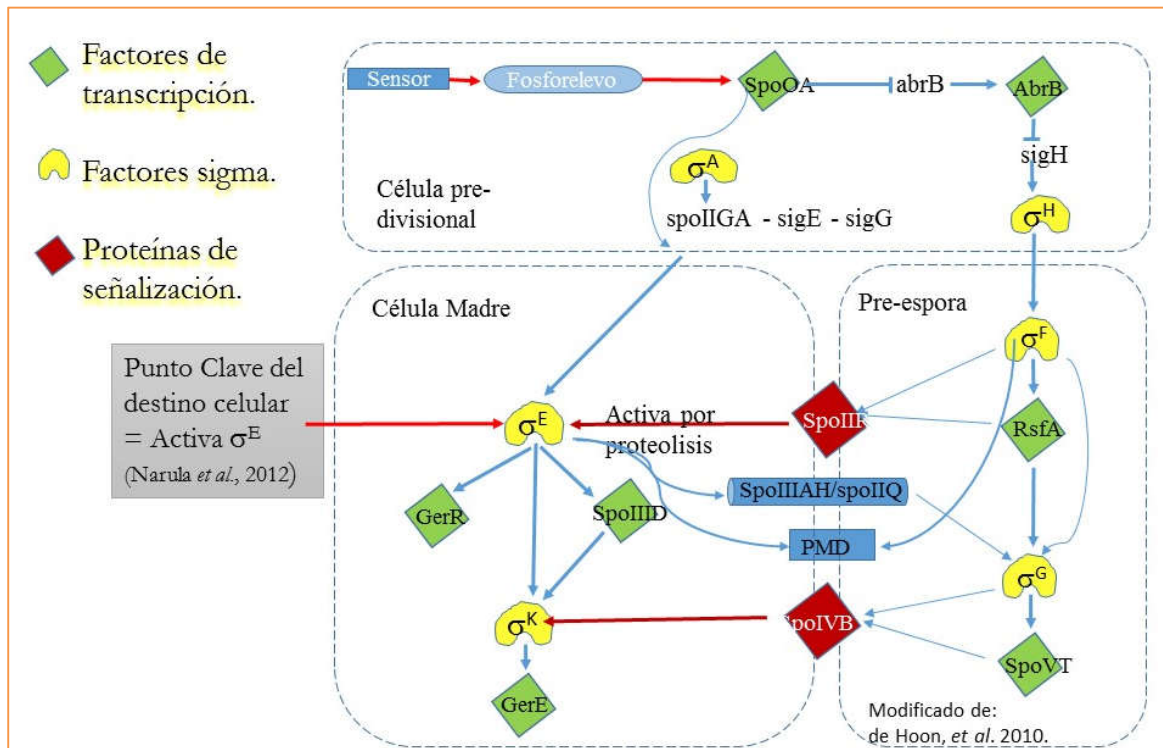


Figura 2. Representación esquemática de la cascada de regulación transcripcional, que se conoce para la generación de una espora madura en *Bacillus subtilis*. Una cascada fosforelevó señala y activa a Spo0A, misma que activa la expresión de los factores sigma específicos de compartimento (Modificada de de Hoon *et al.*, 2010).

El gen *spoIIIG* codifica para el factor σ^G , el cual es responsable de la expresión genética tardía en la prespora. Este factor controla la expresión de genes involucrados en la maduración de la espora y genes importantes para la germinación. En total σ^G controla la expresión de 95 genes (Wang *et al.*, 2006) entre ellos SpoIVFB que señala y ayuda a la activación de σ^K , el último factor sigma en aparecer en la cascada de regulación de la esporulación. El factor σ^K se expresa únicamente en la célula madre, y es sintetizado como

un precursor proteico inactivo, una vez activado estimula la transcripción de los pasos finales de ensamblaje, maduración y liberación de la espora madura.

En un estudio hecho por de Hoon y colaboradores sobre 24 firmicutes representativos formadores de esporas (de Hoon *et al.*, 2010), se estimó que los factores sigma de esporulación están conservados en un 80%, en tanto que el regulador Spo0A se encuentran conservado en todos los organismos formadores de esporas. El mismo estudio, reveló también que estos genes conservan una misma forma de activación por medio de una regulación que se organiza de forma jerárquica y con una lógica funcional (de Hoon *et al.*, 2010). En 2013 Fimlaid y colaboradores reportan la red de regulación que controla la formación de esporas en *Clostridioides difficile* (*C. difficile*), mostrando que a pesar de que los principales estados morfológicos son similares en ambas especies, la regulación jerárquica y secuencial de los factores sigma no, ya que σ^E es parcialmente independiente de σ^F , el factor σ^K no depende de σ^G , ni σ^G de σ^E (Fimlaid *et al.*, 2013).

Otras diferencias en proteínas estructurales también han sido descritas como es el caso de las proteínas de los complejos SpoIIIAA-AH:SpoIIQ (Q:AH) y SpoIIDMP sobre las que más adelante se profundizará en las diferencias estructurales y funcionales de este canal de comunicación entre la célula madre y la prespora.

4 FORMACIÓN DEL SEPTO POLAR Y ACTIVACIÓN DE LOS FACTORES SIGMA.

El inicio de la esporulación puede ser detectado en las células mediante la observación microscópica de la formación de un septo polar, que a diferencia del septo vegetativo es más delgado. Se sabe que algunas bacterias son capaces de iniciar el proceso sin que esté llegue a completarse, ya que la decisión del compromiso está determinada por proteínas que se expresan durante la etapa II y antes de que se complete la etapa III llamada también engullimiento de la prespora por la célula madre. El éxito de la espora reside en la compartimentación, y posterior acoplamiento de la expresión génica realizada por los factores σ^F y σ^E específicos de los compartimentos de la prespora y la célula madre respectivamente (Errington y Illing, 1992), mismos que son activados después que se completa la formación del septo polar. Se sabe que las células que activan σ^F pero no σ^E causan un aborto de la esporulación, mientras que las células que activan ambos factores producen esporas, estas observaciones demuestran que σ^E es el último punto de decisión en la esporulación (Narula *et al.*, 2012).

Para que la membrana de la célula madre logre rodear por completo a la prespora requiere del ensamblaje de dos complejos de proteínas de suma importancia para el engullimiento. Estos complejos son, el complejo SpoIIDMP (DMP) y el canal (Q:AH), el primero tiene funciones de hidrólisis de peptidoglicano y el segundo forma un canal de comunicación entre los dos compartimentos de la espora en desarrollo. Se ha reportado que en *B. subtilis* se requieren los dos complejos para completar el engullimiento, si existe mutación en cualquiera de sus componentes la esporulación se ve afectada, pero aún puede retomar el crecimiento vegetativo formando un estado de “espera o standby” antes de tomar la decisión del compromiso. Una de las proteínas clave para la determinación del compromiso es la proteína SpoIIQ que junto con SpoIIAH forman un canal que comunica ambos compartimentos y que se ha descrito presenta similitud a componentes del sistema de secreción tipo III (Camp and Losick, 2008).

5 EL COMPROMISO PARA ESPORULAR

El término compromiso, usado para cualquier proceso de diferenciación celular es el punto en el cual los cambios inducidos para el desarrollo de la célula se vuelve irreversible. La identificación y caracterización de los genes implicados en el compromiso ha sido un reto importante en biología del desarrollo (Welinder *et al.*, 2011). En *B. subtilis* el compromiso para esporular ha sido estudiado desde 1969 por Sterlini and Mandelstam a través de experimentos en los que se induce el proceso y se observa en que momento la cepa silvestre es incapaz de retomar su crecimiento vegetativo en presencia de nutrientes. Otros experimentos de expresión diferencial mostraron que la decisión del compromiso coincide con el inicio de la activación del factor Sigma F en la prespora y de Sigma E en la célula madre (Parker *et al.*, 1996), estos factores regulan genes de la etapa II y III de la esporulación. En 2005, Dworkin y colaboradores descubren que las proteínas SpoIIP y SpoIIQ juegan un papel importante en el compromiso para esporular en *B. subtilis* (Dworkin and Losick, 2005). En otros organismos no existe evidencia experimental del compromiso, pero si hay un gran avance en el conocimiento acerca de los componentes requeridos para completar el engullimiento en *C. difficile*. Se ha propuesto que una sola maquinaria compuesta por SpoIIDP interactuando con el canal Q:AH permite el engullimiento, por lo tanto indirectamente están relacionados con el compromiso para esporular (Dembek *et al.*, 2018). En cuanto a la regulación de los genes que componen estos complejos de proteínas, se ha demostrado que existen diferencias importantes. El factor σ^F regula la expresión de SpoIIQ en la prespora tanto en *B. subtilis* como en *C. difficile* (Ribis *et al.*, 2018). Sin embargo hay diferencias en la regulación de SpoIIP en

estos organismos. En *B. subtilis* tanto σ^F como σ^E controlan la expresión de SpoIIP, en tanto que en *C. difficile* SpoIIP es únicamente regulada por medio de σ^F . Las ocho proteínas del operon *spoIIIA* son reguladas por σ^E . Es posible que en otros organismos formadores de esporas estas proteínas también tengan un papel importante para el compromiso, ya que mediante estudios genómicos se ha determinado que tanto las proteínas del complejo SpoIIDMP como las del canal Q:AH se encuentran muy conservadas en las bacterias formadoras de esporas (Stragier *et al.*, 2002, Galperin *et al.*, 2012 y Abecasis *et al.*, 2013), aunque como se mencionó los estudios realizados hasta ahora en *C. difficile* muestran algunas diferencias en la regulación del compromiso.

6 PROTEÍNAS RELACIONADAS CON EL ENGULLIMIENTO.

Para poder completar la etapa del engullimiento en *B. subtilis* se requiere la actividad de la maquinaria de degradación de peptidoglicano que como se mencionó anteriormente es realizado por las proteínas SpoIID, SpoIIM y SpoIIP que forman el complejo SpoIIDMP expresado en la célula madre. Este complejo de proteínas es activado por el factor Sigma E (Eichenberger *et al.*, 2003) y también requiere de la proteína SpoIIB para su correcta localización (Perez *et al.*, 2000). Se ha propuesto que estas proteínas son indispensables para el engullimiento pero no para la viabilidad (Higgins y Dworkin, 2012) y se ha demostrado que el gen *spoIIP* es uno de los genes determinantes en el compromiso de la célula madre para esporular (Dwork and Losick, 2005). Cuando la actividad del complejo SpoIIDMP se reduce, la función de SpoIIQ unida a SpoIIIAH se vuelve indispensable para proseguir con el engullimiento funcionando como un segundo mecanismo para facilitar la migración de la membrana (Broder y Pogliano, 2006). Se ha descrito que la interacción de SpoIIQ-SpoIIIAH funciona como un sistema base, es decir que estas proteínas forman los componentes centrales del canal (Meisner *et al.*, 2012) y más recientemente se ha encontrado que en *B. subtilis* se requiere del producto del gen *gerM* para poder ensamblar esta plataforma basal (Rodrigues *et al.*, 2016). En las siguientes secciones se describirá con más detalle cada una de las proteínas aquí estudiadas.

6.2 EL COMPLEJO FORMADO POR LAS PROTEÍNAS SpoIIDMP.

En bacterias, la morfogénesis está directamente relacionada al remodelamiento de la pared celular (Morlot *et al.* 2010). El crecimiento de las bacterias, así como la formación de endosporas requiere de la hidrólisis y expansión del peptidoglicano (PG). Esta actividad es realizada por hidrolasas y sintetasas de PG respectivamente, por lo que la célula debe

modular finamente su expresión. Las proteínas biosintéticas de PG, proteínas de unión a penicilina (PBP) han sido bien estudiadas, sin embargo hay pocos estudios que describan el desensamblaje del PG, el cual es necesario para permitir el crecimiento. Muchas bacterias poseen un alto número de hidrolasas de peptidoglicano, y se ha mostrado que estas enzimas tienen una función redundante. Se ha propuesto que la hidrólisis de peptidoglicano es necesaria para permitir la expansión de la pared celular, la liberación de productos del recambio de peptidoglicano, la separación de las células hijas durante la división y la lisis celular (Shockman y Holtje, 1994; Vollmer *et al.*, 2007). En la bacteria Gram positiva *Staphylococcus aureus* se ha caracterizado un sistema de regulación de dos componentes esencial, este sistema formado por los genes *yycFG* que controla la actividad de varias hidrolasas tanto putativas como probadas (Dubrac y Msadek, 2004). Este mismo sistema YycG/YycF en *B. subtilis* controla la expresión de las hidrolasas CwlO y LytE, que son requeridas para el crecimiento de pared celular (Howell *et al.*, 2003).

El peptidoglicano o mureína es el elemento estructural de la pared celular, está formado por una secuencia de N-acetil-glucosamina (NAG) y Acido N-acetilmurámico (NAM). En *B. subtilis* las cadenas de glicano están unidos por un péptido L-Ala(1)-D-Glu(2)-DAP(3)-D-Ala(4) (DAP, ácido diaminopimelico) unido al grupo carboxil del NAM y con un enlace 4-3 DAP-Ala, como se muestra en la figura 3. En *C. difficile* la estructura del PG es diferente ya que presenta un alto nivel de NAG desacetiladas y las cadenas de azúcar están unidas por un enlace 3-3 DAP-DAP (Peltier *et al.*, 2011).

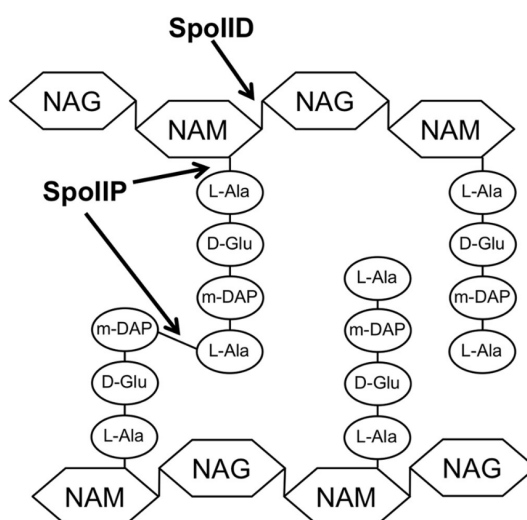


Figura 3. Representación esquemática de la estructura del peptidoglicano en *Bacillus subtilis*.

SpoIIP tiene actividad de endopeptidasa y amidasa, por lo que es capaz de liberar el péptido del NAM o el enlace 4-3 DAP-Ala, mientras SpoIID es una transglucosilasa lítica que corta los enlaces glucosídicos entre NAG y NAM (Morlot *et al.*, 2010).

6.3 FORMACIÓN DEL CANAL Q:AH.

Se sabe que la proteína SpoIIQ se encuentra localizada en el septo asimétrico del lado de la prespora, y junto con la proteína SpoIIIAH forma un canal que permite la comunicación entre el compartimento de la célula madre y la prespora (Rodrigues *et al.*, 2016).

Camp y Losick propusieron un modelo para el canal formado por Q:AH llamado “Tubo de alimentación”, ya que se ha observado que en ausencia de las proteínas del operon *spoIIIA* o de la proteína SpoIIQ la prespora pierde potencial metabólico y muere (Camp and Losick 2009). Aún se desconoce que metabolitos son transportados por este canal, pero se sabe que es esencial para mantener la integridad de la prespora (Fimlaid *et al.*, 2015). En la Figura 4 se muestra el modelo del complejo de secreción Q:AH.

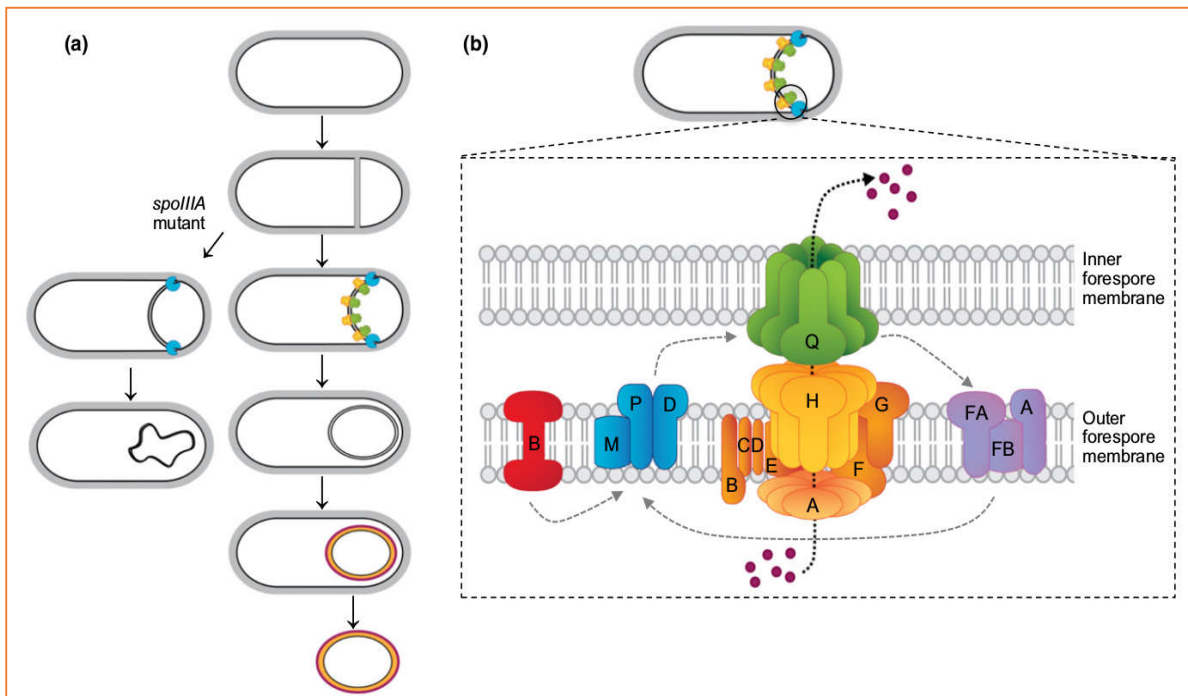


Figura 4. Representación esquemática del complejo de secreción Q:AH en las membranas que rodean la prespora.

- a) muestra los diferentes estado morfológicos de la esporulación. Adicionalmente, se observa que mutantes en proteínas del operon colapsan.
- b) Proteínas que forman complejos, con homología a diversos sistemas de secreción. La señal o metabolitos transportados por este canal es desconocido (Tomado de Crawshaw *et al.*, 2014).

Se indica a continuación el papel de cada componente del complejo Q:AH.

6.3.2 Las proteínas del operon *spoIIA-H*.

El operón *spoIIA-H* codifica para ocho proteínas, desde SpoIIAA hasta SpoIIAH. La mayoría de estas proteínas son asociadas a membrana, y expresadas bajo la regulación del factor σ^E de la célula madre. SpoIIAA presenta homología con la superfamilia de proteínas ATPasas, por lo que es probable que impulse el transporte de metabolitos a través del tubo de alimentación durante la esporulación de *B. subtilis* (Fimlaid *et al.*, 2015). Se sabe que las proteínas homologas de SpoIIAA están involucradas en la exportación de sustratos y en la biogénesis del pilus. SpoIIAB presenta homología con la ATPasa GspF, a pesar de que se conoce su estructura (Zeytuni *et al.*, 2018), su función sigue siendo desconocida. Se cree que su función es permitir la interacción de proteínas solubles y aquellas asociadas a membrana. SpoIIAC y SpoIIAD son proteínas pequeñas asociadas a membrana.

6.3.3 La proteína SpoIIQ.

La proteína SpoIIQ es miembro de la superfamilia Peptidasa M₂₃. Esta familia de proteínas consta de metaloendopeptidasas, endopeptidasas, y lipoproteínas para las que aún no se ha demostrado actividad catalítica (March-Bauer *et al.*, 2017). SpoIIQ es una proteína de 283 aminoácidos, con un peso molecular ~ 30 Kdaltones. Su expresión es controlada por el factor σ^F , por lo que su región promotora fusionada a GFP ha sido utilizada como marcador de esporulación (Dworkin and Losick, 2005). Análisis de la secuencia de aminoácidos de la proteína SpoIIQ muestran que tienen un solo dominio transmembrana (del residuo 21 al 42), un dominio Peptidasa_M23 (117- 216) bien caracterizado (Ledikov *et al.*, 2013), y además parece tener similitud con los motivos RnfC_N y Apocytochr_F_C depositados en la base de datos Pfam. Por su participación en actividades tanto reguladoras como catalíticas se ha propuesto como una proteína multifuncional. Participa en el remodelamiento de la pared celular importante para llevar a cabo el engullimiento y en la activación del factor Sigma G (Londono-Vallejo *et al.*, 1997).

7 LOS SISTEMAS DE SECRECIÓN EN BACTERIAS Y SU RELACIÓN CON EL CANAL DE ALIMENTACIÓN DE *B. subtilis*.

Los sistemas de secreción en bacterias son un sistema de transporte y comunicación. Estos sistemas están presentes en la mayoría de los procariontes, y se han estudiado en gran detalle en las bacterias Gram-negativas, por lo que se pensaba que eran exclusivos de éstas.

Actualmente se ha reportado la existencia de siete sistemas de secreción, encontrados tanto en gram positivas como en gram negativas (Figura 5).

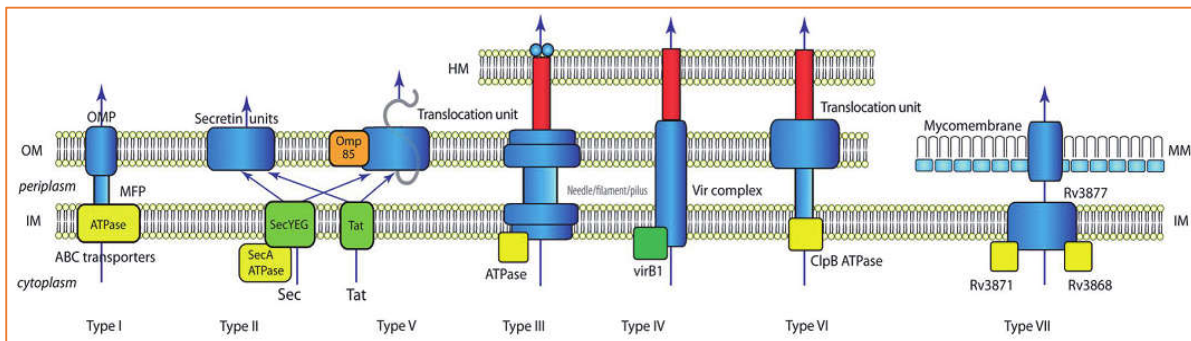


Figura 5. Representación esquemática de los diversos sistemas de secreción. Tomada de Tsai-Tien, *et al.*, 2009.

La diferencia más notable entre los diversos sistemas de secreción es que debido a la naturaleza didérmica de las gram negativas, se requiere de sistemas topológicamente más complejos. Las bacterias patógenas generan un canal de comunicación para establecer la relación huésped-parasito, este canal se forma mediante el ensamblaje de múltiples proteínas que son del sistema de secreción tipo tres (T3SS, por sus siglas en inglés). Estos sistemas son un elemento clave en las asociaciones patogénicas de gran cantidad de bacterias Gram-negativas tanto en animales como en plantas (Mota *et al.*, 2005). Este sistema de secreción constituido por más de 20 proteínas está codificado por los genes *hrp* (hypersensitive response and pathogenicity). Se sabe que en las bacterias gram-negativas las proteínas que conforman la arquitectura del sistema, requieren de una translocación para atravesar las dos membranas (membrana interna MI, y la membrana externa ME) separadas por el espacio periplásmico y la capa de peptidoglicano (Lee y Schneewind, 2001). Por su parte, en las bacterias Gram-positivas la secreción de proteínas hacia el exterior celular requiere del transporte a través de una sola membrana, sin embargo se han encontrado genes que participan en el proceso de secreción con un alto grado de similitud con los componentes centrales de la maquinaria de translocación (denominada sistema Sec) en bacterias Gram-negativas. Como se ha mencionado anteriormente, las proteínas del tubo de alimentación presentes en el operón *spoIIIA-H* durante el proceso de esporulación en *B. subtilis* se ensamblan en un canal formado por proteínas con homología a varios de los sistemas de secreción conocidos. Por su importancia y características peculiares ya descritas, en este proyecto se propone realizar un estudio de genómica comparativa sobre la arquitectura del canal de comunicación intracelular presente en los Firmicutes, tomando como modelo a las proteínas descritas en *B. subtilis* en donde se conoce más

profundamente el funcionamiento y su estructuración, y que ha sido poco estudiada en otros microorganismos endoesporulantes. Estos estudios permitirán en un futuro proponer un modelo evolutivo que indique la forma de cómo se podría haber ido conformando la arquitectura del canal y sugerir si el mecanismo funcional podría ser el mismo.

De manera interesante, se han reportado estudios realizados en algunas Clostridias sobre el complejo Q:AH. En donde SpoIIQ, presenta un dominio LytM intacto, que actuó como metaloendopeptidasa. En los Clostridia, a diferencia de los mecanismos de regulación conocidos en *B. subtilis* la regulación del proceso de esporulación no sigue la misma lógica jerárquica y secuencial.

IV. ANTECEDENTES.

Los factores σ^F y σ^E son determinantes para el compromiso, que es el punto en cual la formación de la espora se vuelve irreversible (Parker and Errington, 1996). Se ha descrito que la pérdida tanto del factor σ^E como del factor σ^F da lugar a un fenotipo diásporico en condiciones limitantes de nutrientes (Lewis *et al.*, 1994) que morfológicamente parecen iguales. Sin embargo, las células que pierden el factor σ^F , y que son diasporicas en condiciones limitantes, cuando se transfieren a medio rico forman un fenotipo denominado “outgrown” que se caracteriza por presentar células alargadas que retoman el crecimiento vegetativo, lo cual fue corroborado al seguir dentro de estos esporangios al factor transcripcional AbrB, que se ha usado como marcador de este tipo de crecimiento.

Por su parte, las células que pierden el factor Sigma E presentan un fenotipo diasporico en condiciones limitantes y que se reconoce como abortivo ya que la esporulación no prosigue y al ser transferidas a un medio rico tampoco puede generar un fenotipo “outgrown” (Parker *et al.*, 1996). Posiblemente a diferencia de la mutante en *sigF* que nunca puede comprometerse, la mutante en *sigE* ya no puede crecer porque ha expresado genes que comprometen a la prespora. Se ha sugerido, que la regulación de estos genes es controlada por σ^F y al perder este factor pierden la capacidad de reactivar la maquinaria necesaria para retomar su crecimiento vegetativo (Dworkin and Losick, 2005).

En un estudio posterior donde se retoma el tema del compromiso con la finalidad de conocer por medio de que genes la célula asegura el compromiso para esporular, se encontró que el fenotipo de la cepa mutante en *sigE*, que produce una célula diasporica abortiva, puede ser rescatado y formar células outgrown mediante la combinación de mutaciones en los genes *spoIIP* y *spoIIQ*, (Dworkin and Losick, 2005). Estos resultados surgen de la hipótesis de que los genes que se encuentran bajo la regulación de σ^F son los responsables del compromiso, ya que *B. subtilis* se compromete después de activar a σ^F y antes de activar a σ^E . Para identificar estos genes se analizaron cepas que además de llevar una mutación en *sigE* contenían una mutación en alguno de los genes del regulon σ^F conocido hasta entonces (Piggot y Losick, 2002). Los resultados, mostraron que ninguna doble mutante es capaz de retomar el crecimiento aún después de ser transferida a medio rico. Posteriormente, se investigó si la acción combinada de más de un gen regulado por σ^F podría revertir el fenotipo de la mutante *sigE*, y se analizó mediante su capacidad para crecer en medio rico a pesar de ya haber entrado a la vía de la esporulación. La combinación de mutaciones en los genes *spoIIP* y *spoIIQ* se analizó porque se tenía la hipótesis de que la mutante *sigE* ya no podía crecer por ser incapaz de desensamblar el

segundo septo y degradar de la pared celular de peptidoglicano, lo cual es esencial para el crecimiento. SpoIIP es una endopeptidasa y SpoIIQ tiene un dominio tipo LytM presente en proteínas metalopeptidasas que degradan la pared celular.

Experimentos previos, demostraron que la proteína SpoIIP impide la formación del segundo septo en la cepa silvestre si se expresa prematuramente y ayuda al adelgazamiento del septo degradando el peptidoglicano (Frandsen and Stragier 1995), también se observó que la pérdida del gen *spoIIP* da lugar a células con múltiples septos. Por su parte, la proteína SpoIIQ que tiene el dominio tipo LytM, en *B. subtilis* no tiene actividad de metalopeptidasa por que pierde los residuos implicados en la unión al metal que son importantes para la catálisis. En 2009 Doan y col. proponen un modelo en el cual SpoIIQ junto con 8 proteínas SpoIIIAA-H forman un novedoso aparato de secreción, con homología al sistema de secreción tipo III, y se ha sugerido que este canal permite a la célula madre nutrir a la prespora (Doan *et al.*, 2009). La decisión final de diferenciarse en una espora en *B. subtilis*, parece requerir más que solo pasar el umbral de Spo0A~P, y se sugiere que existen genes adicionales a *spoIIP* y *spoIIQ* para proceder con la fase II de esporulación, pero aun no es muy claro.

Los estudios para la caracterización de los componentes que promueven la esporulación en organismos formadores de endoesporas, se han realizado principalmente en organismos de la clase Bacilli (*B. subtilis*, *Bacillus anthracis* y *Paenibacillus polymyxa*). Así como en algunos Clostridia como *Clostridium acetobutylicum*, *C. difficile* y *Clostridium perfringens*, (Alsaker and Papoutsakis, 2005; Paredes *et al.*, 2005; Jones *et al.*, 2008; Lawley *et al.*, 2009; Underwood *et al.*, 2009; Steiner *et al.*, 2011). En el caso de otros Firmicutes, los experimentos se reducen a probar la presencia de endosporas en alguna etapa del desarrollo.

En 2012, Galperin y colaboradores hacen un análisis bioinformático del conjunto mínimo de genes para esporular encontrando que en Clostridia no existe ninguna proteína homóloga a SpoIIQ. De igual forma se había descrito la ausencia de proteínas homólogas a Spo0B en Clostridia, sin embargo, en nuestro grupo mediante análisis computacionales se logró demostrar la presencia de una versión de la proteína Spo0B conformada por solo el dominio homólogo Spo0B_a, para la que aún falta demostrar si esta versión es funcional.

Estudios recientemente publicados por Davidson y colaboradores (Davidson *et al* 2018), demostraron experimentalmente estas predicciones. Estas observaciones nos llevaron a realizar este trabajo, en donde proponemos realizar un estudio computacional similar para SpoIIQ y otros componentes propuestos en *B. subtilis* como esenciales en la formación del tubo de alimentación, con el fin de identificar la existencia de ortólogos probables de SpoIIQ en otras bacterias esporulantes.

En un estudio realizado recientemente en nuestro grupo se muestra que es posible localizar las proteínas de la cascada de señalización (quinasas, fosfatasas y fosfotransferasas específicamente) a través del estudio de los dominios que las conforman. Los resultados del trabajo antes mencionado indican que la sofisticada cascada de señalización en *Bacillus subtilis* parece haber sido adquirida mediante la ganancia de genes, ya que es poco conservada en el resto de las bacterias esporulantes (Martínez-Amador, tesis 2017 y Amador *et al.*, 2019).

Para proteínas de señalización intracelular como SpoIIDMP y el canal formado por Q:AH, existe un reporte de Abecasis en donde realizando búsquedas bioinformáticas mediante BlastP como herramienta muestra la distribución de SpoIIIA-H dentro y fuera de los firmicutes, para un set de 71 organismos, que incluía representantes de organismos endosporulantes, exosporulantes, mycobacterias, y no endosporulantes. En este estudio, la proteína SpoIIQ no fue tomada en cuenta. Sabemos que los complejos formados por estas proteínas han sido ampliamente estudiadas en el organismo modelo *B. subtilis*, y en la clase *Clostridia* en menor profundidad, sin embargo poco se sabe de este complejo en otros Firmicutes.

Algunos Firmicutes presentan genes que han sido clasificados como genes de esporulación debido a que presentan un fenotipo deficiente en la esporulación. Incluso algunos genes se han tomado como un marcador de esporulación debido a su importancia como el gen *spo0A*, *sspA* y *dpaAB* (Onyenwoke *et al.*, 2004). Sin embargo, la lista de genes clave o núcleo (core) para la esporulación se actualiza constantemente, conforme incrementa el número de Firmicutes secuenciados.

El presente estudio está enfocado en determinar la distribución y abundancia de las proteínas que conforman en canal de comunicación entre la célula madre e hija, lo que permitirá determinar cuáles de ellas pueden ser consideradas marcadores genómicos de la esporulación. Las proteínas seleccionadas incluyen a SpoIIIA-H, SpoIIQ, las proteínas del complejo DMP, así como SpoIIB y GerM, proteínas definidas en *B. subtilis* como importantes en el compromiso para esporular, o presentan una función indirecta mediante el ensamble y localización de los componentes de estos complejos. Se sabe que SpoIIP es conservado tanto en *Bacilli* como en *Clostridia*, aunque su regulación es un poco distinta, esta proteína es esencial para el engullimiento. Nuestra primera inspección hecha con la herramienta Gene Context tool (desarrollada en el laboratorio), que permite inspeccionar el contexto genómico de un gen y sus relaciones de ortología indica que SpoIIP está distribuido mayoritariamente en *Bacilli* y minoritariamente en *Clostridia*.

Las proteínas SpoIIIA-H, SpoIIP y SpoIIQ que estudiamos son proteínas no reguladoras, que han sido clasificadas como proteínas de señalización intracelular, se sabe que actúan en la etapa II de la esporulación y son reguladas por los factores σ^F y σ^E . Las proteínas seleccionadas para este estudio fueron elegidas debido a que: existe una gran cantidad de información sobre éstas respecto a su función biológica y estructura en el organismo modelo *B. subtilis*, y que con excepción de SpoIIQ se han reportado como proteínas esenciales para la esporulación. Dada su actividad crítica en el proceso de esporulación, en este estudio deseamos corroborar si prevalecen como genes clave para la esporulación dentro de los Firmicutes, o si existe algún o algunos organismos esporulantes que pierdan algún componente, ya que estos genes son llamados como “firmas de esporulación” y se han usado para la predicción de este proceso en organismos que no han podido ser cultivados o secuenciados pero sin evidencia experimental de presentar un fenotipo de esporulación. El análisis detallado de los dominios presentes en los genes involucrados en el compromiso en *B. subtilis*, ayudará a caracterizar de forma más precisa los complejos de las proteínas DMP y Q:AH entre las diferentes clases de organismos formadores de endoesporas.

V. JUSTIFICACIÓN.

Se ha demostrado que el compromiso para esporular depende de los factores σ^F y σ^E , además de proteínas no reguladoras como SpoIIP y SpoIIQ. Esto ha sido ampliamente estudiado en el organismo modelo *B. subtilis*, en tanto que en otras bacterias esporulantes no se ha logrado la identificación de alguna proteína homóloga a SpoIIQ, y por lo tanto no se ha estudiado como se determina el compromiso en estas bacterias. Lo más interesante es que el dominio LytM presente en la proteína SpoIIQ está ampliamente distribuido más allá de los Firmicutes, lo que sugiere que podría tener otras funciones y que su origen es más ancestral. (esto no se ha demostrado). De hecho en el estudio de Adriano O. Henriques *et al* J. Bac. 2013, se muestra que SpoIIP es exclusiva de Firmicutes y en el caso de SpoIIQ, no se ha reportado ningún análisis. En función de lo anteriormente expuesto, proponemos estudiar la distribución filogenética de las arquitecturas de proteínas que forman los complejos DMP y Q:AH en otros genomas y a través de ello correlacionar si su presencia está relacionada únicamente con organismos capaces de esporular. El análisis preliminar de nuestro grupo muestra que algunos dominios de estas proteínas están ampliamente distribuidos en bacterias no esporulantes, pero es importante evaluar si presentan arquitecturas únicas o sitios ya sea motivos o residuos específicos que le provean la especificidad para la generación de un fenotipo esporulante.

VI. PREGUNTA DE INVESTIGACIÓN.

¿Las proteínas de los complejos SpoIIPDM y SpoIIAAH-SpoIIQ, así como las proteínas GerM y SpoIIB se encuentran conservadas dentro de los Firmicutes?

¿Existen dominios homólogos de las proteínas que forman los complejos DMP, el canal Q:AH y de las proteínas GerM y SpoIIB fuera de las bacterias esporulantes?.

VII. HIPOTESIS

La distribución de arquitecturas homologas de las proteínas de los complejos DMP, el canal Q:AH, GerM y SpoIIB es homogénea en los organismos formadores de esporas.

VIII. OBJETIVOS

Objetivo General:

Determinar la distribución y abundancia de arquitecturas homólogas a las proteínas SpoIIB y GerM así como las que forman los complejos SpoIIDMP y Q:AH en bacterias.

Objetivos Específicos:

1. Determinar a las proteínas que están implicadas en el compromiso para esporular tomando como referencia los Firmicutes modelos reportados en la literatura.
2. Determinar las arquitecturas de dominios que definen a cada proteína que forma parte de los complejos de estudio.
3. Evaluar la distribución y abundancia de arquitecturas de las proteínas que forman los complejos DMP y Q:AH, así como las de la proteína GerM y SpoIIB en Firmicutes.
4. Evaluar la distribución y abundancia de arquitecturas de las proteínas que forman los complejos DMP y Q:AH fuera de Firmicutes.

IX. ESTRATEGIA METODOLÓGICA

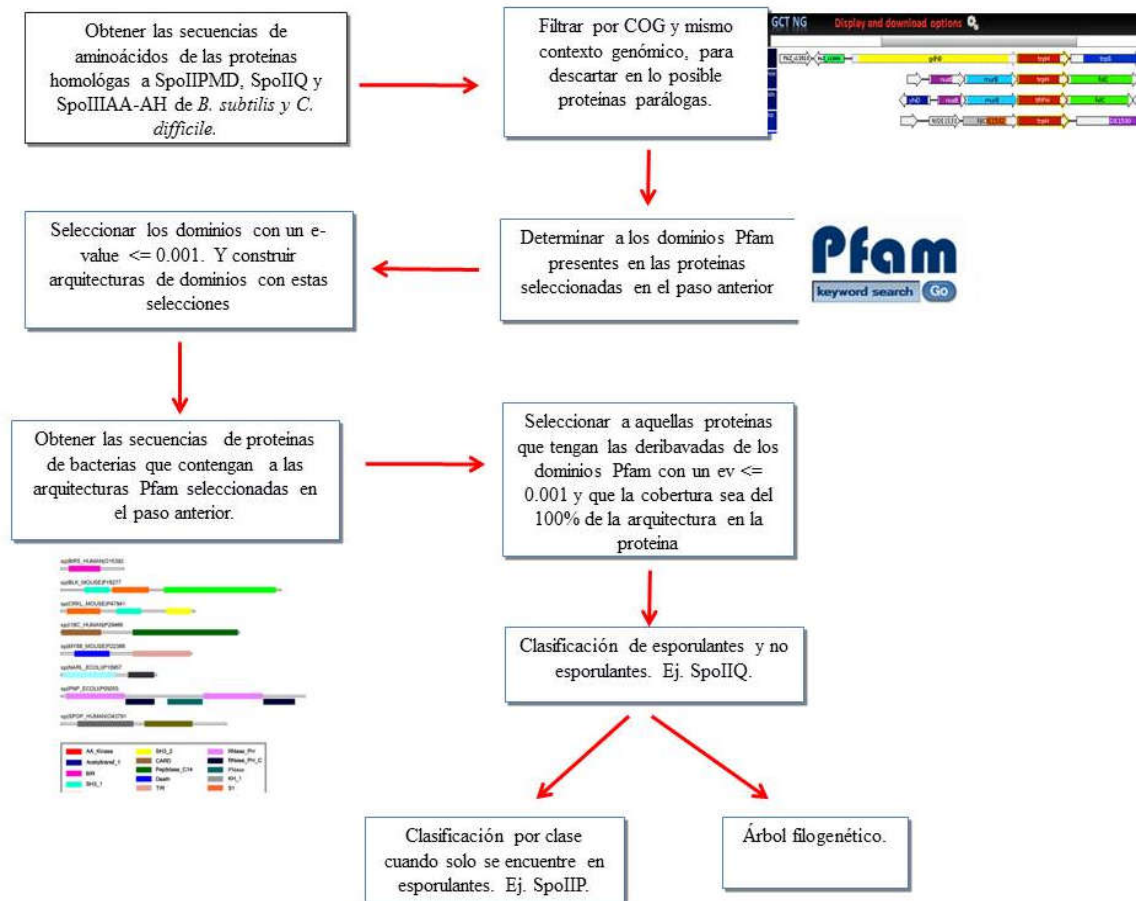


Figura 6. Esquema general de la estrategia metodológica.

X. MÉTODOS.

1. Obtención de arquitecturas de las proteínas de estudio.

1.1. Identificación de proteínas implicadas en el compromiso para esporular.

Se realizó una revisión en la literatura para identificar dentro de los Firmicutes a las proteínas implicadas en el compromiso para esporular. Una vez identificados a los Firmicutes que poseían a estas proteínas con evidencia experimental de ser funcionales y participar en la fase de compromiso elegimos a aquellas que formarían parte del complejo de hidrólisis de peptidoglicano o del tubo que comunica a la célula madre con la prespora, descartando a los factores sigma y a las proteínas de señalización. A las proteínas seleccionadas, se les identificaron los dominios que las componían y las regiones conservadas.

1.2. Evaluación de la composición de dominios.

De las proteínas de la Tabla 1, fueron obtenidas sus secuencias de aminoácidos. Sobre de ellos se hizo un análisis de la composición de dominios de proteína utilizando la herramienta “hmmscan” del paquete HMMER 3.1 y las matrices que describen a los perfiles de dominios de proteína almacenados en la base de datos Pfam A (Finn *et al.*, 2014), los cuales se muestran en la Tabla 2, columna 1. PfamA posee matrices curadas manualmente de estos perfiles, lo que permite hacer búsquedas confiables de dominios. Conservamos a todos aquellos hits con un valor de corte que fue un e-value (expectation value) $\leq 10^{-3}$, posteriormente, si la proteína presenta dos o más dominios superpuestos, se seleccionó el dominio con menor valor e-value y con una mayor cobertura de la proteína. Los dominios seleccionados fueron usados para construir la arquitectura de cada proteína (ver Figura 9). Las proteínas homólogas de SpoIIIAG presentan en Pfam el mismo perfil que las proteínas SpoIIIAF por lo que no puede distinguirse con las matrices Pfam, por lo cual se construyó una matriz específica. Para lo cual se utilizaron las secuencias de proteínas descritas como ortólogas utilizando la herramienta Gen Context Tools 3 desarrollado dentro del grupo que permite observar los contextos genómicos de proteínas homólogas y con ellos seleccionar ortólogos probables. De un total de 73 secuencias, se seleccionaron arbitrariamente a 27 secuencias no redundantes utilizando el programa CD-HIT con un valor de corte de 0.8. Las 27 secuencias fueron alineadas usando el programa MUSCLE, y convertidas a formato Stockholm que sirvió para alimentar el programa “hmmesbuild” del paquete HMMER 3.1, que permitió construir la matriz específica para SpoIIIAG.

2. Búsqueda de dominios homólogos en firmicutes.

Una vez identificadas a las proteínas se procedió a realizar una búsqueda de proteínas homologas. Para ello, utilizamos a los homólogos precalculados en la base de datos KEGG (Kanehisa *et al.*, 2018), lo cuales fueron consultados a través del servidor Gene Context Tool 3. Las consultas pueden realizarse buscando el nombre de la proteína, el resultado de esta consulta es una lista de proteínas homologas asociadas cuando existe a un valor COG (Cluster of Orthologous Genes) (Tatusov *et al.*, 2003), que indica la pertenecía de la proteína a un grupo de ortología. Una ventaja del uso de nuestro servidor es que ofrece una versión gráfica que permite auscultar con facilidad los resultados ya que resalta con el mismo color a todas las proteínas que comparten un COG. Más aún, permite observar de manera gráfica el contexto genómico asociado a cada homologo. No obstante, la versión actual tiene un total de 3880 genomas ya que no ha sido actualizado desde el 2013. Por lo que se decidió utilizar estos resultados como base para identificar los dominios comunes a los ortólogos y extender en una fase posterior esta búsqueda a los 4852 genomas totalmente secuenciados que actualmente tiene la base de datos KEGG. La última versión de estos datos de la base fue descargada en Abril 2019. Con esta última actualización obtuvimos un subgrupo de bacterias del phylum Firmicutes que consta de 953 miembros totalmente secuenciados.

3. Identificación de contextos genómicos conservados.

Dado que algunas de proteínas suelen tener copias de ellas en el genoma y más aún dado que los dominios que poseen algunas de nuestras proteínas de estudio están ampliamente distribuidos dentro y fuera de los Firmicutes. Se realizó un segundo filtrado utilizando como herramienta la conservación del contexto genómico. Basandonos en la idea de que los genes de organismos filogenéticamente cercanos conservan a sus genes vecinos y en ocasiones incluso de manera sinténica (Junier y Rivoire, 2013). Con este objetivo se identificaron a los genes vecinos de cada proteína en los Firmicutes y se les clasificó utilizando la asignaciones como ortólogos propuesta por KEGG, a través de los valores KO (ortólogos KEGG) (Kanehisa *et al.*, 2018). Los KO's de referencia (ver Tabla 5), muestran los KO's seleccionados dada su frecuencia. Aquellos KO's vecinos con una frecuencia de aparición mayor o igual a 2, fueron tomados como base para seleccionar a los mejores homólogos en Firmicutes y otros clados bacterianos. Una consideración importante, es que dentro de los KO's seleccionados, se mantuvieron aquellos KO's presentes en los contextos observados en *B. subtilis* (*Bacillus subtilis* subsp. *subtilis* 168) y *C. difficile* (*Clostridioides difficile* 630 (*Clostridium difficile* 630) (PCR-ribotype 012).

4. Búsqueda de dominios homólogos en bacterias.

Para identificar los dominios homólogos a proteínas que forman el canal de esporulación se extrajeron las matrices Pfam de cada dominio identificado que formaba parte de las arquitecturas, para búsquedas posteriores. Con las matrices seleccionadas se realizó un escaneo sobre todas las proteínas de los 4852 genomas bacterianos totalmente secuenciados presentes en la base de datos KEGG. Se consideró como un resultado positivo a aquellas proteínas con dominios en donde: a) la longitud de las proteínas homólogas fueran similar (entre el conjunto de proteínas ortólogas, la de mayor longitud) a la longitud de las proteínas modelo y b) que solo presentara las arquitecturas definidas, (es decir que si presenta otros dominios adicionales). De los modelos propuestos para *Bacillus* y *Clostridia*, se identificaron sus KO's, así como los KO's vecinos, los cuales se muestran en la Tabla 2, columna 4 y 5 respectivamente. A diferencia de los COG's, el KO identifica con mayor precisión al ortólogo del modelo. Los KO's de las proteínas del canal solo pertenecieron en general a Firmicutes. Así que para la identificación de homólogos dentro de Firmicutes se usaron los criterios descritos anteriormente y la conservación de al menos un KO en su contexto tomando arbitrariamente 3 genes río arriba y 3 genes río abajo. Se buscó además la conservación de las proteínas SpoIIIA-H en operón y se estimó la frecuencia de esta conservación.

5. Análisis filogenético

Para representar la distribución de las arquitecturas de proteínas que forman parte de los complejos DMP y Q:AH, así como SpoIIB y GerM se construyó un árbol de especies. El árbol se obtuvo de la siguiente manera: se obtuvieron las secuencias nucleotídicas correspondientes al gen 16S ribosomal, de los Firmicutes en los cuáles se localizaron las proteínas de estudio. Posteriormente para elegir organismos representativos, las secuencias 16S de 414 Firmicutes fueron filtradas usando el programa CDHIT v4.6 (Li et al. 2001) con un valor de corte de 0.9. Las secuencias seleccionadas fueron alineadas usando el programa SSU-ALIGN, y del alineamiento se generó una matriz de distancia mediante el programa DNAdist (Felsenstein et al., 2005) del paquete PHYLIP. Posteriormente se estimó la filogenia mediante 2 métodos, el Neighbor Joining y el Kitsch, ambos algoritmos también se encuentran en el paquete PHYLIP version 3.69. La visualización del árbol se realizó mediante en servidor iTol (Letunic y Bork 2019).

XI. RESULTADOS Y DISCUSIÓN:

Análisis de las proteínas que forman los complejos SpoIIDMP y SpoIIIAA-AH/SpoIIQ, así como aquellas que favorecen su localización en organismos modelo.

Las proteínas que se eligieron para este estudio son de gran importancia para el proceso de desarrollo de una espora, su función es esencial para completar la etapa del engullimiento en *Bacillus subtilis*. Se ha determinado que las proteínas del operon *spoIIIA* forman parte de un canal de comunicación entre la célula madre y la prespora durante la etapa II de esporulación. Este canal está formado principalmente por la proteína SpoIIQ localizada en el septo en la membrana de la prespora y SpoIIAH localizada en el septo del lado de membrana de la célula madre. Sin embargo se requiere de la presencia de todas las proteínas del operon *spoIIIA*, así como de la hidrólisis de peptidoglicano (realizada por el complejo DMP) para poder ensamblar y activar el complejo de secreción transmembrana (Rodríguez *et al.*, 2013). Por su parte, SpoIIB se sabe que juega un papel importante para la correcta localización del complejo DMP en *B. subtilis* (Aung *et al.*, 2007). SpoIIB se ha reportado como una proteína conservada en el grupo *Bacilli*, en tanto que el grupo *Clostridia* carece de la misma (Galperin *et al.*, 2012).

Otra proteína agregada a nuestro estudio es el producto del gen *gerM*, que codifica para la proteína GerM ya que se encontró que además de tener una función esencial en la germinación de la espora, también es importante para la localización tanto de SpoIIQ como de SpoIIAG en *Bacillus subtilis* (Rodrigues *et al.*, 2016), sin tener evidencia experimental de su función en otros organismos.

Investigaciones previas han sugerido que el sistema de comunicación entre la célula madre y la prespora (Q:AH y DMP) ha sido ganado vía transferencia horizontal, desde bacterias didérmicas como las proteobacterias. Una evidencia que apoya esta hipótesis es el hecho de que este sistema está compuesto de proteínas con similitud a diferentes sistemas de secreción (Doan *et al.*, 2009). Los sistemas de secreción están involucrados en funciones importantes, como la decodificación y resistencia a antibióticos, translocación de factores de virulencia, también están relacionados con la biogénesis de estructuras como pili y flagelos. Recientemente, Abby y colaboradores reportan un estudio en el cual se identifica las proteínas de secreción en genomas bacterianos de organismos didérmicos basados en modelos computacionales, por lo que el objetivo del presente proyecto será complementar esta información tomando como referencia los modelos biológicos del sistema SpoIIA-H y SpoIIQ, para poder determinar la distribución y abundancia de los dominios homólogos de estas proteínas de transporte, de secreción o de comunicación intracelular y analizar cuál es su patrón de variación entre las diferentes clases de Firmicutes. Se ha sugerido que aún falta

por descubrir nuevos sistemas ya que algunos organismos ambientales parecen no contar con ninguno de los sistemas de secreción descritos (Abby *et al.*, 2016).

Para iniciar nuestro estudio, definimos organismos modelo para los cuáles existe información sobre la regulación y expresión de las proteínas que forman estos complejos. Se tomaron a los organismos *B. subtilis* y *C. difficile* (ver Tabla 1). En ambos organismos se han estudiado las proteínas que forman el complejo DMP y las que forman el canal. En un análisis comparativo de las proteínas que forman los complejos en *Bacillus* y sus parientes más cercanos se ha observado que en general se requiere de todo el operón *spoIIIA* y *spoIIQ* para poder actuar como un sistema. Las proteínas SpoIIB y GerM se han estudiado únicamente en *B. subtilis* de forma experimental, y mediante análisis computacionales se ha reportado que no hay ortólogos de SpoIIB en Clostridias y que GerM por el contrario se encuentra ampliamente distribuida en otros fila incluyendo Actinobacterias, Cyanobacterias y Proteobacterias (Rigden y Galperin 2008).

Proteína	KEGG ID <i>B. subtilis.</i>	KEGG ID <i>C. difficile.</i>	Similitud con:
SpoIIIAA	BSU24430	CD630_11920	Superfamilia de ATPasas, involucradas en sistemas tipo II y IV.
SpoIIIAB	BSU24420	CD630_11930	Familia GspF y TadB/C. Sistema Tipo II y IV.
SpoIIIIAC	BSU24410	CD630_11940	Permeasas tipo ABC. Sistema Tipo I.
SpoIIIIAD	BSU24400	CD630_11950	Permeasas tipo ABC. Sistema Tipo I.
SpoIIIIAE	BSU24390	CD630_11960	Permeasas tipo ABC. Sistema Tipo I.
SpoIIIIAF	BSU24380	CD630_11970	Familia FlhB (YscU). Sistema Tipo III
SpoIIIIAG	BSU24370	CD630_11980	Familia YscJ/ FliF. Sistema Tipo III
SpoIIIIAH	BSU24360	CD630_11990	Familia YscJ/ FliF. Sistema Tipo III
SpoIIIQ	BSU36550	CD630_01250	Familia LytM.
SpoIID	BSU36750	CD630_01240	Familia transglicosilasas líticas (Nocadello <i>et al.</i> , 2016).
SpoIIM	BSU23530	CD630_12210	Familia proteínas integrales de membrana DUF95
SpoIIP	BSU25530	CD630_24690	

SpoIIB	BSU28060	* Sin ortólogo	
GerM	BSU28380	* Sin ortólogo	

Tabla 1. Proteínas seleccionadas para este estudio.

Tanto las proteínas del sistema de comunicación como las hidrolasas de peptidoglicano están asociadas a membrana y tienen similitud con algunos componentes de los sistemas de secreción reportados en bacterias gram negativos. Sin embargo, hasta el momento no hemos encontrado reporte además del análisis realizado por Abecasis, que defina si todos los componentes de este canal son genes exclusivos de esporulación, por lo que la clasificación de esporulantes y no esporulantes será clave para responder esta pregunta, además nos permitirá definir y clasificar cada proteína como esencial o accesoria.

Los componentes clave para completar engullimiento en *Bacillus* y en *Clostridia*.

La etapa de engullimiento ha sido mejor documentada en *B. subtilis* y en *C. difficile*. Observamos diferencias importantes en el número de componentes clave para completar el engullimiento, la forma en cómo se regulan y localizan las proteínas que forman los complejos DMP y Q:AH, e incluso la función de algunas proteínas. Estas notables diferencias son mejor visualizadas y resumidas usando dos modelos distintos para completar el engullimiento, como se puede ver la figura siguiente.

Con base en la información tomada de la literatura, definimos dos modelos para explicar las diferencias en cuanto a estructura y función del canal híbrido de esporulación. En el modelo *Bacilli* se requiere la actividad de dos complejos de proteínas, ya que el complejo DMP realiza una función distinta de la función del canal, y ninguna de las proteínas del canal es capaz de suplir o complementar la función de las proteínas DMP. Sin embargo, sorprendentemente en el modelo *Clostridia*, parece ser que se requiere de un sistema más sencillo para completar el engullimiento. La causa principal de estas diferencias podría ser debido a la actividad de hidrolasa de peptidoglicano que ha sido sugerido en la proteína SpoIIQ de *C. difficile* (Serrano *et al.*, 2016; Dembek *et al.*, 2018).

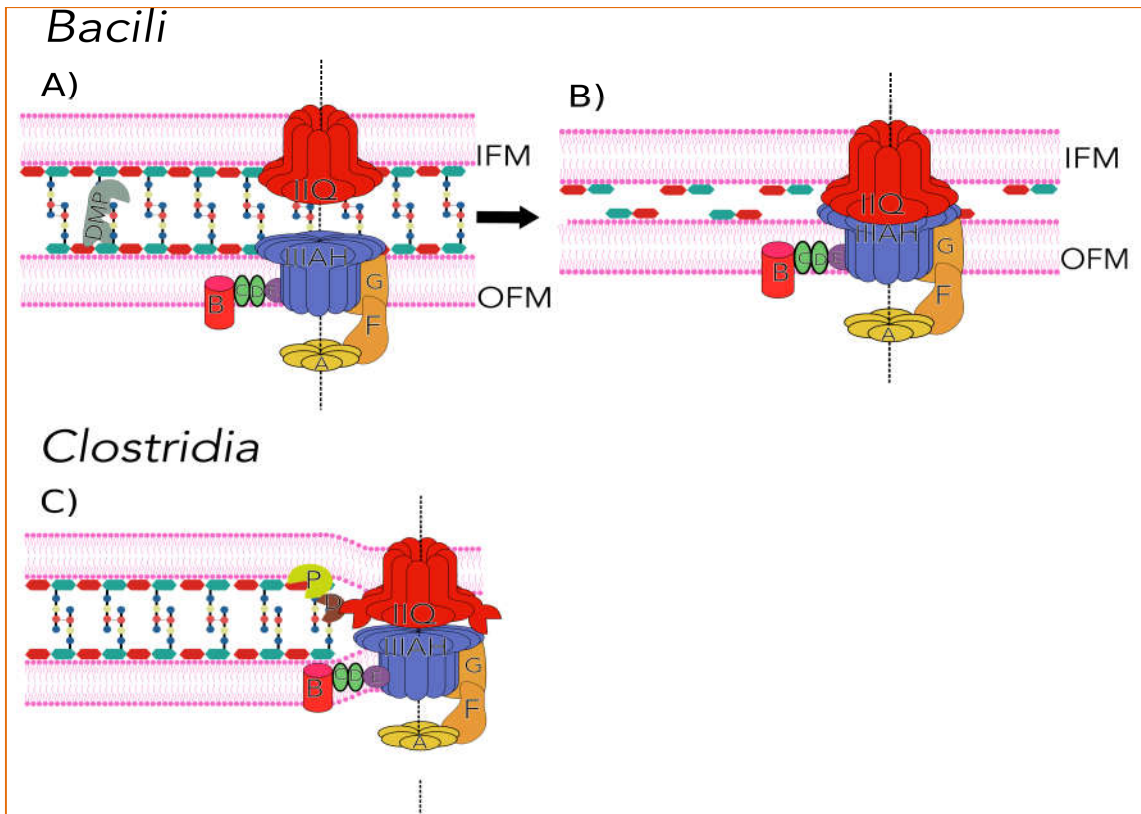


Figura 7. Modelo esquemático de las proteínas requeridas para completar el engullimiento en *B. subtilis* y *C. difficile*.

Modelo *Bacillus subtilis*: Requiere la expresión de dos complejos de proteínas, primero actúa el complejo *SpoIIDMP* que hidroliza el peptidoglicano del septo y posteriormente el canal de esporulación *SpoIIIAH-SpoIIQ* puede ensamblarse permitiendo que *SpoIIQ* y *SpoIIIAH* hagan contacto en el espacio intermembrana.

Modelo *Clostridium difficile*: No se requiere la expresión de los dos complejos, ya que *SpoIIQ* tiene actividad de hidrolasa de peptidoglicano. IFM (Membrana Interna de la prespora), OFM (Membrana Externa de la prespora). Adaptada de Crawshaw *et al.*, 2014.

Todas las proteínas que forman los complejos antes mencionados con excepción SpoIIB, SpoIIQ y GerM, han sido reconocidas como una firma de esporulación. Las 50 proteínas que conforman la firma, han sido utilizadas para la predicción del fenotipo de esporulación y para encontrar nuevos genes implicados en la esporulación (Abecasis *et al.*, 2013). Sin embargo con el reciente aumento en el número de genomas secuenciados, en particular de los Firmicutes, resulta interesante extender el análisis antes mencionado.

Identificación de proteínas homologas al operon SpoIIIAA-AH en Firmicutes.

Se realizó una búsqueda de las proteínas de estudio de *B. subtilis* mediante el servidor Gene Context Tool (GC), que nos permite visualizar el contexto genómico de un gen o un grupo de genes y sus ortólogos en organismos secuenciados, también incorpora información sobre los COG's asociados lo cual es de gran utilidad para la selección de ortólogos verdaderos. Seleccionamos el grupo de secuencias homologas a las proteínas modelo obtenidas mediante el GC. A continuación se muestra un ejemplo gráfico del contexto genómico de las proteínas de estudio. (Figura 8). En A (Arriba) se muestra el contexto del operon SpoIIIAA-AH y en B (abajo) el contexto de la proteína SpoIIQ.

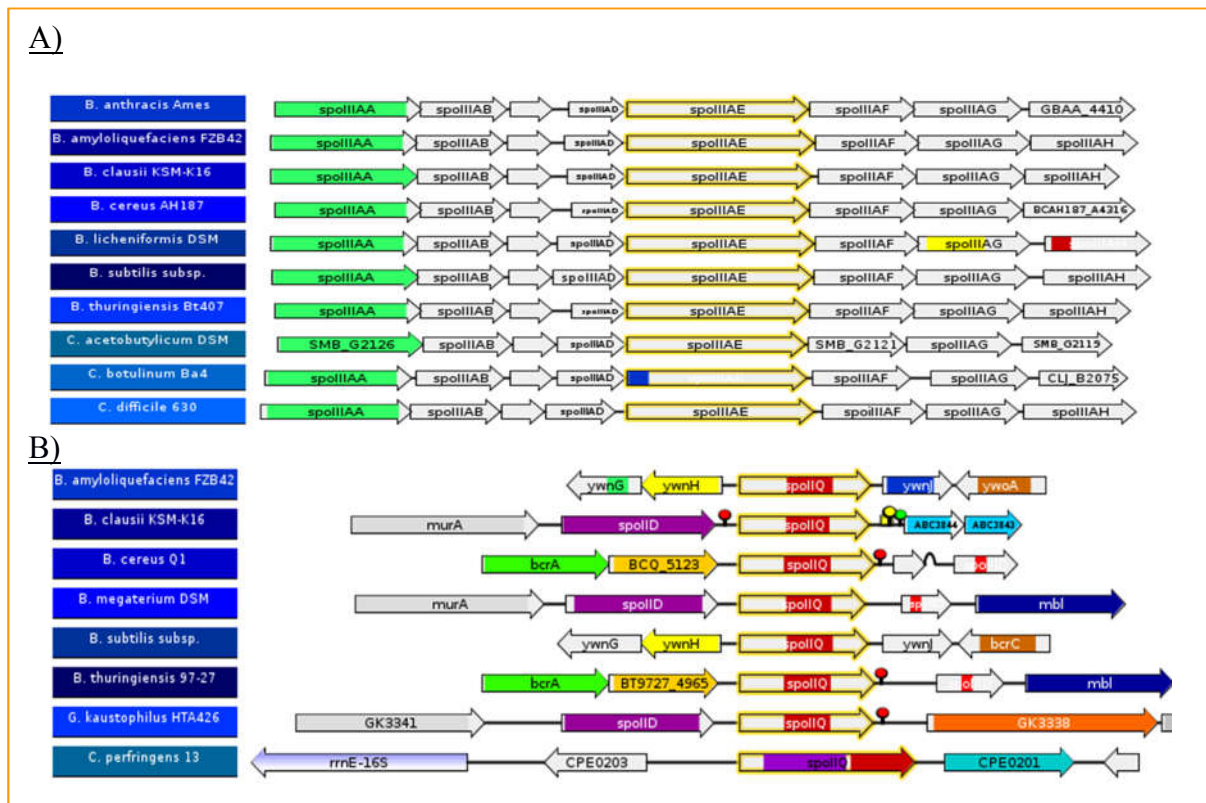


Figura 8. Contexto genómico de las proteínas que forman en canal híbrido de esporulación.

- A) Proteínas del operon SpoIIIAA-AH; Se encuentran altamente conservadas tanto en cepas de *Bacillus* como en *Clostridias*.
- B) Proteína SpoIIQ: Presenta un contexto genómico menos conservado.

Como se puede observar los genes del operon *spoIIIA* se encuentran sumamente conservados, y parecen ser muy específicos de Firmicutes, mientras que la proteína SpoIIQ de *B. subtilis* (bsu-BSU36550) y sus secuencias ortólogas mostraron un arreglo cromosomal muy variable entre especies distintas (Figura 8B), lo que podría indicar que

está implicado en procesos más generales o que tiene una función más específica en una especie y diferente función en otra especie. En las bacterias muchos genes son codificados en operones, como los genes que conforman los diferentes sistemas de secreción. Esta característica ha facilitado su identificación en distintos genomas (Huynen *et al.*, 2000).

La proteína SpoIIQ no se codifica en operón, pero su función la realiza uniéndose con la proteína SpoIIAH para formar un canal que comunica el compartimento de la célula madre con la prespora. Este canal es requerido para nutrir el desarrollo de la espora, así como para la correcta activación del factor *sigma G* (Kehl-Fie *et al.*, 2009). Es importante mencionar que el gen *spoIIQ* ha sido descrito como un reemplazo genético no ortólogo (Galperin *et al.*, 2012), es decir que probablemente haya sido adquirido por transferencia horizontal, sin embargo esta observación no está bien definida. El objetivo de este estudio es analizar mediante genómica comparativa la distribución y evolución de los componentes necesarios para completar el engullimiento en los organismos formadores de esporas.

Para generar un modelo que represente la arquitectura de cada proteína, se identificaron los dominios que las componen, usando las matrices Pfam (base PfamA 3.1 actualizada hasta 2017). De manera general se siguieron los siguientes pasos para la identificación de dominio en las proteínas: se obtuvo el conjunto de secuencias homologas, se identificaron todos los dominios que pueden estar contenidos en cada proteína, y se seleccionó el dominio con el mejor valor esperado *e-value* y la mayor cobertura. Los dominios seleccionados se analizaron para finalmente crear las arquitecturas, que definimos como arreglos de dominios encontrados en la proteína.

La arquitectura que describe a cada proteína del operón *spoIIA*, del complejo DMP, de SpoIIB y GerM se muestra en la Figura 9.

Conocer la arquitectura de cada proteína modelo ayuda a definir las características de cada una, como cuáles son los dominios que la componen, si se encuentra en la región amino o carboxilo terminal, si el dominio abarca toda la proteína, etc.

Se encontró que las proteínas SpoIIAA se encuentran representadas por tres familias de ATPasas, estas son la familia NTPase_1 (perfil Pfam PF03266), AAA (perfil Pfam PF0004) (ATPasas Associated with diverse cellular Activities), y T2SSE (perfil Pfam PF00437). El resto de las proteínas presentan un solo dominio con excepción de SpoIIAD y GerM que están compuestas por dos dominios continuos SpoIIAC (perfil Pfam PF06686), y dos dominios Germane (perfil Pfam PF10646) respectivamente.

Para SpoIIAE, se encontró un subconjunto de proteínas que presentan un dominio adicional SHP, sin embargo este dominio no se tomó en cuenta debido a que el dominio

SHP se encuentra siempre contenido dentro del dominio Spore_III_AE, además de presentar un *e-value* poco significativo (mayor a 10^{-3}).

Proteínas Modelo

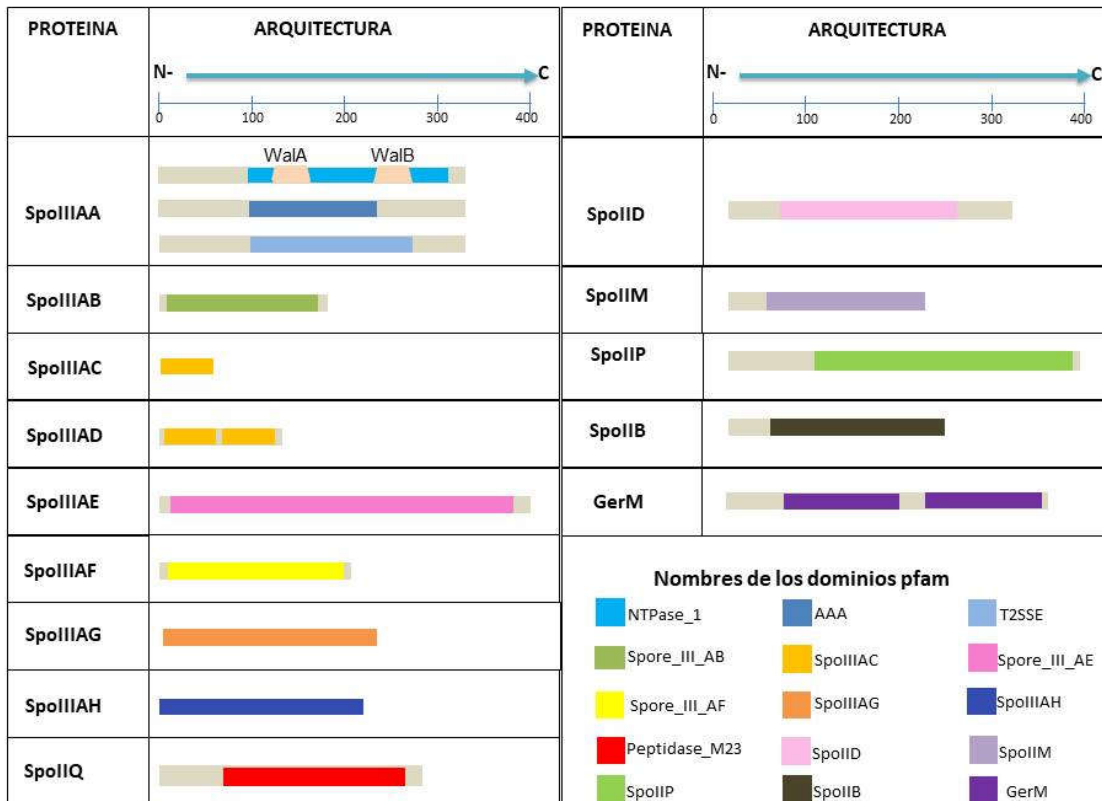


Figura 9. Representación esquemática de las diferentes arquitecturas observadas por proteína usando la base de datos PfamA 3.1.

Los dominios que conforman la arquitectura de cada proteína modelo, son representados mediante las matrices de dominio. La base de datos PfamA, tiene una colección de 16,712 matrices. Las matrices depositadas en Pfam son construidas con un alineamiento semilla usando modelos ocultos de Markov y han demostrado ser eficaces para la identificación de dominios de proteínas, además de ser especialmente útiles para el análisis de proteínas con bajo porcentaje de identidad. Estas matrices resultaron ser adecuadas para identificar los dominios homólogos de todas las proteínas de estudio, con excepción de la proteína SpoIIIIAG para la cual la base *Pfam* tiene un modelo de dominio que identifica por igual a las proteínas SpoIIIIAF y SpoIIIIAG. Para lograr separar a los homólogos de SpoIIIIAG de los de SpoIIIIAF, construimos una matriz específica tal como se mencionó en los métodos. Las 14 matrices de *Pfam* junto con la matriz construida para SpoIIIIAG (Tabla 2, columna

1) fueron usadas para buscar los dominios homólogos en los 4852 genomas actualmente secuenciados almacenados en la base de datos KEGG 2019. Los resultados de esta búsqueda mostraron dominios homólogos aún en otras proteínas que probablemente tengan otras arquitecturas (Tabla 2, columna 7), por ejemplo proteínas además del dominio homólogo presenten otro dominio adicional. También se obtuvieron proteínas codificadas por genes con una organización genética distinta, es decir, que el gen se localiza en una región distinta del cromosoma y por lo tanto presenta diferentes genes vecinos. Para identificar a los presumibles ortólogos (**PO**) de cada proteína usamos como criterio de selección tanto a la arquitectura de la proteína como su contexto genómico.

Observamos la presencia de varios homólogos en cada genoma por lo que para mejorar la predicción de PO, seleccionamos las proteínas que tuvieran un KO asignado (Tabla 2 columnas 5 y 6).

Proteína	Dominios Pfam	Arquitectura	Organización Genómica en Bacilli (operon)	KO que representa la proteína modelo	KOs* vecinos.	Proteínas con el dominio	PO
SpoIIIAA	NTPase_1 (PF03266)	NTPase_1 NTPase_1 x 2	spoIIIAA-AH	K06390	US K02356 K01262 DS K06391- K06397	22723	217
SpoIIIAA	AAA (PF00004)	AAA	spoIIIAA-AH	K06390	US K02356 K01262 DS K06391- K06397	146685	106
SpoIIIAA	T2SSE (PF00437)	T2SSE	spoIIIAA-AH	K06390	US K02356 K01262 DS K06391- K06397	33957	57
SpoIIAB	Spore_III_AB (PF09548)	Spore_III_AB	spoIIIAA-AH	K06391	US K06391 DS	4544	391

					K06392- K06397		
SpoIIIAC	SpoIIIAC (PF06686)	SpoIIIAC	spoIIIAC-AH	K06392	US K06390 K06391 DS K06393- K06397	1832	415
SpoIIIAD	SpoIIIAD (PF06686)	SpoIIIAC x 2	spoIIIAC-AH	K06393	US K06391 K06392 DS K06394- K06397	1832	397
SpoIIIAE	Spore_III_AE (PF09546)	Spore_III_AE	spoIIIAC-AH	K06394	US K06392 K06393 DS K06395- K06397	610	409
SpoIII AF	Spore_III_AF (PF09581)	Spore_III_AF	spoIIIAC-AH	K06395	US K06393 K06394 DS K06396- K06397	1658	368
SpoIIIAG	SpoIIIAG (*build)	SpoIIIAG	spoIIIAC-AH	K06396	US K06394 K06395 DS K06397- K0639	4448	403
SpoIII AH	SpoIII AH (PF12685)	SpoIII AH	spoIIIAC-AH	K06397	US K06395 K06396 DS K01961 K02160 K03625	4540	408
SpoIIQ	Peptidase_M23 (PF01551)	Peptidase_M23	not in operon	K06386	US K06283 K03569	39818	374

					DS K06381 K02392		
SpoIID	SpoIID (PF08486)	SpoIID	not in operon	K06381	US K00790 K21028 DS K06283, K03569.	1853	362
SpoIIM	SpoIIM (PF01944)	SpoIIM	not in operon	K06384	US K03315 K00027 DS K04763, K01839.	2628	363
SpoIIP	SpoIIP (PF07454)	SpoIIP	not in operon	K06385	US K02340 K06012 DS K03596 K00974	1076	341
GerM	Germane (PF10646)	Germane x2	gerM-rph- ysnA-ysnB	K06298	US K01776 K01994 DS K07095 K02428	12620	267
SpoIIB	SPOR (PF05036)	SPOR	spoIIB	K06380	US K11754 K01873 DS K06287 K03630	2085	218

Tabla 2. Resumen de las características usadas para seleccionar presumibles ortólogos (PO) de las proteínas involucradas en el engullimiento y el compromiso para esporular.

Se encontró que los dominios homólogos la proteína SpoIIIAA son los más ampliamente distribuidos, encontrándose en casi todos los genomas. Esto se debe a que la célula necesita energía para realizar una gran cantidad de funciones, desde abrir la doble hélice para comenzar la replicación, transportar sustancias dentro y fuera de la célula, etc. Dentro de

Firmicutes encontramos 1880 homólogos de SpoIIIAA, lo que indica la presencia de varias copias parálogas. SpoIIIAA es clasificada como una ATPasa putativa funcional y forma parte de la superfamilia de ATPasas.

La organización genómica del operon *spoIIIA* así como el análisis de los KOs vecinos fueron características útiles para localizar ortólogos verdaderos (Tabla 2, columna 7). Recuperamos 380 proteínas SpoIIIAA agrupadas en el KO K06390. La secuencia de una proteína puede tener similitud con varias matrices, como el caso de SpoIIIAA, donde la mayoría de las veces se identificaron a las tres familias de ATPase. Cuando se tienen proteínas con similitud a diversas familias de proteínas es difícil clasificar con certeza a dicha proteína, pero si es posible seleccionar la matriz que represente la familia más probable, seleccionando la matriz con el menor valor de *e-value*. Otro criterio de selección de dominio, cuando la proteína tiene más de un dominio por proteína es tomar el que tenga una mayor cobertura. De esta selección se obtuvieron 218 proteínas con el dominio NTPase_1, siendo esta familia la mayor representada, seguida por la familia AAA con 106 proteínas y finalmente la T2SSE que mostró una abundancia menor con 57 proteínas.

Con la finalidad de conocer que diferencias existen entre las proteínas que presentan una arquitectura definida con el dominio NTPase_1, con las que tienen el dominio AAA o el T2SSE, se realizaron alineamientos múltiples por dominio con cada grupo. El dominio NTPase_1 a su vez mostró diferencias en las regiones y/o motivos que lo caracterizan, como se muestra en la figura 10.

Definimos 4 arquitecturas, la primera solo contiene el motivo Walker A (14 proteínas), la segunda solo contiene el Walker B (19 proteínas), la tercera presenta ambos motivos pero se obtiene un par de coordenadas por motivo (52 proteínas), el motivo Walker A se localiza frecuentemente entre los aminoácidos 140-160 y el Walker B entre el 240 y 260. La última arquitectura es la más frecuente, contiene ambos motivos y se detecta con un par de coordenadas (133 proteínas).

Aparentemente existen diferencias entre los dominios NTPase_1, AAA y T2SSE, sin embargo al inspeccionar las secuencias encontramos que en realidad los tres dominios presentan ambos, el motivo Walker A y el Walker B, y que los tres dominios definen casi las mismas regiones de la proteína SpoIIIAA. Adicionalmente encontramos las cajas de Histidina y Aspártico muy conservadas, mismas que ha sido reportado se encuentran conservadas en otras familias de proteínas como en las familias PilB y PilT (Jakovljevic *et al.*, 2008).

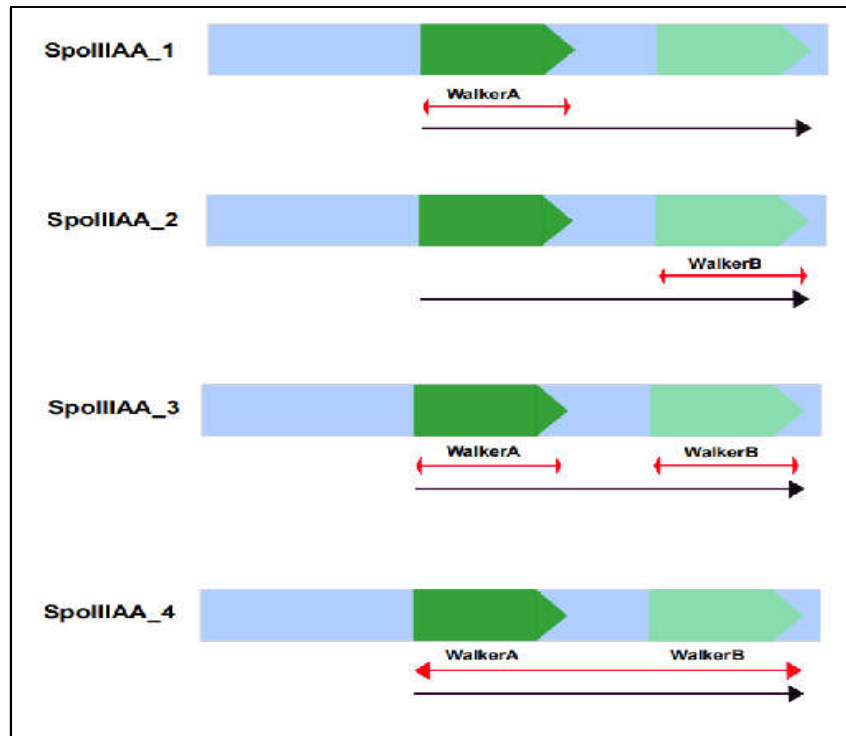


Figura 10. Representación que ejemplifica las arquitecturas propuestas para SpoIIIAA. SpoIIIAA_1 contiene el motivo Walker A y se encuentra usando las matrices de los dominios NTPase_1 y AAA. SpoIIIAA_2 contiene el motivo Walker B se encuentra en los dominios NTPase_1 y AAA. SpoIIIAA_3 contiene un dominio NTPase_1 duplicado que cubre ambos motivos y SpoIIIAA_4 que cubre ambos motivos y se encuentra usando las matrices de los tres dominios. Los motivos Walker se muestran en color verde.

Posteriormente nos preguntamos si existen diferencias en las proteínas SpoIIIAA de las diversas clases de Firmicutes. Se realizó un análisis clasificando primero los dominios en la clase *Bacilli* y la clase *Clostridia*, se alinearon las secuencias como se puede observar en la Figura 11.

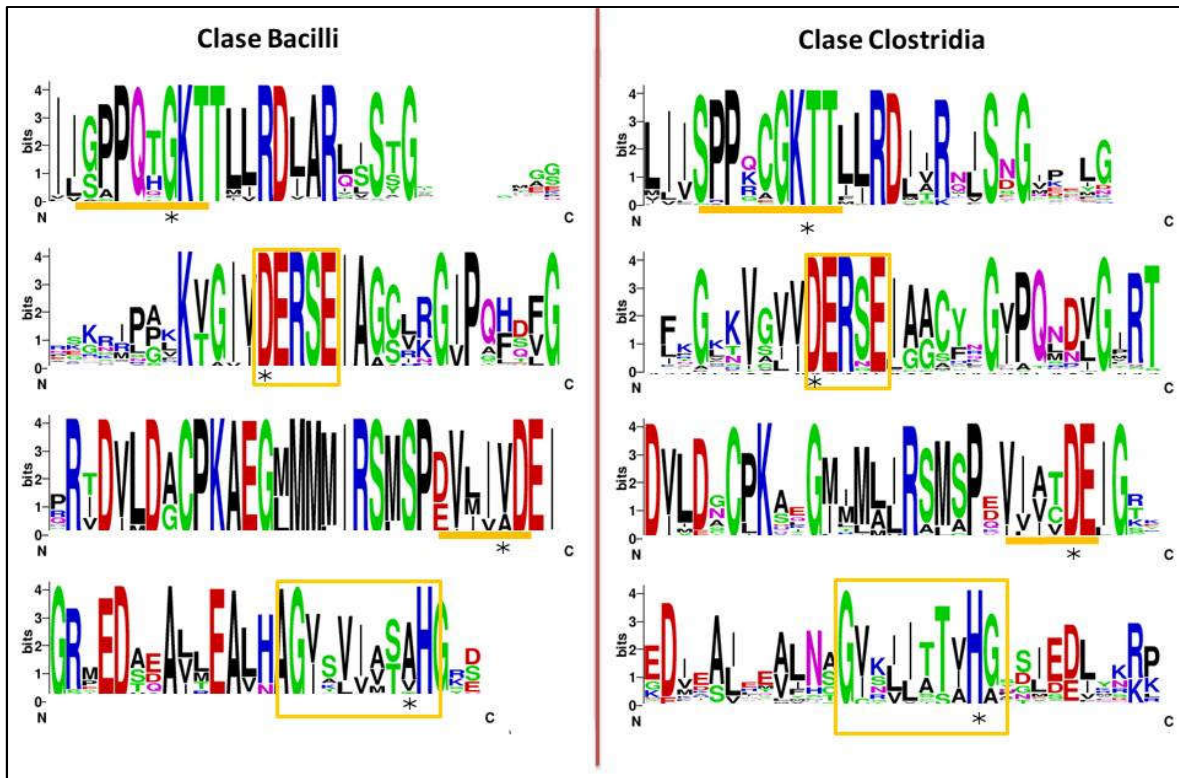


Figura 11. Especificidad de los motivos Walker A y Walker B (mostrados con barra amarilla). También se muestran las cajas de H y D (recuadros amarillos) encontrados en las clase Bacilli y Clostridia que corresponden a proteínas SpoIIIAA.

Encontramos diferencias entre el motivo Walker A de la clase *Bacilli* y la *Clostridia*, y lo más importante es que el motivo ha cambiado respecto a la secuencia consenso reportada que es G-x(4)-GK-[TS]. En el caso de *Clostridia*, en la posición de la primer Glycina hay una Serina, es probable que el cambio haya sido de G a S, ya que las NTPases conservan la Gly y análisis filogenéticos indican que el último ancestro común universal (LUCA) de todas las formas de vida celular ya codificaba múltiples y diversas NTPasas (Leipe *et al.*, 2002). Adicionalmente se creó un perfil de secuencia que representara el prototipo ancestral del motivo Walker A, el cuál presenta una Gly (Romero *et al.*, 2018) y se demostró que una corta secuencia de proteína con el motivo Walker A es funcional.

Se sabe que la proteína SpoIIIAA de *B. subtilis* y de *Clostridias* difiere en cuanto a la función de los dominios Walker A y Walker B. Mientras que en *Bacillus* se requiere la presencia de los dos para su función, en *Clostridias* solo el motivo Walker A es esencial (Fimlaid *et al.*, 2015). Los motivos Walker A y Walker B, junto con las cajas de His y Asp son una firma que caracteriza a las proteínas SpoIIIAA, y a pesar de que se ha reportado que el motivo Walker B no es esencial en *C. difficile* (Fimlaid *et al.*, 2015), éste motivo parece estar bien conservado en el grupo.

Distribución del operon *spoIIIA*, y las proteínas SpoIIQ, SpoIIB, GerM y SpoIIDMP en Firmicutes.

En un trabajo previo se reportó la extensión filogenética de proteínas de la maquinaria de endoesporulación, incluyendo las proteínas involucradas en el engullimiento, en un conjunto de 72 organismos, de los cuales 31 especies son endoesporulantes. Los resultados mostraron que tanto las proteínas del operon *spoIIIA* como las proteínas del complejo DMP se encuentran conservadas en los organismos formadores de esporas, siendo estas proteínas catalogadas como una firma de esporulación (Abecasis *et al.*, 2013). Para corroborar si estas observaciones permanecen, investigamos la distribución de PO de las proteínas involucradas en el engullimiento utilizando un conjunto de 953 Firmicutes secuenciados y las arquitecturas mostradas en la Tabla 1.

Estas arquitecturas fueron suficientes para identificar PO en las seis clases de Firmicutes y de manera sorprendente, en una proteobacteria, como se muestra en la Figura 12.

Las proteínas SpoIIIAA presentan la arquitectura del dominio NTPase_1, AAA o T2SSE. Proteínas con el dominio NTPase_1 son distribuidas en las clases Bacilli, Clostridia y Negativicutes. La arquitectura conformada por el dominio AAA se distribuyó entre la clase Bacilli y Clostridia cubriendo 28% de las proteínas. Por su parte, la arquitectura T2SSE identificada (57 proteínas) 15%, se encontró en Bacilli, Clostridia y Limnochordia. Notablemente, ninguno de estos dominios fue capaz de identificar SpoIIIAA en las clases Erysipelotrichia y Tissierellia, lo que sugiere que si estas bacterias esporulan probablemente tienen una arquitectura alternativa de la ATPasa.

Las arquitecturas que describen el PO desde SpoIIAB a SpoIIAH no se distribuyeron homogéneamente entre los Firmicutes, lo que resulta sorprendente considerando la organización en operon descrita experimentalmente en *B. subtilis*, *C. difficile* y los resultados bioinformáticos mostrados por Abecasis (Abecasis *et al.*, 2013).

La proteína SpoIIAB esta anotada en la base de datos Pfam con el perfil identificado como PF09548 Spore_III_AB. La arquitectura de SpoIIAB se encontró en 391 Firmicutes, se desconoce la función de esta proteína en el complejo, pero estudios previos sugieren que esta proteína sirve como un enlace estructural entre componentes proteicos unidos a membrana y componentes solubles del canal Q:AH (Zeytuni *et al.*, 2018).

SpoIIAB es una proteína con similitud a la familia GspF y TadB/C del sistema tipo II. La superfamilia de proteínas GspF, está compuesta de proteínas de membrana interna, altamente hidrófobas, involucradas en una vía de secreción general (general secretion pathway; GSP) para la exportación de proteínas.

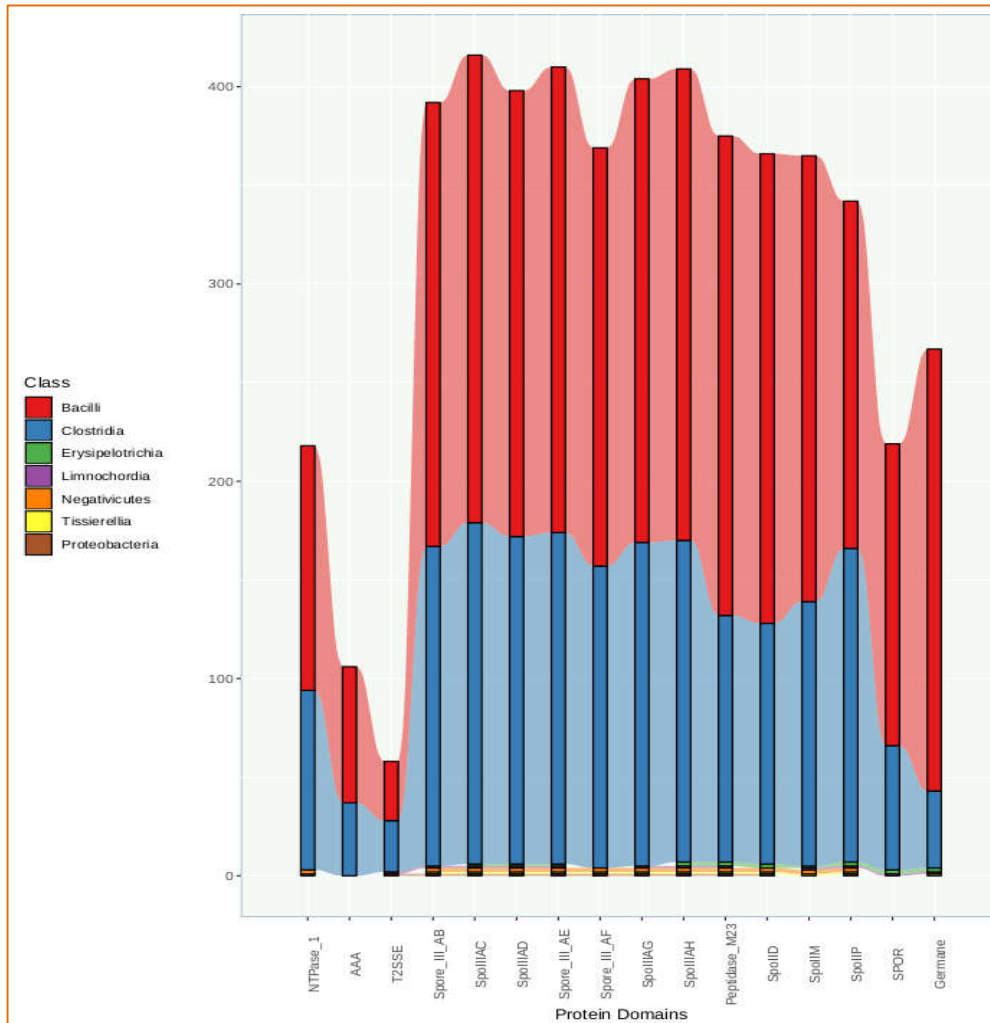


Figura 12. Distribución de dominios homólogos a las proteínas del engulfosoma en los proteomas analizados. La figura presenta la distribución de arquitecturas de proteínas que forman el tubo de alimentación Q:AH, el complejo DMP y las proteínas SpoIIB y GeM. Se detectaron en las clases Bacilli, Clostridia, Negativicutes, Erysipelotrichia, Tissierella, Limnochordia y una Proteobacteria en la que encontramos varias arquitecturas del tubo de alimentación y una proteína del complejo DMP.

SpoIIIC fue la proteína para la cual se identificó un número más significativo de PO. Se ha sugerido que SpoIIIC y SpoIIIC pueden funcionar como un aparato de exportación de solo dos componentes, que pueden desempeñar un papel central en el reconocimiento y secreción del sustrato (Zeytuni *et al.*, 2019). Encontramos que SpoIIIC y SpoIIIC están distribuidos heterogéneamente (Figura 12). Comparando los organismos que presentan proteínas SpoIIIC y SpoIIIC, se puede ver que 237 Bacilli tienen SpoIIIC en contraste con las 226 observaciones de SpoIIIC. Obtuvimos resultados similares en

Clostridia con 173 observaciones de SpoIIIAC versus 166 en SpoIIAD. Erysipelotrichia, Limnochordia, Negativicutes y Tissierellia tienen PO de ambas proteínas. La arquitectura que describe SpoIIAF fue menos identificada, en contraste con SpoIIAE, SpoIIAG y SpoIIAH. Las arquitecturas que describen SpoIIAC, SpoIIAD, SpoIIAE y SpoIIAH fueron apropiadas para identificar al menos una proteína en cada una de las clases evaluadas. Trabajos previos proponen a la proteína SpoIIQ como parte del canal en *B. subtilis* y *C. difficile*. En estos trabajos, los autores realizan mutantes en genes del operon *spoIIA*, *spoIIQ* de *B. subtilis* que producen graves defectos celulares, que incluyen la deslocalización de las proteínas SpoIID y SpoIIP involucradas en el engullimiento (Aung *et al.*, 2007), y alteración en la regulación de los factores sigma de esporulación SigK, SigF y SigG que también participan en la regulación del proceso (Doan *et al.*, 2005, Flanagan *et al.*, 2016). Para buscar SpoIIQ en los 953 Firmicutes utilizamos el perfil Pfam Peptidase_M23 (PF01551) y los KO's vecinos K06283, K03569, K06381, K02392. Sorprendentemente se logró identificar 243 PO de SpoIIQ en Bacilli, sin embargo, tuvimos 239 proteínas SpoIIAH. En Clostridia, detectamos 163 PO de SpoIIAH y 125 PO de SpoIIQ. Los proteomas de Erysipelotrichia, Limnochordia, Negativicutes y Tissierellia tienen una relación uno a uno de las proteínas SpoIIAH y SpoIIQ. Adicionalmente, las diferencias en la función de SpoIIQ entre *B. subtilis* y *C. difficile* pueden ser explicadas tanto por los cambios específicos en Bacillus, ya que se sabe ha perdido su actividad endopeptidasa porque carece de los residuos importantes para la catálisis, y porque al ser una proteína multifuncional, puede compartir con las Clostridia su papel estructural en la formación de canal y específicamente en el grupo Bacilli haber adquirido otras funciones. Varios estudios han demostrado que las proteínas SpoIIQ y SpoIIAH son esenciales para el engullimiento en *C. difficile* (Serrano *et al.*, 2016, Fimlaid *et al.*, 2015). Mas recientemente se caracterizó la maquinaria de degradación en *C. difficile* (Dembek *et al.*, 2018, Ribis *et al.*, 2018), mismos que apoyan la hipótesis de que SpoIIDP forma parte de un mismo complejo con las proteínas del canal de esporulación en *C. difficile* y que estos se han separado en *B. subtilis* y organismos relacionados. En la Figura 12, mostramos que los perfiles utilizados para identificar DMP, como en los casos anteriores, presentan diferentes proporciones de las proteínas del complejo, excepto en los proteomas de Negativicutes y Tissierellia.

Los perfiles Pfam Germane (PF10646) y SPOR (PF05036), se utilizaron para encontrar a las proteínas GerM y SpoIIB, respectivamente. GerM se reportó como un factor requerido para la localización de SpoIIQ, formando junto con SpoIIAH una plataforma tripartita para el ensamblaje del complejo Q:AH y una plataforma base para la localización de proteínas de la célula madre y la prespora (Rodrigues *et al.*, 2016). Nuestros resultados

mostraron diferencias significativas en el número de proteínas entre Q:AH y GerM. GerM resultó estar significativamente menos distribuida que Q:AH, en Clostridia. Observamos resultados similares en el caso de SpoIIB, que fue ampliamente extendido en Bacilli pero no en Clostridia. Las proporciones de las proteínas exhibidas en la Figura 12 muestran claramente una distribución heterogénea de las proteínas que forman los complejos en los organismos analizados. Las razones que explican este comportamiento podrían ser: a) la presencia de bacterias no esporulantes en el conjunto, que presentan una o más de estas proteínas; b) los modelos de referencia propuestos para *B. subtilis* y *C. difficile* no son suficientes para describir los complejos en algunos Firmicutes esporulantes; c) anotaciones incompletas en genomas secuenciados. El siguiente paso fue evaluar si las bacterias que carecen de estas proteínas son están reportadas como esporulantes en la literatura con la finalidad de probar las hipótesis mencionadas anteriormente.

Conservación de las proteínas que forman el tubo de alimentación entre los formadores de endosporas.

El análisis realizado por Abecasis y colaboradores mostró que las proteínas de los complejos Q:AH y DMP deben considerarse como marcadores de esporulación, ya que se conservan en los 31 formadores de endosporas analizados (Abecasis et al., 2013). En la sección anterior, buscamos cada proteína en 953 Firmicutes para determinar su distribución. Si estas proteínas son marcadores de esporulación, las identificaremos en los organismos formadores de endosporas. Para probar esta hipótesis, buscamos en la literatura para hacer un compendio de formadores de endosporas. Utilizamos dos fuentes de información usando bases de datos que ofrecen fenotipos curados como BacMap, Patric y GOLD (Cruz et al., 2012, Wattam et al., 2017, Mukherjee et al., 2018) e inspección bibliográfica del organismo faltante en las bases de datos. Además, buscamos información reciente para corroborar que los fenotipos ofrecidos por las bases de datos estuvieran actualizados y descubrimos que la mayoría de las entradas fueron actualizadas recientemente.

Los 953 Firmicutes, se clasificaron en 5 niveles diferentes de evidencia como se muestra en la Tabla 3. Los fenotipos Esporulante y no esporulante corresponden a especies para las cuales se han realizado rigurosas caracterizaciones experimentales. Otras dos clases muestran especies para las cuales se describe el genoma secuenciado, y el fenotipo esporulante o no esporulante se infirió de las morfologías típicas descritas para la Familia o Género filogenético como en el caso de la familia Streptococcaceae que claramente es un

grupo no esporulante (manual Bergeys, Vos *et al.*, 2011). Finalmente, el grupo desconocido presenta organismos para los cuales no pudimos asignar un fenotipo ya que la Familia o Género posee bacterias esporulantes y no esporulantes. De acuerdo con un estudio reciente realizado por Silva y colaboradores, determinar la presencia de los genes que constituyen la firma de esporulación puede ayudar a predecir el fenotipo de estos organismos (Silva *et al.*, 2019), por lo que convendría ampliar y extrapolar el estudio de Abecasis a nuestro conjunto de datos, que incluye organismos que hasta el momento no han sido estudiados.

Mostramos en la Tabla 3 el número de especies clasificadas en cada grupo.

Clase	No esporulante	Esporulante	Desconocido	No esporulante deducido/desconocido	Esporulante deducido/desconocido
Bacilli	245	187	52	214	28
Clostridia	56	114	19	2	4
Erysipelotrichia	6	0	3	0	0
Negativicutes	9	2	3	3	0
Limnochordia	0	1	0	0	0
Tissierellia	2	1	2	0	0
Total	318	305	79	219	32

Tabla 3. Clasificación de los Firmicutes. Resumen del proceso de curación.

Como se muestra en la Tabla 3, las especies formadoras de endosporas verificadas experimentalmente incluyen 187 bacterias que pertenecen a la clase Bacilli, 114 a Clostridia, dos especies que pertenecen a Negativicutes y una especie que pertenece a Limnochordia y Tissierellia. De estos, solo 42 especies carecen de una a seis de las ocho proteínas del complejo SpoIIIAA-AH.

En *Clostridium tetani* E88 se reportó la ausencia de SpoIIIAC y SpoIIIAD (Abecasis *et al.* 2013). Al inspeccionar la región intergénica entre SpoIIIAB y SpoIIAE, notamos la presencia de una región de 616 pb, en donde hipotetizamos que estas proteínas pequeñas pueden alojarse. Para confirmar esta hipótesis, realizamos una búsqueda BlastX (Altschul *et al.*, 1990), utilizando 1000 pb, que incluyó 384 pb de la región de codificación SpoIIIAB y la región intergénica entre SpoIIIAB y SpoIIAE. La comparación dio coincidencia con las proteínas SpoIIIAC (16-209) y SpoIIIAD (301-610). Luego incluimos las secuencias traducidas en nuestro conjunto de datos para realizar un "hhmscan" utilizando los perfiles SpoIIIAC y D,

que dieron resultados positivos para ambas proteínas. Teniendo en cuenta esta observación, realizamos una búsqueda de las proteínas faltantes del operón *spoIIIA*, usando la secuencia de las regiones intergénicas en los genomas curados de los Firmicutes formadores de esporas. Los resultados son resumidos en la Tabla 4.

Código de organismo	Especies	Clase	Fenotipo	Proteína ausente	Método de anotación
aad	<i>Alicyclobacillus acidocaldarius</i> subsp. <i>acidocaldarius</i> Tc-4-1	Bacilli	Esporulante	SpoIIIAG	BlastX
baco	<i>Bacillus</i> sp. <i>OxB-1</i>	Bacilli	Esporulante	SpoIIIAA	(---)
baco	<i>Bacillus</i> sp. <i>OxB-1</i>	Bacilli	Esporulante	SpoIIIAB	(---)
baco	<i>Bacillus</i> sp. <i>OxB-1</i>	Bacilli	Esporulante	SpoIIIAC	(---)
baco	<i>Bacillus</i> sp. <i>OxB-1</i>	Bacilli	Esporulante	SpoIIIAD	(---)
baco	<i>Bacillus</i> sp. <i>OxB-1</i>	Bacilli	Esporulante	SpoIIIAF	(---)
baco	<i>Bacillus</i> sp. <i>OxB-1</i>	Bacilli	Esporulante	SpoIIIAG	BlastX
bcef	<i>Bacillus cereus</i> FT9	Bacilli	Esporulante	SpoIIIAB	BlastX
bmyc	<i>Bacillus pseudomycooides</i> 219298 (<i>Bacillus mycooides</i> 219298)	Bacilli	Esporulante	SpoIIIAG	BlastX
btb	<i>Bacillus thuringiensis</i> BMB171	Bacilli	Esporulante	SpoIIIAC	BlastX
btc	<i>Bacillus thuringiensis</i> serovar <i>chinensis</i> CT-43	Bacilli	Esporulante	SpoIII AH	BlastX
bts	<i>Kyrpidia tusciae</i> DSM 2912	Bacilli	Esporulante	SpoIII AH	BlastX
cad	<i>Gottschalkia acidurici</i> 9a (<i>Clostridium acidurici</i> 9a)	Tissierellia	Esporulante	SpoIII AA	PF05621 *
cbol	<i>Clostridium bolteae</i> ATCC BAA-613	Bacilli	Esporulante	SpoIII AA	PF13191 *
cbol	<i>Clostridium bolteae</i> ATCC BAA-613	Bacilli	Esporulante	SpoIII AF	BlastX
chy	<i>Carboxydotherrmus hydrogenoformans</i> Z-2901	Clostridia	Esporulante	SpoIII AA	BlastX

chy	<i>Carboxydotherrmus hydrogenoformans Z-2901</i>	Clostridia	Esporulante	SpoIIIAH	BlastX
cso	<i>Clostridium saccharolyticum K10</i> cf.	Bacilli	Esporulante	SpoIIIAA	BlastX
cso	<i>Clostridium saccharolyticum K10</i> cf.	Bacilli	Esporulante	SpoIIIBAB	BlastX
cso	<i>Clostridium saccharolyticum K10</i> cf.	Bacilli	Esporulante	SpoIIIIDAD	BlastX
csr	<i>Clostridium saccharoperbutylacetonicum NI-4(HMT)</i>	Clostridia	Esporulante	SpoIIIAH	(---)
ctc	<i>Clostridium tetani E88</i>	Clostridia	Esporulante	SpoIIIIAC	BlastX
ctc	<i>Clostridium tetani E88</i>	Clostridia	Esporulante	SpoIIIIDAD	BlastX
ddl	<i>Desulfotobacterium dichloroeliminans LMG P-21439</i>	Clostridia	Esporulante	SpoIIIAA	BlastX
dhd	<i>Desulfotobacterium hafniense DCB-2</i>	Clostridia	Esporulante	SpoIIIAA	PF13401 *
dmi	<i>Desulfosporosinus meridiei DSM 13257</i>	Clostridia	Esporulante	SpoIIIAA	PF13481 *
dsy	<i>Desulfotobacterium hafniense Y51</i>	Clostridia	Esporulante	SpoIIIAA	PF13401 *
gel	<i>Geobacillus sp. LC300</i>	Bacilli	Esporulante	SpoIIIAE	BlastX
hmo	<i>Heliobacterium modesticaldum Ice1</i>	Clostridia	Esporulante	SpoIIIAA	BlastX
ibu	<i>Intestinimonas butyriciproducens AF211</i>	Clostridia	Esporulante	SpoIIIAA	BlastX
ibu	<i>Intestinimonas butyriciproducens AF211</i>	Clostridia	Esporulante	SpoIIIAF	BlastX
jeo	<i>Jeotgalibacillus malaysiensis D5</i>	Bacilli	Esporulante	SpoIIIBAB	(---)
jeo	<i>Jeotgalibacillus malaysiensis D5</i>	Bacilli	Esporulante	SpoIIIIAC	(---)
jeo	<i>Jeotgalibacillus malaysiensis D5</i>	Bacilli	Esporulante	SpoIIIIDAD	(---)
jeo	<i>Jeotgalibacillus malaysiensis D6</i>	Bacilli	Esporulante	SpoIIIAE	(---)
jeo	<i>Jeotgalibacillus malaysiensis D7</i>	Bacilli	Esporulante	SpoIIIAF	(---)

lgy	<i>Lysinibacillus varians</i> (<i>Lysinibacillus sp. GY32</i>)	Bacilli	Esporulante	SpoIIIAA	(---)
lgy	<i>Lysinibacillus varians</i> (<i>Lysinibacillus sp. GY32</i>)	Bacilli	Esporulante	SpoIIIAB	(---)
lgy	<i>Lysinibacillus varians</i> (<i>Lysinibacillus sp. GY32</i>)	Bacilli	Esporulante	SpoIIIID	(---)
lgy	<i>Lysinibacillus varians</i> (<i>Lysinibacillus sp. GY32</i>)	Bacilli	Esporulante	SpoIIIAF	(---)
lpil	<i>Limnochorda pilosa HC45</i>	Limnochordia	Esporulante	SpoIIIAF	BlastX
lsp	<i>Lysinibacillus sphaericus C3-41</i>	Bacilli	Esporulante	SpoIIIAA	(---)
lsp	<i>Lysinibacillus sphaericus C3-41</i>	Bacilli	Esporulante	SpoIIIAB	BlastX
lsp	<i>Lysinibacillus sphaericus C3-41</i>	Bacilli	Esporulante	SpoIIIID	BlastX
lsp	<i>Lysinibacillus sphaericus C3-41</i>	Bacilli	Esporulante	SpoIIIAF	(---)
lyb	<i>Lysinibacillus sp. B2A1</i>	Bacilli	Esporulante	SpoIIIAA	(---)
lyb	<i>Lysinibacillus sp. B2A1</i>	Bacilli	Esporulante	SpoIIIAB	(---)
lyb	<i>Lysinibacillus sp. B2A1</i>	Bacilli	Esporulante	SpoIIIID	(---)
lyb	<i>Lysinibacillus sp. B2A1</i>	Bacilli	Esporulante	SpoIIIAF	(---)
lys	<i>Lysinibacillus sp. YS11</i>	Bacilli	Esporulante	SpoIIIAA	(---)
lys	<i>Lysinibacillus sp. YS11</i>	Bacilli	Esporulante	SpoIIIAB	(---)
lys	<i>Lysinibacillus sp. YS11</i>	Bacilli	Esporulante	SpoIIIID	(---)
lys	<i>Lysinibacillus sp. YS11</i>	Bacilli	Esporulante	SpoIIIAF	(---)
paea	<i>Paenibacillus sp. FSL R7-0273</i>	Bacilli	Esporulante	SpoIIIAF	BlastX
pae	<i>Paenibacillus sp. FSL R7-0331</i>	Bacilli	Esporulante	SpoIIIAF	BlastX
paef	<i>Paenibacillus sp. FSL R5-0345</i>	Bacilli	Esporulante	SpoIIIAF	BlastX
paeh	<i>Paenibacillus sp. FSL H7-0357</i>	Bacilli	Esporulante	SpoIIIAF	BlastX

paej	<i>Paenibacillus sp. FSL H7-0737</i>	Bacilli	Esporulante	SpoIIIAF	BlastX
paek	<i>Paenisporosarcina</i>	Bacilli	Esporulante	SpoIIIAA	(---)
paek	<i>Paenisporosarcina</i>	Bacilli	Esporulante	SpoIIIAB	(---)
paek	<i>Paenisporosarcina</i>	Bacilli	Esporulante	SpoIIIID	(---)
paek	<i>Paenisporosarcina</i>	Bacilli	Esporulante	SpoIIIAF	(---)
paen	<i>Paenibacillus sp. FSL P4-0081</i>	Bacilli	Esporulante	SpoIIIAF	BlastX
paeq	<i>Paenibacillus sp. FSL R5-0912</i>	Bacilli	Esporulante	SpoIIIAF	BlastX
pbd	<i>Paenibacillus borealis DSM 13188</i>	Bacilli	Esporulante	SpoIIIAF	BlastX
pbj	<i>Paenibacillus beijingensis DSM 24997</i>	Bacilli	Esporulante	SpoIIIAA	BlastX
pbj	<i>Paenibacillus beijingensis DSM 24997</i>	Bacilli	Esporulante	SpoIIIAE	(---)
pbj	<i>Paenibacillus beijingensis DSM 24997</i>	Bacilli	Esporulante	SpoIIIAF	BlastX
pdu	<i>Paenibacillus durus DSM 1735</i>	Bacilli	Esporulante	SpoIIIAE	BlastX
pdu	<i>Paenibacillus durus DSM 1735</i>	Bacilli	Esporulante	SpoIIIAF	BlastX
pgm	<i>Paenibacillus graminis DSM 15220</i>	Bacilli	Esporulante	SpoIIIAE	BlastX
pgm	<i>Paenibacillus graminis DSM 15220</i>	Bacilli	Esporulante	SpoIIIAF	BlastX
pgm	<i>Paenibacillus graminis DSM 15220</i>	Bacilli	Esporulante	SpoIIIAH	BlastX
ppq	<i>Paenibacillus polymyxa SQR-21</i>	Bacilli	Esporulante	SpoIIIIAC	(---)
pste	<i>Paenibacillus stellifer DSM 14472</i>	Bacilli	Esporulante	SpoIIIAH	BlastX
pste	<i>Paenibacillus stellifer DSM 14472</i>	Bacilli	Esporulante	SpoIIIAE	BlastX
pste	<i>Paenibacillus stellifer DSM 14472</i>	Bacilli	Esporulante	SpoIIIAF	BlastX
pswu	<i>Paenibacillus swuensis DY6</i>	Bacilli	Esporulante	SpoIIIAF	BlastX

rbp	<i>Ruminococcaceae</i> bacterium <i>CPB6</i>	Clostridia	Esporulante	SpoIII AF	BlastX
rst	<i>Rummeliibacillus</i> <i>stabekisii</i> <i>PP9</i>	Bacilli	Esporulante	SpoIII AA	(---)
rst	<i>Rummeliibacillus</i> <i>stabekisii</i> <i>PP9</i>	Bacilli	Esporulante	SpoIII AB	(---)
rst	<i>Rummeliibacillus</i> <i>stabekisii</i> <i>PP9</i>	Bacilli	Esporulante	SpoIII AC	(---)
rst	<i>Rummeliibacillus</i> <i>stabekisii</i> <i>PP9</i>	Bacilli	Esporulante	SpoIII AD	(---)
rst	<i>Rummeliibacillus</i> <i>stabekisii</i> <i>PP9</i>	Bacilli	Esporulante	SpoIII AF	(---)
rst	<i>Rummeliibacillus</i> <i>stabekisii</i> <i>PP9</i>	Bacilli	Esporulante	SpoIII AG	BlastX
sgy	<i>Syntrophobotulus</i> <i>glycolicus</i> <i>DSM 8271</i>	Clostridia	Esporulante	SpoIII AA	BlastX
siv	<i>Solibacillus silvestris</i> <i>StLB046</i>	Bacilli	Esporulante	SpoIII AA	(---)
siv	<i>Solibacillus silvestris</i> <i>StLB047</i>	Bacilli	Esporulante	SpoIII AB	(---)
siv	<i>Solibacillus silvestris</i> <i>StLB048</i>	Bacilli	Esporulante	SpoIII AC	(---)
siv	<i>Solibacillus silvestris</i> <i>StLB049</i>	Bacilli	Esporulante	SpoIII AD	(---)
siv	<i>Solibacillus silvestris</i> <i>StLB049</i>	Bacilli	Esporulante	SpoIII AF	(---)
siv	<i>Solibacillus silvestris</i> <i>StLB050</i>	Bacilli	Esporulante	SpoIII AG	BlastX
spsy	<i>Sporosarcina</i> <i>psychrophila</i> <i>DSM 6497</i>	Bacilli	Esporulante	SpoIII AA	(---)
spsy	<i>Sporosarcina</i> <i>psychrophila</i> <i>DSM 6498</i>	Bacilli	Esporulante	SpoIII AB	(---)
spsy	<i>Sporosarcina</i> <i>psychrophila</i> <i>DSM 6499</i>	Bacilli	Esporulante	SpoIII AC	(---)
spsy	<i>Sporosarcina</i> <i>psychrophila</i> <i>DSM 6500</i>	Bacilli	Esporulante	SpoIII AD	(---)
spsy	<i>Sporosarcina</i> <i>psychrophila</i> <i>DSM 6499</i>	Bacilli	Esporulante	SpoIII AF	(---)
spsy	<i>Sporosarcina</i> <i>psychrophila</i> <i>DSM 6500</i>	Bacilli	Esporulante	SpoIII AG	BlastX
ssil	<i>Solibacillus silvestris</i> <i>DSM</i> <i>12223</i>	Bacilli	Esporulante	SpoIII AA	(---)

ssil	<i>Solibacillus silvestris</i> DSM 12223	Bacilli	Esporulante	SpoIIIAB	(---)
ssil	<i>Solibacillus silvestris</i> DSM 12223	Bacilli	Esporulante	SpoIIIAC	(---)
ssil	<i>Solibacillus silvestris</i> DSM 12223	Bacilli	Esporulante	SpoIIIAD	(---)
ssil	<i>Solibacillus silvestris</i> DSM 12223	Bacilli	Esporulante	SpoIIIAF	(---)
ssil	<i>Solibacillus silvestris</i> DSM 12223	Bacilli	Esporulante	SpoIIIAG	BlastX
sure	<i>Sporosarcina ureae</i> P32a	Bacilli	Esporulante	SpoIIIAA	(---)
sure	<i>Sporosarcina ureae</i> P32a	Bacilli	Esporulante	SpoIIIAF	(---)
sure	<i>Sporosarcina ureae</i> P32a	Bacilli	Esporulante	SpoIIIAG	BlastX
tmr	<i>Thermaerobacter marianensis</i> DSM 12885	Clostridia	Esporulante	SpoIIIAE	(---)
tmr	<i>Thermaerobacter marianensis</i> DSM 12885	Clostridia	sporulating	SpoIIIAF	BlastX
tmr	<i>Thermaerobacter marianensis</i> DSM 12885	Clostridia	sporulating	SpoIII AH	(---)

Tabla 4. Proteínas del operon SpoIII AA reanotadas.

Nuestros resultados incluyen 50 anotaciones nuevas en 49 genomas que muestran de 1 a 6 proteínas no detectadas. En los 49 genomas evaluados, las proteínas faltantes eran principalmente anotaciones incorrectas. Este análisis también mostró que en 5 especies, los dominios Pfam; PF05621, PF13401, PF13191 y PF13481 describen la ATPasa que representa a los dominios SpoIII AA. Los perfiles de estos dominios no fueron seleccionadas porque no cumplían con los criterios de selección propuestos, este fue el caso de *Gottschalkia acidurici* 9a (*Clostridium acidurici* 9a), *Clostridium bolteae* ATCC BAA-613, *Desulfitobacterium hafniense* DCB-2, *Desulfosporosinus meridiei* DSM 13257 y *Desulfitobacterium hafniense* Y51.

También encontramos un grupo de siete genomas de *Heliobacterium modesticaldum* Ice1, *Carboxydotherrmus hydrogeoformans* Z-2901, *Clostridium cf. saccharolyticum* K10, *Intestinimonas butyriciproducens* AF211, *Desulfitobacterium dichloroeliminans* LMG P-21439, *Syntrophobotulus glycolicus* DSM 8271 y *Paenibacillus beijingensis* DSM 24997, para los cuales la búsqueda de BlastX, dio coincidencia significativa con las proteínas homólogas del estado III de esporulación. De todo el conjunto de formadores de

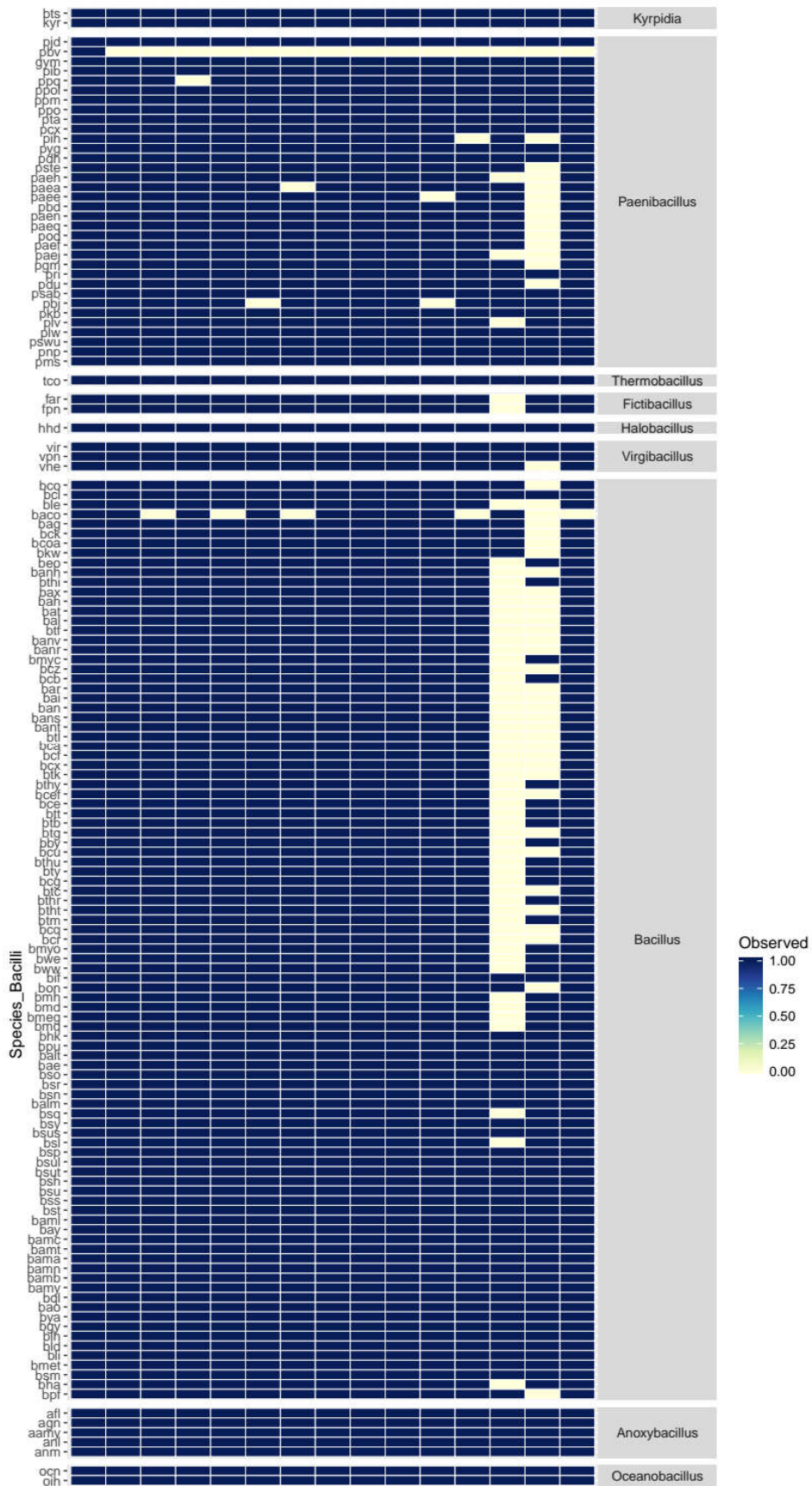
endosporas, el proteoma de *Lysinibacillus sphaericus* C3-41 carece de cuatro proteínas del operón (SpoIIIAA, SpoIIIBAB, SpoIIIBAD y SpoIIIBAF), realizamos una búsqueda de BlastX tomando 4000 pb río arriba del ATG de spoIIIBAG. Como resultado, solo encontramos la proteína SpoIIIBAC. Aplicamos el mismo procedimiento a las otras cuatro especies de *Lysinibacillus*; *Lysinibacillus varians* (*Lysinibacillus* sp. GY32), *Limnochorda pilosa* HC45, *Lysinibacillus* sp. B2A1 y *Lysinibacillus varians* (*Lysinibacillus* sp. GY32) (ver Tabla 4), para todos los casos, no obtuvimos resultados que reconocieran las proteínas faltantes. Una característica importante que enmarca las predicciones de SpoIIIAH que carecían de anotación y que refuerza la presencia del gen en el genoma es la presencia de una acetil-CoA carboxilasa (K01961) como vecino de spoIIIAH, una característica también compartida por 249 genomas poseedores de SpoIIIAH.

En la Tabla 4 mostramos 5 especies; *Bacillus thuringiensis* serovar *chinensis* CT-43, *Kyrpidia tusciae* DSM 2912, *Carboxydotherrmus rogenoformans* Z-2901, *Paenibacillus graminis* DSM 15220 y *Paenibacillus stellifer* DSM 14472, en las cuales se anotó una proteína SpoIIIAH.

Análisis recientes revelaron que SpoIIQ, SpoIIIAA y SpoIIIAH regulan múltiples etapas de la formación de esporas en *C. difficile*, incluyendo el engullimiento, la localización apropiada del abrigo y el mantenimiento de la espora (Fimlaid *et al.*, 2015). En dicho estudio, se observó un engullimiento que terminaba con una estructura similar a un abrigo y que parecía anclarse al borde de la membrana pero que no se adhería a la interfaz de célula madre y la prespora en la mayoría de las células mutantes (Fimlaid *et al.*, 2015). El análisis realizado en este trabajo mostró que falta el PO de SpoIIQ en 20 Clostridias reportadas como formadoras de esporas (Figura 13b); algunas de las especies que carecen de SpoIIQ son *Candidatus Desulforudis audaxviator* MP104C, *Desulfallas gibsonia* DSM 7213 y *Sulfobacillus acidophilus* TPY. De las 20 especies que carecen de SpoIIQ todas presentan al operón *spoIIIAA-AH* completo, Galperin y sus colaboradores sugieren que la ausencia de SpoIIQ en algunos Clostridia es un desplazamiento genético no ortólogo (Galperin *et al.*, 2012). Estos resultados sugieren que el modelo descrito para *Clostridium difficile* podría ser diferente en otros Clostridia, como ha sido sugerido utilizando herramientas experimentales y bioinformáticas (Paredes *et al.*, 2005; de Hoon *et al.*, 2010; Dembek *et al.*, 2018). Se debe hacer un análisis más profundo de las especies para corroborar la hipótesis mencionada anteriormente.

El engullimiento necesita a los complejos DMP y Q:AH en *Bacillus subtilis* (Crawshaw *et al.*, 2014). Recientemente se estudió el papel del complejo DMP y su interacción con el complejo Q:AH en *Clostridium difficile*, encontrando que SpoIIM no es necesaria para el engullimiento ya que produce una esporulación eficiente (Dembek *et al.*, 2018).

Paenibacillus bovis, *Acetoanaerobium sticklandii* DSM 519 y *Lachnoanaerobaculum umeaense* DSM 23576 = CCUG 58757 fueron las únicas especies en las que no pudimos detectar ninguna de las proteínas del complejo DMP, a pesar de que se reportan con un fenotipo esporulante. En el caso de *Paenibacillus bovis* reportado como formador de endosporas (Liu *et al.*, 2016), Spo0A está presente en el genoma, en contraste, no encontramos Spo0A en *Acetoanaerobium sticklandii* DSM 519 y *Lachnoanaerobaculum umeaense* DSM 23576 = CCUG 58757 (Hedberg *et al.*, 2012). *Acetoanaerobium sticklandii* DSM 519 también conocido como *Clostridium sticklandii*, es una especie donde se ha reportado que la formación de esporas es rara y se ve únicamente en cultivos viejos (Stadtman *et al.*, 1957), lo que indica que es posible que la bacteria no sea esporulante, no podemos explicar nuestros resultados y la relación con el fenotipo observado. Nuestros resultados sugieren la ausencia de una o dos PO del complejo SpoIIDMP en 90 formadores de endosporas **Figura 13a-b**. Los experimentos realizados en *Clostridium difficile* demostraron que mutantes en SpoIIM pueden proceder con el engullimiento, pero si requiere de las proteínas SpoIIP-SpoIID (Dembek *et al.*, 2018).



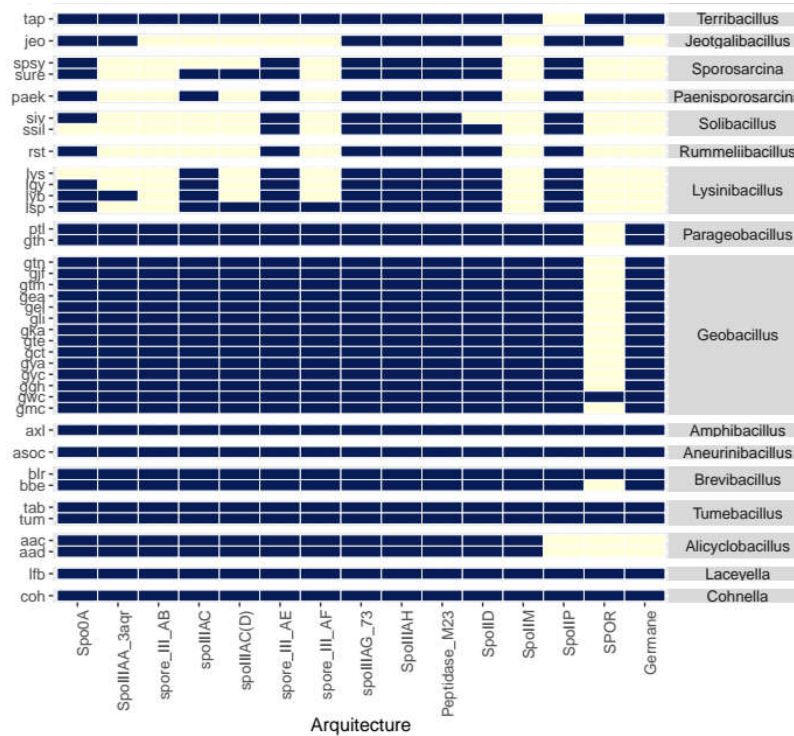


Figura 13. Distribución por género de proteínas del engullimiento en Bacilli y Clostridia endoesporulantes. a) El mapa de calor presenta la distribución de proteínas involucradas en la estructuración del tubo de alimentación, el complejo SpoIIDMP, SpoIIB y GerM entre diferentes géneros de la clase Bacilli.

Los resultados encontrados en *Intestinimonas butyriciproducens AF211*, sugieren que en esta especie el complejo Q:AH podría ser suficiente para completar la esporulación. Vale la pena mencionar que la descripción morfológica de la cepa reveló que a pesar de presentar esporas, la pared celular no es típica de Gram-positivos, y las células se tiñen de Gram-negativas (Kläring *et al.*, 2013), una característica que puede estar afectando la disposición y organización de los complejos involucrados en la formación de esporas.

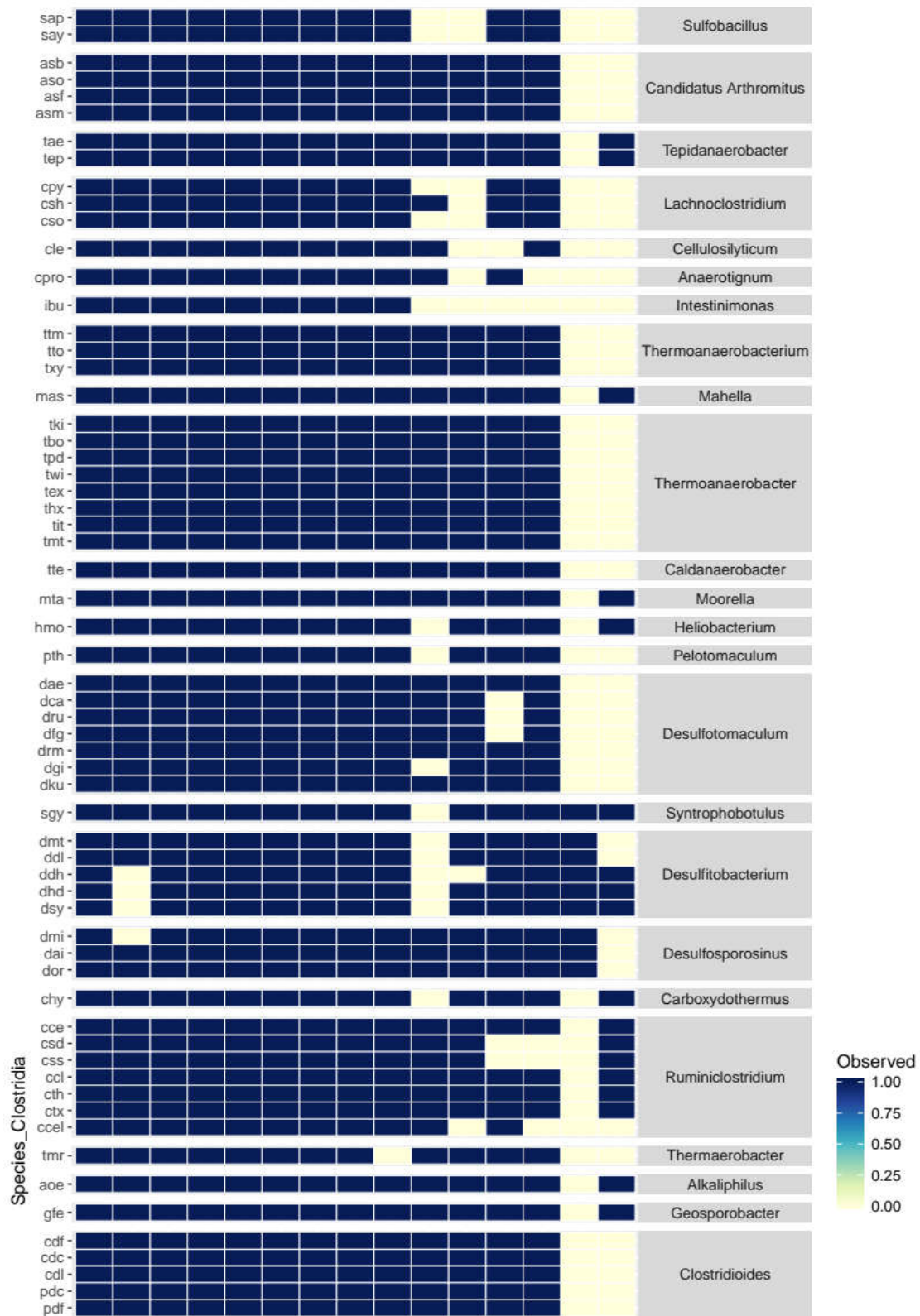
La proteína GerM, una lipoproteína previamente implicada en la germinación de esporas, se requiere para ensamblar la plataforma basal del complejo Q:AH durante la esporulación en *Bacillus subtilis*. Ya que GerM y SpoIIAH derivan de la célula madre, y SpoIIQ de la espora, estas proteínas tienen dependencias de localización recíproca (Rodrigues *et al.*, 2016). En las especies de bacilos la proteína GerM está ausente en quince organismos. Todas estas especies, excepto *Alicyclobacillus acidocaldarius subsp. acidocaldarius DSM 446* y *Alicyclobacillus acidocaldarius subsp. acidocaldarius Tc-4-1* presentan SpoIIP. Los trece Bacilli restantes no presentan ni gerM, ni SpoIID o SpoIIM (Figura 13a). Los resultados en la Figura 13b muestran que 92 Clostridia esporulantes y dos Negativicutes, *Pelosinus sp. UFO1* y *Pelosinus fermentans JBW45*, no tienen PO GerM. Este hallazgo es

importante ya que GerM reduce la actividad del factor sigma G en *Bacillus subtilis* (Rodrigues *et al.*, 2016), el cual a su vez modifica la expresión de genes involucrados en la activación de sigma K, que induce la expresión de proteínas de germinación (Fimlaid *et al.*, 2013). En contraste con los hallazgos experimentales en *Bacillus subtilis*, el papel de Sigma G como modulador de Sigma K en las especies de Clostridia, incluido *Clostridium difficile*, queda por dilucidar. Otras funciones asociadas con GerM están relacionadas con la acumulación de ácido dipicolínico en la espora y la hidrólisis de la corteza al germinar. Alternativamente, GerM podría influir indirectamente en la germinación promoviendo el ensamblaje adecuado del abrigo (McKenney *et al.*, 2012), por lo que la ausencia de esta proteína en tantas especies de Clostridia muestra que una gran cantidad de mecanismos están controlando la esporulación en diferentes especies.

SpoIIB muestra débil homología con el dominio de especificidad de sustrato de CwlM, una amidasa de *Bacillus licheniformis* (Margolis *et al.*, 1993), lo que sugiere que interactúa con la pared celular. Estudios recientes mostraron que la mutante nula de *spoIIB* en *B. subtilis* muestra un fenotipo de engullimiento transitorio no descrito anteriormente y similar al de otras mutantes del engullimiento que no degradan el peptidoglicano del septo. La cepa mutante *spoIIB* puede completar el engullimiento, lo que sugiere que el defecto afecta solo la velocidad del proceso (Perez *et al.*, 2000). Los resultados obtenidos muestran que esta proteína probablemente no es esencial para el proceso de esporulación en una gran cantidad de Firmicutes esporulantes. Encontramos que 78 Bacilli no tienen un PO de SpoIIB. Esta proteína resultó faltar en 70 Clostridias esporulantes, dos Negativicutes (*Pelosinus sp. UFO1* y *Pelosinus fermentans JBW45*) (ambos carecen de GerM), y un Tissierellia; *Gottschalkia acidurici 9a* (*Clostridium acidurici 9a*).

Los resultados mostrados en las Figuras 13a-b indican que la distribución de los componentes en las especies que están más estrechamente relacionadas se conservan con frecuencia. No obstante, encontramos algunos ejemplos donde las especies no conservan la misma distribución a pesar de la distancia filogenética como en el genero de Lysinibacillus discutido anteriormente. La clasificación de las especies en el género, como se muestra en la Figura 13a-b, reveló que todas las especies agrupadas en Alicyclobacillus, Anoxybacillus, Fictibacillus, Parageobacillus y Tumebacillus de la clase Bacilli conservaron la misma disposición de proteínas. La misma clasificación realizada dentro de Clostridia expuso el mismo comportamiento para las especies agrupadas en los géneros: Candidatus Arthromitus, Clostridioides, Tepidanaerobacter, Thermoanaerobacter y Thermoanaerobacterium. Vale la pena señalar, que a diferencia de los bacilos en los que las especies agrupadas tienen solo dos miembros, las especies agrupadas dentro de Clostridia tienen una mayor cantidad de miembros. Es notable ver que los géneros con una

mayor cantidad de miembros no conservan una distribución homogénea de las proteínas ni las proteínas del tubo de alimentación Q:AH ni las proteínas DMP.



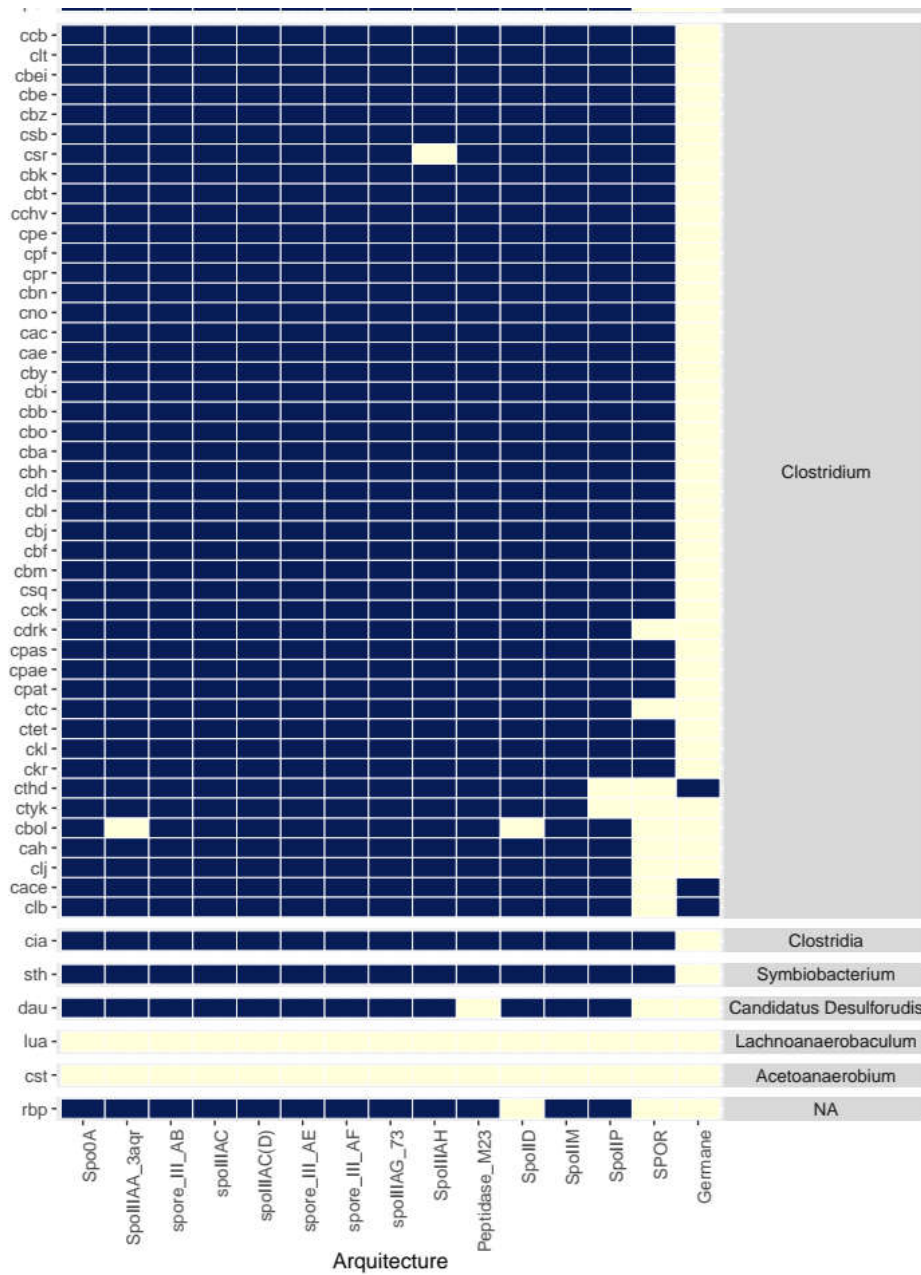


Figura 13. Distribución por género de proteínas del engullimiento en Bacilli y Clostridia endoesporulantes. b) Distribución de las proteínas involucradas en la estructuración del los complejos Q:AH y DMP, y las proteínas SpoIIB y GerM entre diferentes géneros de la clase Clostridia.

Nuestra curación manual también mostró un grupo de especies para las cuales no encontramos evidencia convincente que muestre un fenotipo esporulante, pero se propone que todas se comportan como formadores de endosporas, debido a que estas especies se distribuyen siguiendo patrones similares a los de las especies confirmadas como

formadoras de esporas (Datos no mostrados). Como en los resultados anteriores, encontramos que cuando más de dos especies se agrupan en un género, aumenta la probabilidad de que falte una proteína. El género *Virgibacillus* y *Selenomonas* fueron las únicas excepciones encontradas en este conjunto. Teniendo en cuenta la disponibilidad de los datos de secuencia de los genes 16S ribosomal y que los genomas completos aumentan, creemos que para algunas especies, en particular, las de la clase Clostridia que pueden reorganizarse en un nuevo género, familia, orden, clase o incluso filo según lo propuesto por Kunisawa (Kunisawa *et al.*, 2015). Las otras dos posibilidades se basan en la construcción de modelos bioinformáticos para obtener otras arquitecturas que permitan encontrar las proteínas faltantes o esperar experimentos que describan mecanismos en especies no probadas diferentes *B. subtilis* y *C. difficile*. Con base en los resultados presentados, el siguiente paso natural fue organizar las proteínas identificadas en un árbol filogenético y observar si las especies con mayor cercanía conservaban la misma organización.

Distribución de los componentes del engullimiento en Firmicutes representativos.

Organizamos los resultados de PO con los datos de esporulación curados manualmente en un árbol de especies, usando secuencias de ARNr 16S. Probamos la metodología descrita para construir el árbol reproduciendo el árbol hecho por Davidson (Davidson *et al.*, 2018). El árbol presentado en la Figura 14, muestra la distribución filogenética y de arquitectura de 147 Firmicutes representativos, elegidos utilizando los siguientes criterios: a) un representante de un grupo construido con el algoritmo CD-HIT; cuando el grupo posee especies con diferente organización de proteínas, también se seleccionó; cuando la proteína tiene la misma organización de proteínas, pero diferente fenotipo reportado (como formador de endosporas o no endosporas), también se seleccionó. Incluimos en el árbol las predicciones calculadas para Spo0A para 259 Firmicutes (Amador *et al.*, 2019), y fueron complementadas con el K07699 asignado en KEGG a Spo0A. Este último criterio se aplicó para asegurar lo más posible una condición de esporulación. El árbol muestra 110 especies que representan formadores de endosporas, 31 no formadores de endosporas y 6 organismos para los cuales no se ha demostrado un fenotipo de esporulación (Figura 14).



Figura 14. Árbol filogenético basado en las secuencias del gen 16S rRNA y proteínas del engullimiento identificadas en 149 proteomas representativos. Seleccionamos secuencias representativas del gen 16S rRNA del conjunto completo de 953 Firmicutes como se describe en los métodos, para la construcción de un árbol de especies. El árbol se complementa (panel derecho), con arquitecturas que describen catorce proteínas involucradas en el engullimiento y el compromiso para esporular.

El árbol muestra que algunas especies de Firmicutes no esporulantes, particularmente las de la clase Clostridia, tienen proteínas del complejo Q:AH, DMP y SpoIIB o GerM. Para algunas especies como, por ejemplo, *Syntrophomonas wolfei subsp. wolfei str. Goettingen* se ha reportado que su genoma codifica para proteínas de varias etapas de esporulación (Sieber *et al.*, 2010), pero no se ha logrado encontrar la condición en la que este Clostridium esporule. De manera similar, la especie *Syntrophothermus lipocalidus DSM 12680* carece de SpoIIIAF y SpoIIAA, si prevalecen los modelos presentados para *B. subtilis* o *C. difficile*, la ausencia de SpoIIQ o SpoIIIAA podría ser la explicación del fenotipo observado.

Thermosediminibacter oceani DSM 16646 y *Natranaerobius thermophilus JW / NM-WN-LF* tienen un conjunto completo de proteínas que forman el tubo de alimentación, así como DMP y GerM. La ausencia de SpoIIB en estas especies como se describe en un párrafo anterior solo podría afectar a la velocidad del engullimiento (Perez *et al.*, 2000), por lo que teóricamente podrían esporular. La inspección en el genoma de otros componentes propuestos por Abecasis arrojará algo de luz para intentar probar una amplia gama de condiciones de esporulación, para especies aisladas de ambientes extremos (Zhao *et al.*, 2011). Otros Clostridios no esporulantes también tienen un conjunto completo de proteínas que forman el tubo de alimentación, pero carecen de las otras proteínas analizadas.

No encontramos el PO de SpoIIIAF y SpoIID en la especie *Limnochorda pilosa HC45* reportada como endosporulante. Todas las especies pertenecientes a la clase Erysipelotrichia como se muestra en la Figura 14 no son esporulantes o tienen un fenotipo desconocido, no detectamos PO de más de una de las proteínas analizadas en este estudio. Las observaciones obtenidas para la clase Negativicutes muestran que las especies esporulantes (Pelosinus) tienen las proteínas que forman el tubo de alimentación y el complejo SpoIIDMP, pero carecen de SpoIIB y GerM. Las especies no esporulantes de esta clase no tienen ninguna de las PO buscadas. Las observaciones realizadas en este trabajo sugieren que el conjunto de proteínas que deben considerarse como firmas de esporulación deben seleccionarse cuidadosamente y tal vez solo estar distribuidas específicamente a nivel taxonómico de género.

Proteínas que forman el canal de esporulación presentan contextos genómicos muy conservados.

Con la finalidad de extender nuestro análisis para buscar proteínas ortólogas fuera de firmicutes, utilizamos el análisis del contexto genómico para determinar que genes vecinos podrían ser utilizados como un “gen marcador vecino”. Se buscaron los ortólogos de cada proteína usando la clasificación de KO’s, y se recuperaron también los datos para los tres genes vecinos corriente arriba y abajo, los resultados se pueden observar en la tabla 5.

SpoIIIAA (KO6390)			SpoIIIAB (KO6391)			SpoIIIAC (KO6392)		
Id KO's	Frec	Distrib	Id KO's	Frec	Distrib	Id KO's	Frec	Distrib
K-----	552	0.25	K-----	424	0.19	K06393	384	0.16
K06392	365	0.16	K06393	376	0.17	K06394	383	0.16
K06393	362	0.16	K06390	376	0.17	K06390	381	0.16
K06391	360	0.16	K06392	375	0.16	K06391	375	0.16
K02356	239	0.11	K06394	370	0.16	K-----	359	0.15
K01262	140	0.06	K02356	210	0.09	K06395	352	0.15
K03786	31	0.01	K01262	25	0.01	K02356	47	0.02
K00928	27	0.01	K00928	14	0.006	K06396	16	0.007
K08097	12	0.01	K08097	11	0.005	K09769	8	0.003
K05979	9	0.004	K18682	8	0.004	K00928	7	0.003
K18682	8	0.004	K09769	8	0.004	K04041	5	0.002
K09769	8	0.004	K12574	6	0.003	K01262	5	0.002
K06394	8	0.004	K04041	5	0.002	K12574	4	0.002
K03565	7	0.003	K09807	4	0.002	K00459	4	0.002
K04041	5	0.002	K06395	4	0.002	K06397	3	0.001
K00891	5	0.002	K06306	3	0.001	K06306	3	0.001
K09807	4	0.002	K03786	3	0.001	K01887	3	0.001
K01887	4	0.002	K09768	2	0.001	K09807	2	0.001
K12574	3	0.001	K08221	2	0.001	K09768	2	0.001
K06395	3	0.001	K07491	2	0.001	K07448	2	0.001
K06306	3	0.001	K07025	2	0.001	K06399	2	0.001
K01271	3	0.001	K06396	2	0.001	K03786	2	0.001
K00014	3	0.001	K06133	2	0.001	K03553	2	0.001
K09768	2	0.001	K02015	2	0.001	K03530	2	0.001
K07491	2	0.001	K01990	2	0.001	K03466	2	0.001
K07025	2	0.001	K01876	2	0.001	K03438	2	0.001
K03091	2	0.001	K01077	2	0.001	K02897	2	0.001
K01990	2	0.001	K00014	2	0.001	K01876	2	0.001
						K01077	2	0.001
						K00014	2	0.001
SpoIIIID (KO6393)			SpoIIIAE (KO6394)			SpoIIIAF (KO6395)		
Id KO's	Frec	Distrib	Id KO's	Frec	Distrib	Id KO's	Frec	Distrib
K06392	387	0.17	K06392	383	0.16	K06393	355	0.16
K06394	379	0.16	K06397	378	0.16	K06396	353	0.16
K06390	378	0.16	K06396	378	0.16	K06394	353	0.16

K06391	376	0.16	K06393	376	0.16	K06392	352	0.16
K06396	374	0.16	K06391	370	0.16	K06397	349	0.16
K06395	354	0.15	K06395	352	0.15	K02160	176	0.08
K-----	55	0.02	K-----	70	0.03	K-----	161	0.07
K06397	14	0.01	K02160	21	0.01	K01571	19	0.01
K02356	5	0.002	K06390	8	0.003	K06391	4	0.002
K01262	3	0.001	K00459	6	0.003	K02837	4	0.002
K03565	2	0.001	K02356	5	0.002	K06390	3	0.001
K03553	2	0.001				K03413	3	0.001
						K01961	3	0.001
						K18682	2	0.001
						K03625	2	0.001
						K03565	2	0.001
						K03553	2	0.001
						K02968	2	0.001
SpoIIAG (KO6396)			SpoIIAH (KO6397)			SpoIIQ (KO6386)		
Id KO's	Frec	Distrib	Id KO's	Frec	Distrib	Id KO's	Frec	Distrib
K06394	379	0.16	K-----	513	0.22	K-----	461	0.33
K06397	378	0.16	K06394	368	0.16	K06283	145	0.10
K06393	374	0.16	K06396	377	0.16	K03569	125	0.09
K06395	353	0.15	K06395	348	0.15	K06381	123	0.09
K-----	241	0.10	K01961	249	0.11	K00790	106	0.08
K02160	220	0.09	K02160	225	0.10	K02392	80	0.06
K01961	220	0.09	K03625	92	0.04	K03823	56	0.04
K03625	83	0.04	K01491	38	0.02	K01990	53	0.04
K01571	24	0.01	K03601	28	0.01	K01992	48	0.03
K06392	16	0.01	K01571	23	0.01	K19302	44	0.03
K02759	4	0.002	K01409	16	0.01	K04751	28	0.02
K03413	3	0.001	K06393	11	0.005	K03588	24	0.02
K02837	3	0.001	K02837	4	0.002	K00681	16	0.01
K06391	2	0.001	K02759	4	0.002	K07240	9	0.01
K06147	2	0.001	K06392	3	0.001	K02372	8	0.006
K06012	2	0.001	K03602	3	0.001	K20742	6	0.004
K03601	2	0.001	K03413	3	0.001	K03320	6	0.004
K02968	2	0.001	K02968	3	0.001	K02033	5	0.004
K01409	2	0.001	K06397	2	0.001	K02028	5	0.004
			K06147	2	0.001	K03088	4	0.003
			K06012	2	0.001	K02034	4	0.003
			K01572	2	0.001	K01358	4	0.003
			K01448	2	0.001	K06131	3	0.002
						K21467	2	0.001
						K19802	2	0.001
						K13954	2	0.001
						K01304	2	0.001
						K00138	2	0.001

Tabla 5. Lista de KO's vecinos seleccionados, usando como referencia el KO central (mostrado entre parentesis). En rojo se muestran los KO's del operon y en azul los conservados en la vecindad.

Estos resultados muestran que cada proteína de estudio puede tener un conjunto de genes vecinos muy amplio, pero parecen ser especie específicos, por lo que el análisis realizado fuera de Firmicutes nos ha permitido constatar que efectivamente estas proteínas son casi exclusivas de Firmicutes y que pocas ortólogas se encuentran fuera de ellos.

Más allá de Firmicutes: *Ruegeria sp. NKC1-1* a Protoebateria que preserva Spo0A y la maquinaria del engullimiento.

La cascada de esporulación en bacterias formadoras de endosporas se ha estudiado ampliamente de forma experimental y bioinformática en Firmicutes. Dado que la mayoría de las proteínas parecen ser altamente específicas para completar las funciones de esporulación, algunas proteínas de las histidina cinasas del fosforelevio y SpoIIIAA que desempeñan un papel en el proceso del engullimiento se han comparado con modelos de proteínas similares fuera de Firmicutes. Nuestros resultados revelan que las proteínas SpoIIP y Spo0A, y las proteínas del tubo de alimentación, excepto SpoIIIAF, están presentes en el genoma de *Ruegeria sp. NKC1-1* (identificador KEGG rue - T05571-). Todas las proteínas tienen el KO correspondiente al definido para las proteínas del engullimiento, así como los Kos de genes vecinos. *Ruegeria sp. NKC1-1* se describe en la base de datos de taxonomía NCBI como en la base de datos KEGG, como una Alphaproteobacteria aislada de kimchi (Ref-Seq GCA_003336365.1). El manuscrito que describe el fenotipo y la distribución filogenética propuso que el aislado puede ser distinguido de las cepas de *Natribacillus halophilus HN30T* y *Geomicrobium halophilus BHIT* más estrechamente asociadas a los Bacilli. Por lo tanto, la cepa NKC1-1 podría considerarse una nueva especie de Gram positiva, aeróbica, en forma de bacilo, no móvil y sin esporas (PMID: 30361979). La observación de PO en el proteoma y la clasificación fenotípica de NKC1-1 como Bacilli, sugieren dos formas posibles de clasificar la cepa: a) NKC1-1 es un Bacilli que sigue el modelo de *Bacillus subtilis*, y no puede esporular porque carece de los genes que codifican SpoIIIAF, SpoIIQ y SpoIIDM; b) Es una Alphaproteobacteria que incorpora algunos genes esporulantes de otros Firmicutes presentes en el kimchi. Para probar cualquiera de las hipótesis, se requiere más evidencia experimental.

XII. CONCLUSIONES.

1. El estudio realizado muestra que los nuevos avances en bioinformática y la disponibilidad de bases de datos nuevas y más amplias son herramientas útiles para inferir los componentes de la compleja maquinaria de esporulación en Firmicutes.
2. La estrategia para el análisis de las proteínas implicadas en el compromiso para esporular y el engullimiento utilizando perfiles de proteínas, así como el contexto genómico, es muy apropiada para recuperar PO con mayor precisión.
3. Los métodos empleados revelaron proteínas aparentemente ausentes, lo que sugiere la existencia de variantes de proteínas o mecanismos alternativos aún no dilucidados que describen el proceso del engullimiento.
4. La observación de una homogeneidad más amplia de PO que sirve para formar el tubo de alimentación en géneros, así como la observación de la presencia de componentes de otras etapas de esporulación se pueden usar como base para clasificar los genomas recién secuenciados en formadores de endosporas probables o no formadores de endosporas.
5. El establecimiento de firmas de esporulación ha sido una cuestión abierta que aún queda por dilucidar, sin embargo, ahora está claro para un pequeño grupo de genes como Spo0A, algunas de las proteínas del complejo Q:AH y los factores sigma de esporulación están bien conservados en Firmicutes, estas y otras proteínas involucradas en etapas posteriores, deben volver a analizarse para integrar un grupo de genes que sirven como marcadores genómicos, considerando no solo su presencia individual sino la presencia de varios componentes de la ruta de esporulación.
6. Finalmente, mayor trabajo experimental y la mejora de nuevos experimentos en condiciones de esporulación no probadas, expondrá los mecanismos que cada género filogenético de Firmicutes ha desarrollado para diferenciarse en esporas y otros destinos relacionados.

XIII. PERSPECTIVAS.

- Realizar un análisis de datación molecular y proponer una hipótesis evolutiva.
- Determinar que residuos son importantes para identificar firmas para cada proteína implicada en la esporulación.
- Determinar la distribución de otras proteínas involucradas en el compromiso como los factores de regulación, factores sigma y otras proteínas de señalización que no interactúan con el canal.
- Hacer un análisis de la conservación de la regulación de la expresión para evaluar su conservación y diferencias en las cascadas de regulación que nos pudieran hablar de las diferentes estrategias adquiridas por las distintas especies.

REFERENCIAS:

1. Abecasis, A.B., Serrano, M., Alves, R., Quintais, L., Pereira-Leal, J.B., y Henriques, A.O. (2013). **A genomic signature and the identification of new sporulation genes.** *J Bacteriol.* 195: 2101–2115.
2. Amador PM, Castañeda N, Loza A, Soto L, Merino E, Maria R, et al. (2019). **Prediction of protein architectures involved in the signaling σ pathway initiating sporulation in Firmicutes.** *BMC Res Notes.* 10–2. doi.org/10.1186/s13104-019-4712-3
3. Altschul S.F., Gish W., Miller W., Myers E.W., Lipman D.J. (1990). **Basic local alignment search tool.** *Mol. Biol.* vol.215: 403-41).
4. Anagnostopoulos C., y Spizizen J. (1961). **Requirements for transformation in *Bacillus subtilis*.** *J. Bacteriol.* 81:741–746.
5. Aung S, Shum J, Abanes-De Mello A, Broder DH, Fredlund-Gutierrez J, Chiba S, Pogliano K. (2007). **Dual localization pathways for the engulfment proteins during *Bacillus subtilis* sporulation.** *Mol Microbiol* 65:1534–1546.
6. Broder DH y Pogliano K. (2006). **Forespore engulfment mediated by a ratchet-like mechanism.** *Cell.* 126: 917–928.
7. Camp AH, Losick R. (2008). **A novel pathway of intercellular signalling in *Bacillus subtilis* involves a protein with similarity to a component of type III secretion channels.** *Mol Microbiol.* 69:402–417.
8. Camp AH, Losick R. (2009). **A feeding tube model for activation of a cell-specific transcription factor during sporulation in *Bacillus subtilis*.** *Genes Dev* 23:1014–1024.
9. Chastanet A., Vitkup D., Yuan GC., Norman T. M., Liu J. S., y Losick R. M. (2010). **Broadly heterogeneous activation of the master regulator for sporulation in *Bacillus subtilis*.** *Proceedings of the National Academy of Sciences of the United States of America.* 107 18: 8486-91.
10. Crawshaw, A.D., Serrano, M., Stanley, W.A., Henriques, A.O., and Salgado, P.S. (2014). **A mother cell-to-forespore channel: current understanding and future challenges.** *FEMS Microbiol Lett.* 358: 129–136.
11. Cruz J, Liu Y, Liang Y, Zhou Y, Wilson M, Dennis JJ, Stothard P, Van Domselaar G, Wishart DS. (2012). **BacMap: an up-to-date electronic atlas of annotated bacterial genomes.** *Nucleic Acids Res.* Vol. 40:D599-D604.
12. de Hoon MJ, Eichenberger P, Vitkup D. (2010). **Hierarchical evolution of the bacterial sporulation network.** *Curr Biol.* 20:R735-R745.
13. Davidson P, Eutsey R, Redler, Hiller NL, Laub MT, Durand D. (2018). **Flexibility and constraint: Evolutionary remodeling of the sporulation initiation pathway in Firmicutes.** *PLoS Genet.* 13;14(9):e1007470.
14. Dembek M, Kelly A, Barwinska-Sendra A, Tarrant E, Stanley W.A, Vollmer D, Biboy J, Gray J, Vollmer W, Salgado P.S. (2018). **Peptidoglycan degradation machinery in *Clostridium difficile* forespore engulfment.** *Mol. Microbiol.* 110: 390-410.

15. Doan T, Marquis KA, Rudner DZ. (2005). **Subcellular localization of a sporulation membrane protein is achieved through a network of interactions along and across the septum.** *Mol Microbiol* 55:1767–1781.
16. Doan T, et al. (2009). **Novel secretion apparatus maintains spore integrity and developmental gene expression in *Bacillus subtilis*.** *PLoS Genet.* 5:e1000566.
17. Dworkin J y Losick R. (2005). **Developmental commitment in a bacterium.** *Cell.* 121: 401–409.
18. Eichenberger P., Jensen S. T., Conlon E. M., van Ooij C., Silvaggi J., Gonzalez-Pastor J. E. et al. (2003). **The sigma E Regulon and the Identification of Additional Sporulation Genes in *Bacillus subtilis*.** *J. Mol. Biol.* 327: 945–972.
19. Errington, J., and Illing, N. (1992). **Establishment of cell-specific transcription during sporulation in *Bacillus subtilis*.** *Mol. Microbiol.* 6: 689–695.
20. Frandsen N., y Stragier P. (1995). **Identification and characterization of the *Bacillus subtilis* spoIIP locus.** *Journal of bacteriology.* 177: 716-722.
21. Fimlaid, K. A. *et al.* (2013). **Global analysis of the sporulation pathway of *Clostridium difficile*.** *PLoS Genet.* 9: e1003660.
22. Fimlaid KA, Jensen O, Donnelly ML, Siegrist MS, Shen A. (2015). **Regulation of *Clostridium difficile* Spore Formation by the SpoIIQ and SpoIIIA Proteins.** *PLoS Genet.* 10: e1005562.
23. Flanagan KA, Comber JD, Mearls E, Fenton C, Wang Erickson AF, Camp AH. (2016). **A membrane-embedded amino acid couples the SpoIIQ channel protein to anti-sigma factor transcriptional repression during *Bacillus subtilis* sporulation.** *J Bacteriol* 198:1451–1463.
24. Galperin MY, Mekhedov SL, Puigbo P, Smirnov S, Wolf YI, Rigden DJ. (2012). **Genomic determinants of sporulation in *Bacilli* and *Clostridia*: towards the minimal set of sporulation-specific genes.** *Environ Microbiol.* 14(11):2870–90.
25. Galperin MY. (2013). **Genome Diversity of Spore-Forming Firmicutes.** *Microbiol Spectr.* 1(2).
26. Hedberg ME, Moore ER, Svensson-Stadler L, Hörstedt P, Baranov V, *et al.* (2012) ***Lachnoanaerobaculum* a new genus in *Lachnospiraceae*; characterization of *Lachnoanaerobaculum umeaense* gen. nov., sp. nov., isolated from human small intestine, *Lachnoanaerobaculum orale* gen. nov., sp. nov., isolated from saliva and reclassification of *Eubacterium saburreum* (Prévot) Holdeman and Moore 1970 as *Lachnoanaerobaculum saburreum* comb. nov.** *Int J Syst Evol Microbiol* 62: 2685–2690.
27. Higgins D, y Dworkin, J. (2012). **Recent progress in *Bacillus subtilis* sporulation.** *FEMS microbiology reviews.* 36 (1): 131-48.
28. Jakovljevic, V., Leonardy, S., Hoppert, M., and Søgaard-Andersen, L. (2008) **PilB and PilT are ATPases acting antagonistically in type IV pili function in *Myxococcus xanthus*.** *J Bacteriol* 190: 2411–2421.

29. Junier I, Rivoire O. (2013). **Synteny in bacterial genomes: inference, organization and evolution.** *ArXiv*. 1307.4291v1.
30. Kläring, K., Hanske, L., Bui, N., Charrier, C., Blaut, M., Haller, D., et al. (2013). ***Intestinimonas butyriciproducens* gen. nov., sp nov., a butyrate-producing bacterium from the mouse intestine.** *Int. J. Syst. Evol. Microbiol.* 63: 4606–4612.
31. Kunst F., Ogasawara N., Moszer I., Albertini A.M., Alloni G., Azevedo V., Bertero M.G., Bessieres P., Bolotin A., Borchert S., et al. (1997). **The complete genome sequence of the gram-positive bacterium *Bacillus subtilis*.** *Nature.* 390: 49–256.
32. Leipe, D. D., Wolf, Y. I., Koonin, E. V. & Aravind, L. (2002). **Classification and evolution of P-loop GTPases and related ATPases.** *J. Mol. Biol.* 317: 41–72.
33. Letunic, I. y Bork, P. (2019). **Interactive tree of life (iTOL) v4: recent updates and new developments.** *Nucleic Acids Research* 47: W256–W259.
34. Lewis PJ, Partridge SR, Errington J. (1994). **σ factors, asymmetry and the determination of cell fate in *Bacillus subtilis*.** *Proc Natl Acad Sci USA.* 91: 3849–3853.
35. Liu, Z., Xu, X., Han, J., Gao, C., Wu, Z., and Hang, F. (2016). ***Paenibacillus bovis* sp. nov., isolated from raw yak (*Bos grunniens*) milk.** *Int. J. Syst. Evol. Microbiol.* 66: 1413–1418.
36. Londono-Vallejo JA, Frehel C y Stragier P. (1997). **SpoIIQ, a forespore-expressed gene required for engulfment in *Bacillus subtilis*.** *Mol Microbiol.* 24:29–39.
37. Margolis P.S., Driks A y Losick R.(1993). **Sporulation gene spoIIB from *Bacillus subtilis*.** *J. Bacteriol.* 175: 528-540.
38. McKenney, P.T. and Eichenberger, P. (2012) **Dynamics of spore coat morphogenesis in *Bacillus subtilis*.** *Mol Microbiol.* 83: 245–260.
39. Meisner, J., Maehigashi, T., Andre, I., Dunham, C.M., and Moran, C.P., Jr. (2012) **Structure of the basal components of a bacterial transporter.** *Proc Natl Acad Sci USA* 109: 5446–5451.
40. Molle V, Fujita M, Jensen S.T, Eichenberger P, Gonzalez-Pastor J.E, Liu J.S, y Losick R. (2003). **The Spo0A regulon of *Bacillus subtilis*.** *Mol. Microbiol.* 50: 1683–1701.
41. Morlot C, Uehara, T, Marquis, K.A, Bernhardt, T.G, y Rudner D.Z. (2010). **A highly coordinated cell wall degradation machine governs spore morphogenesis in *Bacillus subtilis*.** *Genes Dev.* 24: 411–422.
42. Mukherjee S, Stamatis D, Bertsch J, Ovchinnikova G, Katta HY, Mojica A, Chen I-MA, Kyrpidis NC, Reddy T. (2018). **Genomes OnLine database (GOLD) v. 7: updates and new features.** *Nucleic Acids Res.* 8;47(D1): D649–59.
43. Narula, J., Devi, S. N., Fujita, M., and Igoshin, O. A. (2012). **Ultrasensitivity of the *Bacillus subtilis* sporulation decision.** *Proc. Natl. Acad. Sci. U.S.A.* 109: 3513–3522.
44. Paredes C, Alsaker KV, Papoutsakis ET. (2005). **A comparative genomic view of clostridial sporulation and physiology.** *Nat Rev* 3: 969–978.

45. Parker G. F., Daniel R. A., Errington J. (1996). **Timing and genetic regulation of commitment to sporulation in *Bacillus subtilis***. *Microbiology*. 142: 3445–3452.
46. Perez AR, Abanes-De Mello A, Pogliano K. (2000). **SpoIIB localizes to active sites of septal biogenesis and spatially regulates septal thinning during engulfment in *Bacillus subtilis***. *J. Bacteriol.* 182: 1096–1108.
47. Piggot P. J., y Losick R. (2002). **Sporulation genes and intercompartmental regulation. In *Bacillus subtilis* and Its Closest Relative: From Genes to Cells**, A.L. Sonenshein, J.A. Hoch, and R. Losick, eds. (Washington, DC: *ASM Press*). pp. 483–518.
48. Rodrigues, C. D, Ramirez-Guadiana, F. H, Meeske, A. J, Wang, X, y Rudner, D. Z. (2016). **GerM is required to assemble the basal platform of the SpoIIIA–SpoIIQ transenvelope complex during sporulation in *Bacillus subtilis***. *Mol Microbiol.* 102(2): 260-273.
49. Schultz D., Wolynes P. G., Ben Jacob E., y Onuchic J. N. (2009). **Deciding fate in adverse times: sporulation and competence in *Bacillus subtilis***. *Proceedings of the National Academy of Sciences of the United States of America*. 106 (50): 21027-34.
50. Serrano M, Crawshaw AD, Dembek M, Monteiro JM, Pereira FC, de Pinho MG, Fairweather NF, Salgado PS, Henriques AO. (2016). **The SpoIIQ-SpoIIIAH complex of *Clostridium difficile* controls forespore engulfment and late stages of gene expression and spore morphogenesis**. *Mol Microbiol.* 100 204–228.
51. Stadtman TC, McClung LS. (1957). ***Clostridium sticklandii* nov. spec.** *J Bacteriol.* 73: 218-9.
52. Wang S.T., Setlow B., Conlon E.M., Lyon J.L., Imamura D., Sato T., Setlow P., Losick R., y Eichenberger P. (2006). **The forespore line of gene expression in *Bacillus subtilis***. *J. Mol. Biol.* 358: 16–37.
53. Wattam A.R., Davis J.J., Assaf R., Boisvert S., Brettin T., Bun C., Conrad N., Dietrich E.M., Disz T., Gabbard J.L. et al. (2017). **Improvements to PATRIC, the all-bacterial Bioinformatics Database and Analysis Resource Center**. *Nucleic Acids Res.* 45:D535–D542.
54. Wu L.J., y Errington J. (1994). ***Bacillus subtilis* SpoIIIE protein required for DNA segregation during asymmetric cell division**. *Science*. 264: 572–575.
55. Zhao, B., Mesbah, N.M., Dalin, E., Goodwin, L., Nolan, M., Pitluck, S., et al. (2011) **Complete genome sequence of the anaerobic, halophilic alkalithermophile *Natranaerobius thermophilus* JW/NM-WN-LF**. *J Bacteriol* 193: 4023–4024.
56. Zeytuni N, Flanagan KA, Worrall LJ, Massoni SC, Camp AH, Strynadka NCJ. (2018). **Structural characterization of SpoIIAB sporulation-essential protein in *Bacillus subtilis***. *J Struct Biol.* 202: 105–112.
57. Zeytuni, N., and Strynadka, N.C.J. (2019). **A hybrid secretion system facilitates bacterial sporulation: A structural perspective**. *Microbiol. Spectr.* 7.