RC-07-098
REV. 00-3/09

UNIVERSIDAD AUTÓNOMA DE NUEVO LEÓN

FACULTAD DE CIENCIAS BIOLOGICAS

Instituto de Biotecnología

Análisis genómico de *Bacillus thuringiensis*: Estructura Poblacional"

Por

QBP. AMADA TORRES SALAZAR

T E S I S

Como requisito parcial para obtener el Grado de

DOCTOR EN CIENCIAS

con especialidad en BIOTECNOLOGIA

San Nicolás de los Garza, N. L Junio 2012

Der einzige Mensch,
mit dem du ein Leben lang
zusammen bist,
bist du selbst !
Darum höre auch auf dich.....
denn nur du weißt,
was dir gut tut.....
welche Wünsche du hast.....
welche Träume du hast.....
hab dich gern,
denn das ist viel wert.....
tue das,
was du willst.....
was dir gefällt.....
wozu du Lust hast.....
was dir Spaß macht.....
was dich glücklich macht.....
aber tue es.....
denn dieses Leben,
lebst du nur einmal.

Sogar die Wissenschaft kann ohne Schönheit nicht bestehen [...], nicht ein Nagel würde mehr erfunden werden!

Even science cannot exist without beauty [...], not a single nail would be re-invented!

Fjodor M. Dostojewski,
Die Dämonen

AGRADECIMIENTOS

Al Consejo Nacional de Ciencia y Tecnología por el apoyo numero 18533 otorgado en la convocatoria de Becas CONACYT Nacionales Enero-Julio 2008, que comprendió el periodo de Febrero 2008 a Enero 2010, el ajuste de vigencia solicitado en mayo 2011 hasta Enero de 2012. De igual manera por el apoyo otorgado en la Convocatoria de Becas de Inversión en el Conocimiento 2008-Marzo 2009, en el Programa de Becas-Mixtas en el Extranjero para la estancia académica realizada en el Duetsches Krebsforschungszentrum (DKFZ) en la ciudad de Heidelberg, Alemania.

LISTA DE TABLAS

LISTA DE FIGURAS

NOMENCLATURA

DNA	Por sus siglas de inglés: *Deoxyribonucleic Acid*, ácido desoxirribunucleico
RNA	Por sus siglas de inglés: *Ribonucleic Acid*, ácido ribonucleico
ORF	Por sus siglas de inglés: *Open reading frames*. Marco de lectura abierta, que contiene un sitio de inicio y terminación de la traducción en la misma cadena.
Indel	Por sus siglas de inglés: Insertion/Deletion, Combinación del inicio de las palabras inserción y deleción, que identifica mutaciones puntuales en el genoma
MLEE	Por sus siglas de inglés: *Multilocus Enzyme Electrophoresis*, Técnica molecular basada en la diferente movilidad electroforética que presenta un locus o alelo particular de una enzima con actividad conocida. Los alelos de cada uno de los loci define el perfil alélico
AFLP	Por sus siglas de inglés: *Amplified Fragment Length Polymorphism*. polimorfismo de los fragmentos amplificados
MLST	Por sus siglas de inglés: *Multilocus sequence typing*. Es una técnica molecular para la identificación de loci múltiples. Se utiliza la secuencia del DNA de los múltiples fragmentos internos de genes compatibles con la vida. Los alelos de cada uno de los loci define la secuencia tipo
CDS	Por sus siglas de inglés:
CFU	Por sus siglas de inglés, *Colony-Forming Unit*,
primer walks	secuencia se utilizó el método de caminando en el iniciador
kb	Kilobase de DNA
GC	Contenido de Guanina-Citosina en una secuencia de DNA dada
AT	Contenido de Adenina-Timina en una secuencia de DNA dada
Mb	Megabases de DNA

CDS	Regiones codificantes del DNA
UPGMA	Por sus siglas de inglés: *Unweighted Pair Group Method with Arithmetic Averages*. Método de agrupación de datos que utiliza los promedios aritmético no ponderados como medida de la diferencia entre los taxa
pb	par de bases de DNA

TABLA DE CONTENIDO

RESUMEN

La biología de *Bacillus thuringiensis* presenta preguntas sin responder. Una de ellas es la referente a su real nicho ecológico en la naturaleza; una más relacionada a su identidad como especie, mientras que los estudios realizados para conocer la estructura poblacional no son concluyentes y la variabilidad a nivel genómico sigue sin ser apropiadamente cuantificada. Estos elementos genómicos son de particular relevancia para establecer posibles relacione evolutivas o hacer un uso inteligente y racional de todos los beneficios que esta interesante bacteria posee.

De tal suerte que, para obtener información de interés biotecnológico, una comparación sistemática utilizando los perfiles completos de proteínas codificadas en los cromosomas de 3 genomas completos y 15 genomas parciales (*contig*) de las diferentes variedades de *Bacillus thuringiensis* fue elaborada. Por medios de un análisis fenético de la relación existente en una gran cantidad de datos multiestado y la comparación con aquellos que se generan en una codificación como datos binarios se describió la estructura genómico poblacional. Adicionalmente, se probó si la forma de codificación de los datos influye en la representación de la diversidad y estructura genética de la especie.

Los resultados de la variabilidad encontrados son los más altos reportados al momento H = 0,999, mientras que la estructura poblacional encontrada presenta un patrón epidémico. La forma de codificación de los datos, presenta diferencias en el cálculo de la variabilidad cuando utilizamos el índice de Shannon para ello. Por otro lado, la estructura poblacional se conserva igual, independientemente de la forma de codificación de los datos.

Esta aproximación utilizando el contenido completo de cromosomas de *B. thuringiensis* es una herramienta muy poderosa para entender la variación global en la especie. En particular, este análisis presenta la pérdida de algunas proteínas entre las cepas y la variabilidad en el número de copias de genes identificados y anotados.

PALABRAS CLAVE

Variabilidad genómica, variabilidad de proteínas, transferencia horizontal del genes, estructura genómico poblacional

ABSTRACT

Genetic structure and variability of eighteen full-sequenced *Bacillus thuringiensis* strains worldwide collected using *in silico* chromosome protein comparison was investigated. We explore the variability of the protein genome profile in two approaches, encoded as multistate and binary data set. Processed protein-code database contains 3877 entries from 19 chromosomes sequence, 18 of *B. thuringiensis* and one of *B. subtilis* used as out group. 60% of the multistate data was unicopy loci. In contrast, approximately 67% of the binary data was variable loci, with a mean 24% of exclusive proteins. The performed Shannon index showed high values in the approach as multistate (average 7.76 ±0.03) as compared with binary data (average 6.71 ±0.02). In other context, both approaches showed a monophyletic phenograms, the strains of *B thuringiensis* were grouped into three subgroups. The distributions of the strains suggest one epidemic bacterial population structure. *Bacillus thuringiensis* strains making one cohesive taxon with high variability, short genetics distances and epidemical population structure. These strains contain a diverse range of mobile elements that contribute to genome dynamics and that may also have phenotypic and biotechnological impact. This is the first report that formality quantifies protein variability and gives the population structure in the specie using systematic high-throughput approach using two approaches. In particular, this analysis showed the lack of several proteins between the strains and the variability in the copy number of each identified and annotated proteins.

KEY WORDS

Protein variability, genome variability, horizontal gene transfer, genomic population structure.

1. INTRODUCCIÓN

El estudio tradicional de la biología molecular y celular se ha centrado en el uso de unos pocos genes en cada ensayo. De manera contrastante, utilizando las tecnologías de alto desempeño, por una parte, la secuenciación del contenido completo de macromoléculas, como el DNA, RNA o proteínas, y por otra parte; la medición de la expresión de los genes o proteínas y como éstos se regulan, han generan las áreas de la genómica, transcriptomica, proteómica, reguloma y metabolómica que pretenden hacer mediciones simultáneas de un grupo de macromoléculas en una condición particular, todas ellas apoyadas en los análisis numéricos, informáticos y bioinformáticos.

El conocimiento derivado de cada una de estas aproximaciones, es evidentemente, diferente y permiten tener panoramas distintos de los sistemas analizados. Si bien pueden ser complementarios, las herramientas y el esfuerzo para obtener conclusiones son por mucho, más demandante en el segundo caso.

Estamos siendo testigos de los cambios que la tecnología de alto desempeño esta generando en todo el mundo. Los reportes de secuencias genómicas completas están al orden del día, resaltan los estudios que pretenden conocer la variabilidad poblacional en humanos con el programa multicéntrico de los 1000 genomas, impulsados principalmente, por industria farmacéutica en un esfuerzo conjunto para conocer la susceptibilidad a una determinada enfermedad o predecir un efecto adverso a drogas y las potenciales interacciones que se presentan con el ambiente. La gran promesa de este proyecto es la posibilidad de generar una base de datos que sustente la medicina personalizada (Kuehn 2008; Via *et al.* 2010).

De forma similar, la secuenciación intensiva y masiva de organismos con importantes nexos con los humanos, en el denominado "Proyecto Genoma" ha incluido a especies particularmente interesante, en las que se incluye el género *Bacillus*, que por sus connotaciones a: i) la salud pública, por poseer especies patógenas a humanos; ii) la industria en general, por dotarla de especies productoras de metabolitos con aplicaciones a la cadena alimenticia humana; y iii) por supuesto, su utilidad como arma biológica de

destrucción masiva, ha permitido la rápida generación de bases de datos con genomas completos de este versátil genero.

El proyecto del grupo *Bacillus cereus*, desarrollado por el *Bacillus cereus* Group Genome Project (JCVI), tiene como vertientes expandir el entendimiento de la patogenicidad, mediante la secuenciación y anotación de los genomas de diez cepas y acompasar la amplia diversidad que muestra este grupo de organismos.

Adicionalmente, la secuencia genómica de *Bacillus thuringiensis* fue desarrollada con objeto de ayudar a la determinación de los genes responsables de la adaptación a nuevos nichos y el metabolismo celular de la especie, al igual de ser referente de otras especies del grupo, en particular de *B. anthracis* y otros *Bacillus* del Grupo I, en vías que facilite las investigaciones bio-forense en salud pública y defensa. Colateralmente, estas bases de datos, también permitirán obtener un conocimiento ecológico, fisiológico, metabólico y evolutivo de las especies que conforman el complejo grupo de *B. cereus*.

De todos conocido, es que la variabilidad microbiológica es el resultado de la acción de una serie de fuerzas ecológicas y evolutivas que afectan a las poblaciones microbianas, como es la selección natural, flujo génico, deriva génica, mutación y transferencia horizontal de genes (Whittam *et al.* 1983; Selander *et al.* 1985; Selander *et al.* 1986). La suma de todos estos factores, dan características que hacen particularmente diferentes a las poblaciones dentro de una misma especie, ya sean en su morfología o incluso pueden cambiar drásticamente su biología.

Esta variabilidad, por otra parte, nos puede ayudar a conocer que tan estructurada se encuentra una especie en la naturaleza. Su medición, permite describir tres posibles escenarios. Una clásica estructura clonal, caracterizada por genotipos estrechamente relacionados con poca asociación de alelos y variabilidad producida por escasos procesos mutaciones. Una estructura epidémica, en donde es posible distinguir linajes de clones con procesos de recombinación frecuentes, loci pérdidos y otros sobrerrepresentados, estables en localidades separadas en tiempo y distancia. Así, como la típica estructura panmitíca, que recuerda una reproducción sexual, con un continúo intercambio y flujo génico que forma una variabilidad genética compartida entre las poblaciones (Tibayrenc *et al.* 1990; Istock *et al.* 1992; Lenski 1993; Smith *et al.* 1993).

De manera que el reto al que se enfrenta la genómica microbiana como ciencia, es demostrar su valor en cuanto a la generación de información relevante biológicamente funcional. En nuestra primera aproximación en el área, generamos la estructura genómico poblacional de *Bacillus thuringiensis* con base en las secuencias genómicas reportadas.

2 PLANTAMIENTO DEL PROBLEMA y JUSTIFICACIÓN

A la fecha no existe una documentación adecuada de la variabilidad genética o genómica presente en el cromosoma de *Bacillus thuringiensis*, por lo que tampoco es conocido la estructura poblacional que describen las diversas cepas. Estos elementos genómicos son de particular relevancia para establecer posibles relaciones evolutivas o hacer un uso inteligente y racional de todos los beneficios que esta interesante bacteria posee.

Las bases de datos públicas que contienen las secuencia del genoma completo de esta especie, representa una oportunidad inmejorable para poner a prueba una estrategia global de generación de conocimiento biológico que sirva de base a la industria biotecnológica. Sobre todo si se toma en cuenta los altos costos en tiempo, dinero y recursos humanos necesarios para desarrollar una búsqueda clásica de actividades biológicas de gen por gen.

3 HIPOTESIS

Si el contenido genómico del cromosoma de *Bacillus thuringiensis* permite cuantificar la variabilidad presente en la especie, entonces se podrá reconstruir su estructura genómico poblacional.

4. OBJETIVOS

4.1 Objetivo General

Analizar y cuantificar la variabilidad genética de *Bacillus thuringiensis* para construir su estructura genómico poblacional

4.2 Objetivos particulares

1. Obtener de bases de datos públicas las secuencias genómicas cromosomales de *Bacillus thuringiensis* disponibles

2. Utilizando diferentes métricas, cuantificar la variabilidad en la especie

3. Representar gráficamente la estructura poblacional de *Bacillus thuringiensis* descrita por su contenido cromosómico

5 ANTECEDENTES

La biología de *Bacillus thuringiensis* presenta dos preguntas sin respuesta hasta el momento. Una es referente al real nicho ecológico de la especie y la otra relacionada con su identidad. Es decir, al tiempo de escribir él presente, no se ha establecido claramente si *B. thuringiensis* es una variedad de *Bacillus cereus*, o forma una especie aparte. Adicional a esta incertidumbre, se desconoce la función biológica que efectúa en la naturaleza. De forma relevante, la variabilidad a nivel genómico no se ha cuantificado de manera fehaciente y por tanto la estructura genética poblacional también sigue en duda.

5.1 Descripción de la especie

Bacillus thuringiensis es considerada una bacteria ubicua, debido a que ha sido aislada de diversos ecosistemas como suelo (Addison 1993; Meadwos 1993; Iriarte *et al.* 1998), en tierra de cultivo (Bel *et al.* 1997; Arrieta *et al.* 2004) (Dangar *et al.* 2010) ambientes salinos y sumergidos en agua (Das *et al.* 2006); hojas de plantas (Bel *et al.* 1997; Damgaard *et al.* 1998; Mizuki *et al.* 1999; Freitas *et al.* 2008), pastos (Damgaard *et al.* 1998) y granos para alimentación de ganado (Meadows *et al.* 1992) (Apaydin 2004); sin sorprender los hallazgos en el intestino de animales y gusanos (Swiecicka *et al.* 2002), lepidópteros muertos de diferentes partes del mundo (Ohba *et al.* 1994; Iriarte *et al.* 1998; Schnepf *et al.* 1998) y en diferentes órganos del humano(Turnbull *et al.* 2003) . Existen grandes colecciones de aislados de todo tipo de fuentes naturales y agroecosistemas en el mundo entero (Lecadet *et al.* 1999).

B. thuringiensis es un bacilo Gram positivo anaerobio facultativo, quimiorganótrofo, mesofílico, catalasa positivo y móvil por flagelos peritrícos. Las células vegetativas toleran un amplio rango de pH. La forma de espora es resistente a un amplio intervalo de temperatura y condiciones ambientales, incluyendo la radiación UV(Myasnik *et al.* 2001; Nicholson 2002; Saxena *et al.* 2002; Nicholson *et al.* 2005) e incluso presenta resistencia a desinfectantes químicos (Schmidt 1955). Forman inclusiones cristalinas visibles al microscopio óptico durante la esporulación. Las inclusiones han sido denominadas como proteínas parasporales portadoras de toxina. Los aislados pueden ser clasificados de acuerdo a propiedades fisiológicas y al antígeno flagelar H. La reacción antigénica ante el antígeno flagelar H, reporta un total de 82 sero-variedades que se integran en 69 serotipos y 13 sub grupos antigénicos en la colección del International Entomopathogenic *Bacillus* Center, en Paris, Francia; la cual contiene 3500 aislados (Lecadet *et al.* 1999). Por otro lado, el sitio web del NCBI han reportado 124 sero-variedades (http://www.ncbi.nlm.nih.gov/Taxonomy/Browser/wwwtax.cgi). En México la CONABIO, reporta 40 sero-variedades en su colección (CIBIOGEM-CONABIO).

5.2 Taxonomía de *Bacillus thuringiensis*

Los primeros intentos de clasificación de las especies de *Bacillus* se basaron en dos características: el crecimiento aeróbico y la formación de esporas. Esto dio como resultado la unión de bacterias que poseen diferentes tipos de fisiología y la ocupación de una gran variedad de nichos. Por lo tanto la heterogeneidad en la fisiología, la ecología y la genética ha dificultado su organización.

La edición del Bergey de 1986, reconoce 40 especies, sin embargo varias nuevas especies se han designado desde 1986, mientras que basado en la comparación filogenética con ciertas especies de *Bacillus* otras especies se han trasladado a nuevos o existentes géneros de bacterias formadoras de esporas (Fritze 2004). Actualmente se reconocen 246 taxa descritas del género (Brands 2004-2011). Muchas especies en este taxón son de importancia clínica, como es el grupo *B. cereus*; compuesto por *B. cereus*, *B. anthracis*, *B. thuringiensis*, *B. mycoides*, *B. pseudomycoides* y *B. wiehenstephanensis*, los miembros de este grupo comparten una gran cantidad de similitudes morfológicas y bioquímicas (Blackwood *et al.* 2004). Al conjunto de estas especies también se le conoce como *B. cereus sensu scricto*. Mientras que *B. cereus*, *B. anthracis* y *B. thuringiensis* pueden ser consideradas como miembros del grupo de *B. cereus sensu lato* (Daffonchio *et al.* 2000).

El contenido de GC en el DNA en el género varía de 32 a 69%. Los estudios de hibridación de DNA-DNA sugiere una alta similitud ente *Bacillus* cereus, *B. anthracis* and *B. thuringiensis* (Drobniewski 1993).Los estudios comparativos utilizando la secuencia de la sub unidad pequeña del DNA ribosomal reporta un 94% de similitud entre las cepas de *Bacillus* anthracis y *B. cereus*, mientras que *B. thuringiensis* y *B. cereus* solo difieren en nueve nucleótidos (Ash *et al.* 1991).

Estudios adicionales de variabilidad molecular en las especies del grupo de *B. cereus sensu lato* utilizando diversas técnicas, como MLEE (Helgason *et al.* 1998), MLST (Helgason *et al.* 2000b), AFLP y huella genética (Ticknor *et al.* 2001) sugieren que tanto *B. thuringiensis* como *B. anthracis* deben ser clasificadas como subespecies de *B. cereus*. De hecho, el polimorfismo genético del grupo, está principalmente representado por las cepas de *B. cereus* y *B. thuringiensis*, esta amplia similitud en sus conjuntos genéticos son los factores primordiales de complicación para la resolución de la cuestión taxonómica entre estas dos especies. Entre tanto, se reporta una

homogeneidad genética en las cepas de *B. anthracis* (Harrell *et al.* 1995; Read *et al.* 2002)

Debido a él alto grado de variabilidad genética que se ha observado dentro y entre las especies, se ha sugerido que *B. cereus* y *B. thuringiensis* deben considerarse como una sola especie (Carlson *et al.* 1994). Por otro lado, se piensa que los estudios, basados en los resultados de AFLP, MLEE, MLST y el análisis de secuencia de la subunidad pequeña del DNA ribosomal en un número limitado de cepas no es suficiente para clasificar estas especies como una entidad única (Vilas-Boas *et al.* 2007).

De acuerdo con la clasificación formal más actual se establece que *Bacillus thuringiensis* pertenece al Dominio: Bacteria (Haeckel, 1894) Woese, Kandler & Wheelis, 1990; Al Phylum: Firmicutes (Gibbons & Murray, 1978); Clase: "Bacilli"; Orden: Bacillales (Prévot, 1953); Familia: Bacillaceae (Fischer, 1895); Genero: *Bacillus*™ (Cohn, 1872, nom. approb); Especie: *Bacillus thuringiensis* Berliner, 1915. Con el sinómino: "*Bacillus* cereus var. *thuringiensis*" Smith et al. 1952 (Brands 2004-2011).

5.3 Genomas secuenciados de *Bacillus thuringiensis*

La secuencia completa de *B. thuringiensis* fue desarrollada con objeto de ayudar a la determinación de los genes responsables de la patogenicidad celular, la adaptación a nuevos nichos y el metabolismo celular de la especie, al igual de ser referente de otras especies del grupo *B cereus sensu lato*, en particular de *Bacillus anthracis* y otros *Bacillus* del Grupo I.

El sito del NCBI ha reportado las variantes de konkukian 97-27 (Han *et al.* 2006) y Al Hakam (Challacombe *et al.* 2007) proyectos del DOE Joint Genome Institute y la subsp islaerensis ATCC 35646

Bacillus thuringiensis var konkukian cepa 97-27, es una cepa aislada de un caso de necrosis severa en tejido humano de un soldado francés herido por la explosión de una bomba en la antigua Yugoslavia (Hernandez *et al.* 1998) e identificada como serotipo H34. Estudios con MLST mostraron que está fuertemente relacionada con cepas de *B. cereus* (Hoffmaster *et al.* 2006)

Bacillus thuringiensis var Al Hakam fue secuenciada utilizando bibliotecas de plámidos y fosmidos. Para ensamblar y cerrar la secuencia se utilizó el método de *primer walks* y sucesivas amplificaciones por PCR. En esta cepa se describieron 4,969 ORF y al menos 21 pseudogenes (Challacombe *et al.* 2007). Esta variante anteriormente fue relacionada con *B. anthracis* por AFLP y con *B. cereus* por MLST (Hoffmaster *et al.* 2006)

La tercera secuencia corresponde a la subsp isralensis ATCC 35646 (HD522) productora de toxinas con actividad insecticida, hasta 2009 fue reportada en ensamblaje y se alojó en la base de datos del NCBI, sin embargo fue retirada. El proyecto corrió a cargo del Integrated Genomics. A la fecha en el sitio Patric se encuentra esta misma cepa con varios órdenes de repetición.

En mayo de 2010, el grupo Chino del Dr. Ziniu Yu, reportó la finalización y anotación del genoma completo de la cepa mutante de *Bacillus thuringiensis* BMB171 que carece de proteínas de cristal pero tiene una alta frecuencia de transformación. Esta cepa, fue secuenciada utilizando la tecnología paralela y masiva de pirosecuenciación (He *et al.* 2010).

5.4 Diversidad biológica en *Bacillus thuringiensis*

Algunos reportes de la variabilidad de la especie, realizan una medición cualitativa, hecho que evita una comparación con otros reportes y así tener un conocimiento real de este importante atributo ecológico. Por otro lado, la mayoría de estos estudio fueron desarrollados utilizando los resultados de MLEE, en donde una gran parte de ellos se ejecutaron en una mezcla de cepas y especies en las que el límite inferior de variabilidad fue encontrado en cepas aisladas de casos de periodontitis con 0.150 (Helgason *et al.* 2000a) y el límite superior, hasta el momento, fue descrito por cepas aisladas de suelo Noruego con 0.526 (Helgason *et al.* 1998). El único estudio en donde se analizan las especies por separado es lo reportado en poblaciones simpátricas aisladas de muestras de suelo por Vilas-Boas y col. (2002), quienes para *B. thuringiensis* reportaron un valor de 0.283, mientras que para las cepas de *B. cereus* fue de 0.310.

5.5 Estructura genética-genómica poblacional de *Bacillus thuringiensis*

Mientras que los estudios realizados para conocer la estructura poblacional de *B. thuringiensis*, no son concluyentes y en algunos casos se han mezclado con los correspondientes al grupo *B. cereus*, como si estos fueran una sola entidad.

El único reporte utilizando una sola especie, es lo encontrado en *B. thuringiensis* var israelensis H14, en aislados proveniente de muestras de suelo Sueco en donde fue descrita una estructura clonal (Ankarloo *et al.* 2000), similar a la estructura que describe *B. anthracis* (Ticknor *et al.* 2001). El resto de los estudios poblacionales varían en sus resultados desde una estructura poco clonal (Ehling-Schulz *et al.* 2005), no clonal (Priest *et al.* 2004), pasando por una estructura mixta (Helgason *et al.* 1998), hasta lo reportado como una estructura incongruente (Ko *et al.* 2004).

6 MÉTODO

6.1 Unidad de estudio

Debido a la dificultad técnica para separar un solo individuo, el estudio de microrganismos se basa en CFU. Es posible tratar las CFU como poblaciones, si consideramos la definición de ésta, como un grupo de individuos emparentados que existen en una unidad de tiempo y espacio (Hedrick 2009). En congruencia con esta definición, una CFU es una población, por extensión, la sofisticada teoría matemática del estudio de la genética poblacional se utilizó como el fondo de soporte para este estudio.

6.2 Obtención de Secuencias

En febrero de 2009 se encontraban en progreso los proyectos de secuenciación de los aislados: Bt407, IBL200, IBL4222, *andalousiensis* BGSC 4AW1, *berliner* ATCC10792, *huazhongensis* BGSC 4BD1, *kurstaki* T03a001, *monterrey* BGSC 4AJ1, *pakistani* Tl3001, *pondicheriensis* BGSC 4BA1, *pulsiensis* BGSC 4CC1, *soto* T04001, *thuringiensis* T01001 y *tochigiensis* BGSC 4Y1 a cargo del Naval Medical Research Center. Los aislados de las serovariedades *darmstadiensis* y *kurstaki* por la Huazhong Agricultural University de China. México contribuiría con las seroariedades *darmstadiensis* y *entomocidus* realizadas por la Universidad Michoacana de Hidalgo y el CINVESTAV-IPN en Irapuato. Adicionalmente el sitio NCBI albergó 3 secuencias más, dos completas y una en ensamblaje. Para mayo de 2010 obtuvimos 15 secuencias genómica de *B. thuringiensis* y una *B. cereus* de la bases de datos pública PATRIC. En julio de 2011 se actualizaron las secuencias, en donde obtuvimos tres secuencias adicionales de *B. thuringiensis*.

6.3 Análisis genómico estructural

Desarrollamos un análisis estructural comparativo con las secuencias genómicas de *B thuringiensis* como se describe a continuación. Para la comparación del contenido de GC en las cepas utilizamos el sitio Panthema-*Bacillus*, en donde fue calculado el contenido de nucleótidos, así como el porcentaje AT y GC. De igual manera utilizamos el programa Genome Atlas v 3.0 (Hallin *et al.* 2004; Ussery *et al.* 2008), usando las bases de datos del mismo portal. Para las cepas restantes utilizamos el programa Matlab R2008a v 7.4 (The MathWorks 1984-2008).

6.3.1 Tamaño del genoma

En general la densidad en el contenido de los genomas bacterianos, no varía mucho más de aproximadamente un gen por kb de DNA. Claramente, los tamaños genómicos son directamente proporcionales al número de genes que contienen. Calculamos y comparamos el tamaño de los genomas disponibles mediante gráficas de barras.

6.3.2 Contenido de nucleótidos

La composición de bases generalmente muestra una correlación entre el tamaño del genoma y su contenido de Guanina-Citosina (GC). El contenido del nucleótido "C" de una secuencia dada, es la fracción de "C" en la secuencia en cuestión. Este puede tomar valores entre cero (cuando no hay C) hasta 1 (cuando todos los nucleótidos son C), sin embargo se reportan como un porcentaje del contenido. Para una secuencia que contiene un 50% de AT, se puede esperar, rigurosamente un contenido del 25% de C. En forma similar se calculan los contenidos de T, G y A (Jensen *et al.* 1999).

6.3.3 Porcentaje de Adenina-Timina

El porcentaje del contenido de Adenina-Timina (AT) es el promedio del contenido de estas bases, en una ventana de tamaño especifico. Típicamente, para un genoma bacteriano de aproximadamente 5Mb de tamaño, se asume una ventana de 10,000 bases.

6.3.4 Porcentaje Guanina-Citosina

El porcentaje del contenido de Guanina-Citosina (GC) es el promedio del contenido de estas bases, en una ventana de tamaño especifico. Típicamente, para un genoma bacteriano de aproximadamente 5Mb de tamaño, se asume una ventana de 10,000 bases.

6.3.5 Sintenía

El concepto de sintenía se refiere a regiones de multigenes donde la secuencia de DNA y el orden de los genes esta conservada entre dos genomas. Puede ser realizada utilizando gráficas de puntos que representan un alineamiento pareado entre genomas (Eisen *et al.* 2000). De manera adicional, la sintenía se obtuvo de un alineamiento múltiple genómico realizado en Mauve (Darling *et al.* 2008; Darling *et al.* 2010).

6 4 Edición de secuencias

Debido al papel crucial que constituye la identificación de una función particular a partir de una secuencia, se ha desarrollado toda una serie de herramientas y técnicas para realizar esta tarea, lo que ahora denominamos como anotación. Se sabe que en la mayoría de las secuencias nuevas la asignación de la función es hecha a través de la utilización de la anotación. El 97% de las bases de datos realiza esta anotación utilizando evidencia electrónica únicamente. También es conocido que la base de datos de UniProt, solo tiene un 3% de las proteínas con una anotación afirmativa, es decir, los que no están etiquetados como hipotéticos. Estas anotaciones son depositadas en el GenBank donde persiste el evento y eventualmente constituye un error sistemático (Gilks *et al.* 2002) que se conoce como una descripción falsa-positiva con respecto a la función biológica. Los errores sistemáticos asociados con este protocolo de anotación se han estudiado extensivamente y se han dado algunas de las causas de estos errores (Galperin *et al.* 1998; Brenner 1999; Jones *et al.* 2007; Medigue *et al.* 2007; Schnoes *et al.* 2009; Koser *et al.* 2011). Entre ellos podemos separar tres grupos principales: i) los que corresponde a procesos biológicos como, la duplicación de genes, resolución de dominios, fusión y fisión de genes, distancias evolutivas, etc.; ii) los de tipo ortográfico, incluyendo omisiones parciales o totales de la escritura correcta (Brent 2005) (Green *et al.* 2005) y iii) los debido a los algoritmos utilizados y a su poder de predicción (Andorf *et al.* 2007; Jones *et al.* 2007). Por todo ello se ha implementado la denominada curación manual de las bases de datos (Brent 2005). En nuestro caso, nos enfocamos a los errores de tipo dos, una vez que se trabajó con una sola especie. Nos apoyamos en la hoja de cálculo Excel de MsOffice 2007y los manejadores de bases de datos MySQL Server 2000 (Microsoft Co) y Oracle (Edwards 2007). La actualización de algunas regiones codificantes se realizó en función de lo reportado en las bases de datos de referencia: UniProt (http://www.uniprot.org/), Brenda (http://www.brenda-enzymes.org/) y Gene home (http://www.ncbi.nlm.nih.gov/gene), propuestos como los mejores referentes al momento.

6.5 Codificación de CDS de cada secuencia genómica para realizar un análisis numérico

Numerosos autores han propuesto métodos numéricos para la reconstrucción de filogenias conceptualmente más relacionados con la taxonomía fenética, basados en la idea de agrupar los taxones por su similitud global (Wagner 1961) (Edwards *et al.* 1964) (Camin *et al.* 1965; Cavalli-Sforza *et al.* 1967) (Fitch *et al.* 1967) (Sneath *et al.* 1973). En la actualidad, la taxonomía fenética es muy útil para resolver problemas de microtaxonomía como el presentado en este estudio.

Para realizar un análisis numérico de un taxa, a partir de datos obtenidos por técnicas moleculares, es indispensable conocer el patrón de herencia que estos datos presentan. Debido a que de acuerdo a este patrón, se pueden elegir las métricas a utilizar en el análisis. Regularmente los datos generados por MLEE son codificados en un sistema multiestado, con allelos que presentas diferentes frecuencias, se asume que los genes que codifican estas enzimas son selectivamente neutros y siguen un sistema Mendeliano de Codominancia (Selander *et al.* 1985; Selander *et al.* 1986). Por otro lado, los datos generados por RAPD se denominan como el perfecto modelo de Dominancia Mendeliana y son codificados como marcadores binarios. Por la naturaleza de amplificación aleatoria de las moléculas blanco se ha propuesto que los marcadores son neutrales (Williams *et al.* 1990). Los patrones de bandas generados por estas técnicas, son asignados como un perfil, que posteriormente se utiliza para describir las diferencias entre los individuos dentro de una población o entre poblaciones distintas.

De acuerdo con Smith, los resultados de la estructura población dependen de cómo estos datos sean analizados (Smith *et al.* 1993). Por ello, en este estudio, codificamos nuestros datos en los dos sistemas anteriormente mencionados.

6.7 Matrices de Datos

Una vez que contamos con toda la información referida a los estados de variación que están presentes en las secuencias genómicas curadas, procedimos a transformarla en una matriz de datos. Por convención, los taxones son las filas y los caracteres las columnas. Incluimos los datos obtenido de *B. subtilis* como grupos externo, de manera que la reconstrucción puede tener raíz.

La primera matriz de datos, se generó transformando los CDS como caracteres binarios; en el caso de la presencia del CDS en una secuencia particular se asignó el valor de (1), en caso de ausencia el valor fue de (0). De esta manera se ensambla el sistema binario que permite hacer un análisis de Dominancia Mendeliano

La segunda matriz, adicionalmente tomó en cuenta el contenido del total de copias de cada CDS presentes en la secuencia dada, estos fueron transformados en frecuencias relativas para su análisis y presentados en una matriz multiestado que permitirá tratar los datos como si se tratara de un sistema Medeliano de Codominancia.

6.8 Variabilidad genómica

La variabilidad genómica puede ser entendida de muy diferentes formas, por el tamaño de los genomas, el número de proteínas codificadas, por la simple diferencia en el número de copias o por la ausencia-presencia de un gen o proteína específico. Realizamos el análisis fenético de las secuencias reportadas, utilizando la variabilidad intra-especifica vía diferentes índices y programas como PAST v 2.07 (Oyvind *et al.* 2001), SPSS v 13.0 (The Appache Software Foundation 1989-2004) y Matlab R2008a v 7.4 (The MathWorks 1984-2008)

El promedio de heterozigocidad fue calculado utilizando los supuestos del equilibrio de Hardy-Weinberg, $H_j = 1-\sum pi\ 2$, donde *pi* es la frecuencia de un *ith* alelo en el *jth* locus de la cepa (Ward *et al.* 1992).

Correlacionamos la variabilidad obtenida de los alelos variables, exclusivos y polimórficos con la heterozigocidad. Un locus polimórfico se define como el locus con dos o más copias (alelos) en la cepa o genoma.

6.9 Estructura genómica poblacional

Es a través de la variabilidad genética que se puede establecer si las poblaciones de una especie presentan una estructura poblacional particular. La estructura población por tanto, no es sino el saber cómo esta variabilidad está repartida entre las poblaciones. Utilizando los Índices de Jaccard (Jaccard 1901) para la matriz de datos binarios y Mannhatan (Minkowski 1910) para la matriz de datos multiestado Ambos fueron agrupados por UPGMA (Sneath *et al.* 1973) con un valor de beta de 0.05, se reconstruyeron las gráficas que representan la estructura poblacional de *B. thuringiensis*, las métricas se calcularon en el programa NTSYs-pc 2.01i (Rohlf 1993), las representaciones gráficas fueron mejoradas con el programa Dendroscope v 2.7.4 (Huson *et al.* 2010). Probamos la correlación entre dos grupos independientes de datos medidos como distancias pareadas y similitud entre las cepas.

6.10 Recursos computacionales

Todos los análisis se desarrollaron en una computadora portátil Toshiba Dual Core, con 2 G de memoria RAM con dos procesadores de 1.66 GMz con acceso a internet.

7 RESULTADOS

7.1 Obtención de Secuencias

La secuencia completa de *B. thuringiensis* fue desarrollada con objeto de ayudar a la determinación de los genes responsables de la adaptación a nuevos nichos y el metabolismo celular de la especie, al igual que, de ser referente de otras especies del grupo, en particular de *Bacillus anthracis* y otros Bacillus del Grupo I, esto permitió obtener las secuencias públicas de 19 genomas de los sitios: National Center for Biotechnology Information NCBI (http://www.ncbi.nlm.nih.gov/), en el TIGR Comprehensive Microbial Resource (http://cmr.jcvi.org/tigr-scripts/CMR/CmrHomePage.cgi), en el sitio Pathema-Bacillus, Bioinformatics Resource Center (http://pathema.jcvi.org/cgi-bin/Bacillus/PathemaHomePage.cgi), utilizando Ensembl Genome Browse del sitio EMBO-EBI (http://www.ensembl.org/index.html) y del sitio web Pathosystems Resource Integration Center, PATRIC (http://www.patricbrc.org/portal/portal/patric/Home).

Localizamos 28 secuencias genómicas de *B. thuringiensis* en los diferentes sitios web accesados hasta el pasado octubre de 2011, en donde se incluyen los plásmidos y el cromosoma de cada variante. Utilizamos la secuencia del cromosoma de 18 variantes reportada, incluyendo las secuencias parciales de 15 de ellas, adicionalmente la secuencia de *B. subtilis* fue utilizada como grupo externo, una lista de todos ellos y sus principales características se muestra en la tabla 1. Las secuencias completas hasta el momento que se tienen reportadas pertenecen a las variantes B. *thuringiensis* Al Hakam, *konkukian*. 97-27 y BM171 que se incluyen en el análisis, mientas que las variantes *chinensis* CT43 (Jin *et al.* 2011)y *finitimis* YBT-02 (Zhu *et al.* 2011) ya no pudieron ser incluidas.

Tabla I Principales características de las secuencias de *B. thuringiensis y B subtilis*

No	Nombre				Tamaño (kp)	GC-Contenido (%)	Fuente, sitio de aislamiento	Referencia, comentario	Anotación	
	serovariedad	Cepa (clave)	Serotipo H	No. Acceso					PATRIC CDS	Ref Seq CDS
1	*B. subtilis* subtilis*	*JH642*		NZ_CM000489	4187.62	43.49		Hoch *et al*., 1977	4364	4495
2		*BMB171*		NC_014172	5643.05	35.10	Huazhong Agricultural University Stock, Chine	Mutante sin cristal y alta frecuencia de transformación He et al., 2010	5707	5349
3		*Bt407*		NZ_CM000747	6026.84	34.73	Pasteur Institute, Stock, Francia	Mutante sin cristal	6441	6298
4		*IBL 200*		NZ_CM000758	6731.79	34.41	EUNA, humano similar a *israelensis*	Blackburn *et al*., 2011	6846	6693
5		*IBL 4222*		NZ_CM000759	6612.43	34.64	Gato, similar a *israelensis*		6826	6658
6	*andalousiensis*	*BGSC 4AW1*	*37*	ACNG00000000		34.96	España, Suelo	Quesada-Moraga, *et al*., 2004	5760	5546
7	*Berliner*	*ATCC 10792*	*1*	NZ_CM000753	6260.14	34.68	Alemania, *Ephestia kuehniella*	Smit *et al*., 1952	6427	6243
8	*huazhongensis*	*BGSC 4BD1*	*40*	NZ_CM000756	6231.20	34.49	China	Ziniu, Y. 1993	6190	6019
9	*Israelensis*	*ATCC 35646*	*14*	NZ_AAJM00000000	5880.84	35.00	Israel, agua de desecho		6720	6132
10	*konkukian*	*97-27*	*34*	NC_005957	5237.68	35.41	Yugoslavia, Necrosis humana, 1995	Hernández, *et al*., 1998	5577	5197
11	*kurstaki*	*T03a001*	*3a, 3b, 3c*	NZ_CM000751	5527.57	34.76	Francia, *Ephestia kuehniella,* 1961	Priest, *et al*., 2004	5716	5556
12	*monterrey*	*BGSC 4AJ1*	*28a, 28b*	NZ_CM000752	6489.02	34.52	México, Suelo	Lecadet 1994	6720	6490
13	*pakistani*	*T13001*	*13*	NZ_CM000750	6037.51	34.63	Pakistán, *Cydia pomonella*	Shaik R. 1976	6228	6028
14	*pondicheriensis*	*BGSC 4BA1*	*20a, 20c*	NZ_CM000755	6031.48	34.83	India, Suelo	Rajagopalan PK 1983	6285	6053
15	*pulsiensis*	*BGSC 4CC1*	*65*	NZ_CM000757	6002.60	34.76	Pakistán, campos de granos	IEBC donado por A. Khalique	6209	5944
16	*sotto*	*T04001*	*4a, 4b*	NZ_CM000749	6107.75	34.55	Canadá	Angus T. A. 1970	6852	6583
17	*thuringiensis*	*T01001*	*1*	NZ_CM000748	6323.12	34.67	Canadá, *Ephestia kuehniella, 1958*	Heimpel A. M. in Priest, *et al*., 2004	6475	6323
18	*tochigiensis*	*BGSC 4Y1*	*19*	NZ_CM000746	5625.91	34.77	Japón, Suelo	Ohba, *et al*., 1981	5622	5732
19		Al Hakam		NC_008600	5257.09	35.43	Posible arma biológica, Iraq	Helgason, *et al*., 2000	5523	4798

7.2 Análisis genómico estructural

Desarrollamos un análisis estructural comparativo intra-específico con las secuencias genómicas de B *thuringiensis*, con los siguientes resultados:

7.2.1 Tamaño del genoma

Como se muestra en la tabla 1, los tamaños de genomas reportados para *B thuringiensis*, tienen un rango de 52.3 a 67.3 Mb con un promedio de 59.6 Mb.

7.2.2 Contenido de nucleótidos

El contenido de nucleótidos de cada una de las cepas se muestra en el apéndice del mismo nombre. Una comparación de los promedios de las especies del grupo *B. cereus* se presenta en la figura 1, la información de estas gráficas se puede observar en la tabla 2

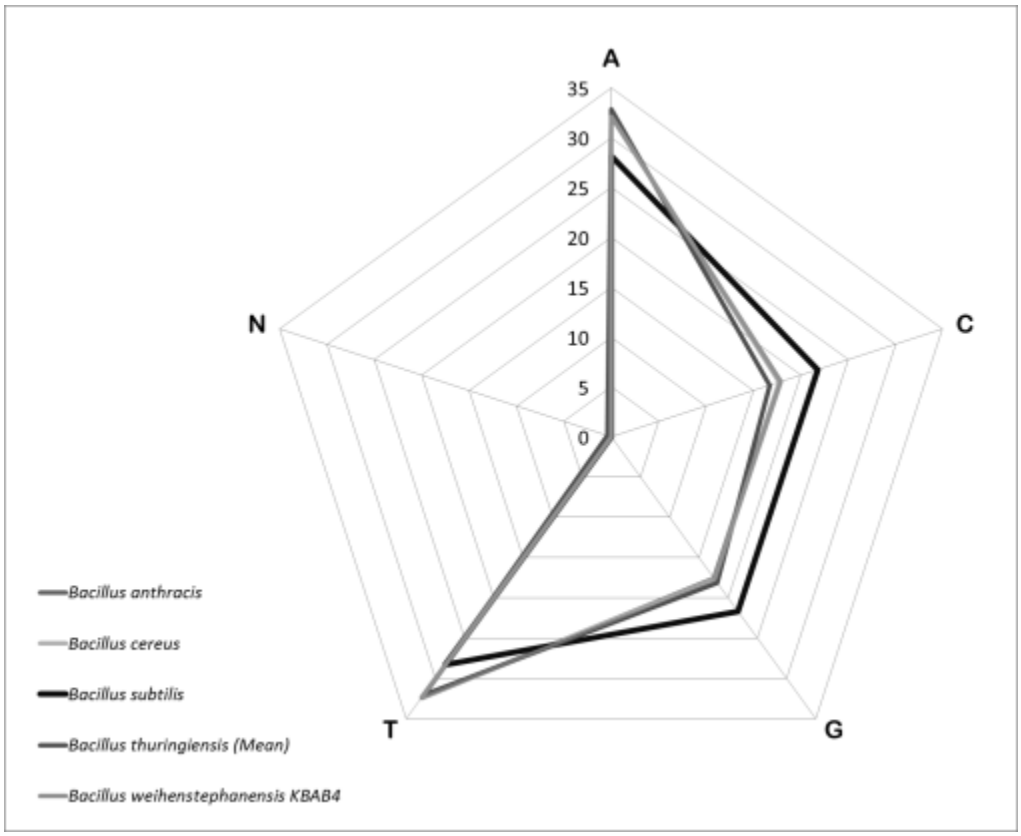


Figura 1 Contenido de Nucleótidos en las cepas del grupo *Bacillus cereus sensu lato* y *Bacillus subtilis*

Tabla II Contenido Promedio de Nucleótidos en las especies del grupo *Bacillus cereus sensu lato*.

No. Access	Nombre	A	C	G	T	N	Porcentaje GC	Porcentaje AT	Tamano cromosomico
AE016879.1	*Bacillus anthracis str. Ames*	32.24	17.79	17.59	32.38	0	35.38	64.62	
AE017334.2	*Bacillus anthracis str. 'Ames Ancestor',*	32.24	17.79	17.59	32.38	0	35.38	64.62	
AE017225.1	*Bacillus anthracis str. Sterne*	32.24	17.79	17.59	32.38	0	35.38	64.62	
	Bacillus anthracis	**32.24**	**17.79**	**17.59**	**32.38**	**0.00**	**35.38**	**64.62**	
CP001177.1	*Bacillus cereus AH187*	32.15	17.86	17.73	32.25	0	35.59	64.41	
CP001283.1	*Bacillus cereus AH820*	32.25	17.77	17.62	32.35	0.01	35.40	64.60	
AE017194.1	*Bacillus cereus ATCC 10987*	32.26	17.76	17.82	32.16	0.01	35.58	64.42	
AE016877.1	*Bacillus cereus ATCC 14579*	32.37	17.62	17.66	32.34	0	35.28	64.72	
CP001176.1	*Bacillus cereus B4264*	32.30	17.69	17.61	32.40	0	35.30	64.70	
CP000001.1	*Bacillus cereus E33L*	32.28	17.75	17.60	32.37	0	35.35	64.65	
CP001186.1	*Bacillus cereus G9842*	32.33	17.69	17.58	32.41	0	35.26	64.74	
CP000764.1	*Bacillus cereus subsp. cytotoxis NVH 391-98*	32.00	18.03	17.85	32.13	0	35.88	64.12	
	Bacillus cereus	**32.24**	**17.78**	**17.67**	**32.32**	**0.00**	**35.44**	**64.56**	
AL009126.2	*Bacillus subtilis*	28.18	21.81	21.71	28.30	0	43.52	56.48	
	Bacillus subtilis subsp. subtilis str. JH642	28.19	21.80	21.69	28.31	0.01	43.49	56.50	4187.62
	Bacillus subtilis	**28.19**	**21.81**	**21.70**	**28.31**	**0.01**	**43.51**	**56.49**	
CP001907	*Bacillus thuringiensis serovar chinensis CT-43*	32.30	17.69	17.69	32.33	0.00	35.38	64.62	
CP002508	*Bacillus thuringiensis serovar finitimus YBT-020*	32.21	17.79	17.75	32.25	0.01	35.54	64.46	
NC_005957	*Bacillus thuringiensis serovar konkukian str. 97-27*	32.23	17.78	17.62	32.36	0.00	35.41	64.59	5237.68
NC_014171	*Bacillus thuringiensis BMB171*	32.29	17.68	17.61	32.42	0.00	35.29	64.71	5643.05
NZ_CM000747	*Bacillus thuringiensis Bt407*	33.17	16.29	18.44	31.79	0.31	34.73	64.96	6026.84
NZ_CM000758	*Bacillus thuringiensis IBL 200*	33.45	15.66	18.74	31.79	0.36	34.41	65.23	6731.79
NZ_CM000759	*Bacillus thuringiensis IBL 4222*	33.18	16.31	18.33	31.60	0.58	34.64	64.78	6612.43
NZ_CM000754	*Bacillus thuringiensis serovar andalousiensis BGSC 4AW1*	32.36	17.50	17.45	32.22	0.47	34.96	64.57	
NZ_CM000753	*Bacillus thuringiensis serovar berliner ATCC 10792*	33.22	16.03	18.65	31.70	0.40	34.68	64.91	6260.14
NZ_CM000756	*Bacillus thuringiensis serovar huazhongensis BGSC 4BD1*	33.64	15.30	19.19	31.42	0.45	34.49	65.06	6231.20
NZ_CM000751	*Bacillus thuringiensis serovar kurstaki str. T03a001*	32.72	16.69	18.07	31.91	0.61	34.72	64.63	5527.57
NZ_CM000752	*Bacillus thuringiensis serovar monterrey BGSC 4AJ1*	**32.68**	**17.20**	**17.32**	**32.41**	**0.39**	**34.52**	**65.09**	**6489.02**
NZ_CM000750	*Bacillus thuringiensis serovar pakistani str. T13001*	32.24	17.56	17.07	32.11	1.02	34.63	64.35	6037.51
NZ_CM000755	*Bacillus thuringiensis serovar pondicheriensis BGSC 4BA1*	33.06	16.56	18.27	31.80	0.31	34.83	64.86	6031.48
NZ_CM000757	*Bacillus thuringiensis serovar pulsiensis BGSC 4CC1*	33.68	15.16	19.60	31.13	0.43	34.76	64.81	6002.60
NZ_CM000749	*Bacillus thuringiensis serovar sotto str. T04001*	33.22	15.78	18.77	31.44	0.79	34.55	64.66	6107.75
NZ_CM000748	*Bacillus thuringiensis serovar thuringiensis str. T01001*	32.90	16.84	17.84	32.04	0.39	34.67	64.94	6323.12
NZ_CM000746	*Bacillus thuringiensis serovar tochigiensis BGSC 4Y1*	32.84	16.59	18.19	31.99	0.39	34.77	64.84	6323.12
NC_008600	*Bacillus thuringiensis str. Al Hakam*	32.30	17.73	17.70	32.27	0	35.43	64.57	5257.09
	Bacillus thuringiensis (Mean)	**32.83**	**16.74**	**18.12**	**31.95**	**0.36**	**34.86**	**64.77**	
CP000903.1	*Bacillus weihenstephanensis KBAB4*	**32.12**	**17.90**	**17.66**	**32.32**	**0**	**35.56**	**64.44**	

A=Adenina,T=Timina, C=Citosina, G=Guanina, N= no identificado
En negro se presentan los promedios por especies.

7.2.3 Porcentaje de Adenina-Timina

El promedio de porcentaje de AT fue de 64.77%, como se puede observar en la tabla 2, adicionalmente una gráfica de barras se muestra en la figura 2.

7.2.4 Porcentaje de Guanina-Citosina

El promedio de porcentaje de GC fue de 34.80%, como se puede observar en las tablas 1 y 2, adicionalmente una gráfica de barras se muestra en la figura 2.

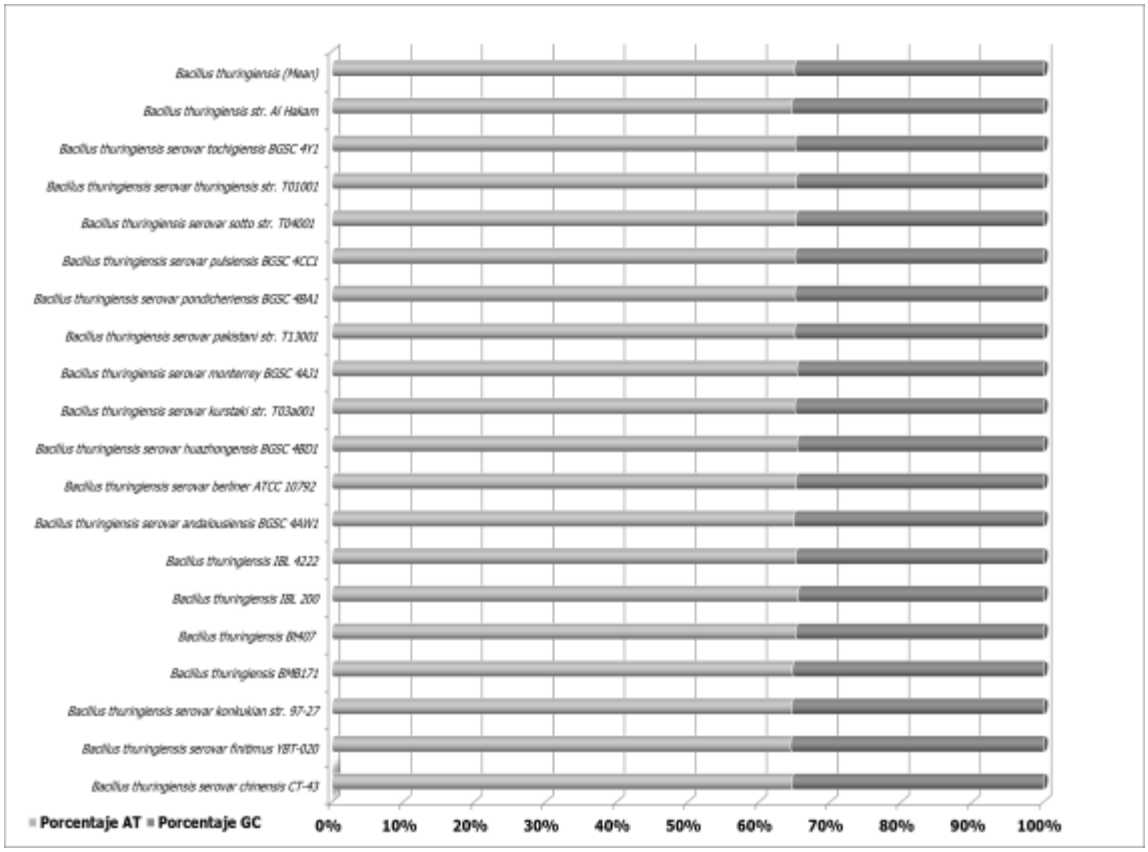


Figura 2 Contenido de AT y GC de los genomas de *Bacillus thuringiensis*, en gris se observan los contenidos de AT, en negro el contenido de GC de los diferentes genomas.

7.2.5 Sintenia

La sintenia se obtuvo de un alineamiento múltiple genómico realizado en Mauve (Darling *et al.* 2008; Darling *et al.* 2010) como puede observarse en la figura 3.

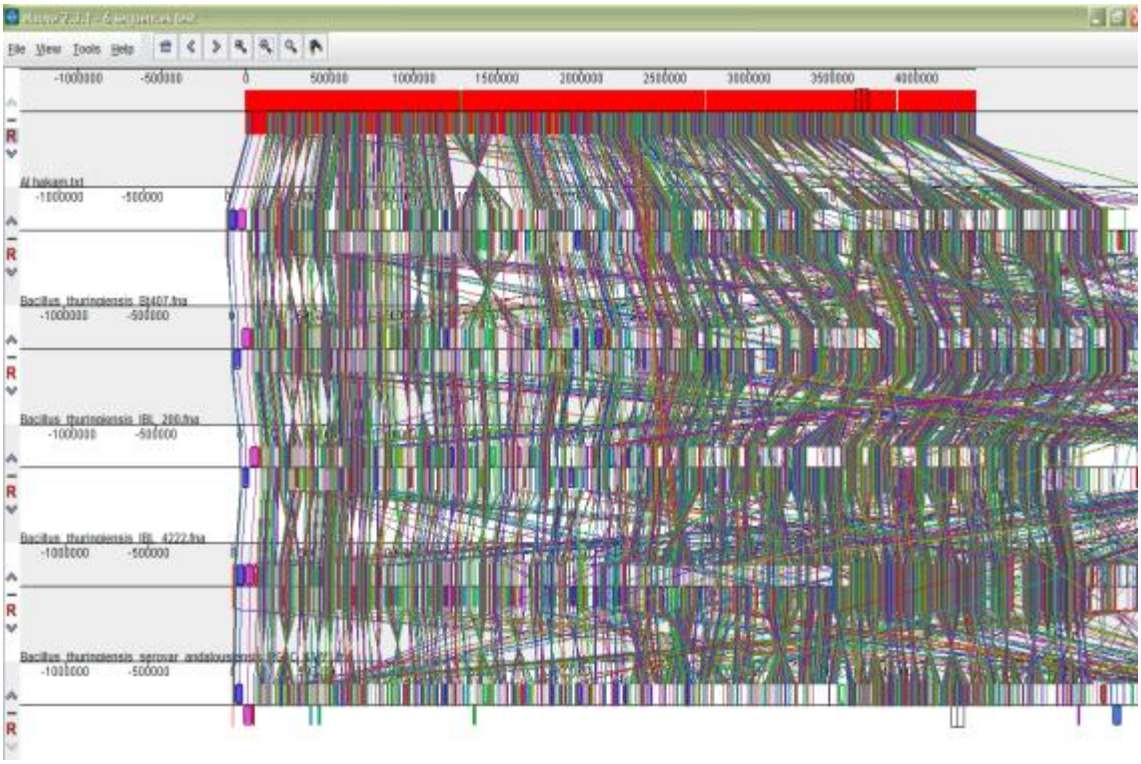


Figura 3 Alineamiento múltiple genómico de las secuencias de *Bacillus thuringiensis*

7.3 Edición de secuencias

Encontramos diferentes entradas duplicada con errores gramaticales, incongruente o diferente anotación para una proteína particular como se muestra en la tabla III. Los errores gramaticales mas frecuentemente encontrados fueron, el uso de la coma, guiones y paréntesis, por ejemplo: *"ABC transporter ATP-binding proteins = ABC transporter (ATP-binding protein), ABC transporter, ATP binding protein o ABC transporter, ATP-binding protei*n"; todas las entradas con este tipo de errores fueron corregidas manualmente y consolidadas en una sola notación.

La asignación incongruente de las entradas como: "*like", "Antigen", "protein", "Similarity", "Transporter", "Complete genome", "function not yet clear", "no significant homology", "distant similarity"*, etc. fueron eliminadas del análisis. Otros ejemplos mas pueden observarse en la tabla IV.

De forma similar los genes con anotación: *hypothetical, putative, unknown, unnamed* o *uncharacterized*, también fueron removidos del análisis.

La edición y actualización de algunos CDS se realizó en función de lo reportado en las bases de datos de referencia: UniProt (http://www.uniprot.org/), Brenda (http://www.brenda-enzymes.org/), Gene home (http://www.ncbi.nlm.nih.gov/gene) propuestos como los mejores referentes al momento. Los resultados de muestran en la tabla V.

La matriz final contiene 3877 entradas por 19 secuencias cromosómicas, 18 de *Bacillus thuringiensis* y una de *B. subtilis* provenientes de diferentes ecosistemas. Todos los marcadores son regiones codificantes y distribuidos con diferentes frecuencia entre las cepas y corresponden a cerca del 36% ± 4.2% del tamaño total de las anotaciones reportadas.

Tabla III Anotaciones que Presentan Errores Sistemáticos del Tipo 2

Anotación Utilizada	**Anotaciones con errors**
2',3'-cyclic-nucleotide 2'-phosphodiesterase	2',3'-cyclic-nucleotide 2'-phosphodiesterase / 5'-nucleotidase
3-oxoacyl-(acyl carrier protein) synthase	3-oxoacyl-[acyl-carrier-protein] synthase
4-carboxymuconolactone decarboxylase	4-carboxymuconolactone decarboxylase domain/alkylhydroperoxidase AhpD family core domain protein
4-Hydroxy-2-oxoglutarate aldolase	4-Hydroxy-2-oxoglutarate aldolase / 2-dehydro-3-deoxyphosphogluconate aldolase
4-oxalocrotonate tautomerase	4-oxalocrotonate tautomerase; Xylose transport system permease protein xylH
ABC transporter ATP-binding protein	ABC transporter (ATP-binding protein) ABC transporter, ATP binding protein ABC transporter, ATP-binding protein
Acetyltransferase	Acetyltransferase, GNAT family acetyltransferase, GNAT family protein Acetyltransferase(EC:2.3.1.-) Acetyltransferase (GNAT family) SAS0976
Acyltransferase	acyltransferase family protein
Aldehyde dehydrogenase	aldehyde dehydrogenase family protein
alkaline serine protease	alkaline serine protease, subtilase family
Alpha/beta hydrolase	alpha/beta hydrolase fold
Amino acid ABC transporter, permease protein	amino acid ABC transproter, permease protein VC0009 [imported]
Amino acid permease	Amino acid permease family protein
aminoglycoside phosphotransferase family protein	aminoglycoside phophotransferase family protein Aminoglycoside phosphotransferase
Ankyrin repeat protein	ankyrin repeat domain protein Ankyrin repeat Ankyrin
Antibiotic biosynthesis monooxygenase	Antibiotic biosynthesis monooxygenase domain-containing protein
Arginyl-tRNA synthetase	Arginyl-tRNA synthetase-related protein
Arsenate reductase	Arsenate reductase family protein
Aspartate racemase	aspartate racemase family protein
ATP/GTP binding protein	ATP/GTP-binding protein, SA1392 homolog
ATPase, histidine kinase-, DNA gyrase B-, and HSP90-like domain protein	ATPase, histidine kinase-, DNA gyrase B-, and HSP90-like domain protein protein
ATP-dependent DNA helicase	ATP-dependent DNA helicase pcrA
ATP-dependent RNA helicase	ATP-dependent RNA helicase BA2475
Bacteriophage	Bacteriophage-related protein
Bacterial luciferase family protein	Bacterial luciferase family protein YtmO, in cluster with L-cystine ABC transporter

Continuación Tabla III

Anotación Utilizada	Anotaciones con errores
Beta-glucosidase	Beta-glucosidase; 6-phospho-beta-glucosidase
CAAX amino terminal protease	CAAX amino terminal protease family
capsular polysaccharide biosynthesis protein	Capsular polysaccharide biosynthsis protein capsular polysaccharide synthesis enzyme
Carbohydrate kinase, PfkB	carbohydrate kinase, PfkB family
Carbonic anhydrase	carbonic anhydrase, family 3 carbonic anhydrase, prokaryotic type, putative
Cell surface protein IsdA1	Cell surface protein IsdA, transfers heme from hemoglobin to apo-IsdC
Cell wall hydrolase	cell wall hydrolase/autolysin
Cell wall surface anchor family protein	Cell wall surface anchor family protein, LPXTG motif
Cephalosporin hydroxylase	cephalosporin hydroxylase family protein
choloylglycine hydrolase collagen-like protein	choloylglycine hydrolase family protein Collagen-like triple helix repeat protein
Cytochrome d ubiquinol oxidase subunit II	cytochrome d ubiquinol oxidase, subunit II-related protein
Dehydrogenase	Dehydrogenases with different specificities (related to short-chain alcohol dehydrogenases)
Diadenosine tetraphosphatase	Diadenosine tetraphosphatase and related serine/threonine protein phosphatases
Dienelactone hydrolase family	Dienelactone hydrolase and related enzymes
DNA integration/recombination/inversion protein	DNA integration/recombination/invertion protein
DNA polymerase III alpha subunit	DNA polymerase III, alpha subunit
DNA polymerase III delta prime subunit	DNA polymerase III, delta prime chain DNA polymerase III delta subunit
DNA replication protein dnaC	DNA replication protein dnaC
DNA-directed RNA polymerase beta subunit	DNA-directed RNA polymerase beta' subunit
ErfK/YbiS/YcfS/YnhG family protein	ErfK/YbiS/YcfS/YnhG superfamily
Esterase	Esterase/lipase
Exodeoxyribonuclease V alpha chain	Exodeoxyribonuclease V alpha chain ## RecD Exodeoxyribonuclease V alpha chain(EC:3.1.11.5)
FkbH	FkbH domain
Gamma-glutamyltranspeptidase	Gamma-glutamyltranspeptidase PgsD/CapD, catalyses PGA anchorage to peptidoglycan
glucosamine--fructose-6-phosphate aminotransferase (isomerizing) (glmS)	Glucosamine--fructose-6-phosphate aminotransferase [isomerizing]
Glutamine ABC transporter, periplasmic glutamine-binding protein (TC 3.A.1.3.2)	Glutamine ABC transporter, periplasmic glutamine-binding protein (TC 3.A.1.3.2) / Glutamine transport system permease protein GlnP (TC 3.A.1.3.2)
glutaredoxin family protein	glutaredoxin-like protein NrdH-related protein
Glycerophosphoryl diester phosphodiesterase	Glycerophosphoryl diester phosphodiesterase family protein

Continuación Tabla III

Anotación Utilizada	Anotaciones con errores
Glycine betaine ABC transport system, permease protein OpuAB	Glycine betaine ABC transport system, permease protein OpuAB / Glycine betaine ABC transport system, glycine betaine-binding protein OpuAC
Glycosyltransferase	glycosyl transferase glycosyl transferase and polysaccharide deacetylase fusion Glycosyl transferase Family 1 Glycosyl transferase, group 1 Glycosyl transferase, group 1 family Glycosyl transferase, group 1 family, anthrose biosynthesis
Glyoxalase/Bleomycin resistance protein/Dioxygenase family protein	Glyoxalase, Glyoxalase/Bleomycin resistance protein/Dioxygenase superfamily Glyoxalase family protein
Glyoxylate reductase / Hydroxypyruvate reductase / 2-ketoaldonate reductase, broad specificity	Glyoxylate reductase / Glyoxylate reductase / Hydroxypyruvate reductase Glyoxylase family protein
GTP-binding protein HflX	GTP-binding protein -to HflX
Histidine kinase	Histidine kinase of the competence regulon ComD
Histidyl-tRNA synthetase	Histidyl-tRNA synthetase, archaeal-type paralog
HIT family hydrolase	HIT family protein
Isochorismatase	Isochorismatase family protein
Late competence protein ComC, processing protease / Leader peptidase (Prepilin peptidase) / N-methyltransferase	Late competence protein ComC, processing protease; Leader peptidase (Prepilin peptidase) / N-methyltransferase
Lipase(EC:3.1.1.3)	Lipase lipase family protein Lipase precursor
Long-chain-fatty-acid--CoA ligase	Long-chain-fatty-acid--CoA ligase / Acetoacetyl-CoA synthetase [leucine] Long-chain-fatty-acid--CoA ligase associated with anthrachelin biosynthesis @ Long-chain-fatty-acid--CoA ligase of siderophore biosynthesis
Macrolide efflux protein	Macrolide-efflux protein
major facilitator family transporter	major facilitator family transporter; possible multidrug efflux pump
Malonyl CoA-acyl carrier protein transacylase	Malonyl CoA-acyl carrier protein transacylase; Enoyl-[acyl-carrier-protein] reductase [FMN]
Maltose phosphorylase	Maltose phosphorylase / Trehalose phosphorylase
Maltose/maltodextrin transport ATP-binding protein MalK	Maltose/maltodextrin transport ATP-binding protein MalK; Multiple sugar ABC transporter, ATP-binding protein
mandelate racemase/muconate lactonizing enzyme(EC:5.1.2.2)	Mandelate racemase/muconate lactonizing enzyme family protein
metal-dependent hydrolase	Metal-dependent hydrolase COG0491 with rhodanese-homology domain (RHOD)

Continuación Tabla III

Anotación Utilizada	Anotaciones con errores
Metallo-beta-lactamase family protein	Metallo-beta-lactamase superfamily protein PA0057
Methionyl-tRNA synthetase	Methionyl-tRNA synthetase, clostridial paralog
Methylmalonate-semialdehyde dehydrogenase	Methylmalonate-semialdehyde dehydrogenase [inositol]
Methyltransferase(EC:2.1.1.-)	Methyltransferase methyltransferase domain protein methyltransferase type 11
mobilization protein	Mob
Molybdenum cofactor biosynthesis protein MoaD	Molybdenum cofactor biosynthesis protein MoaD; Molybdopterin converting factor subunit 1
Molybdopterin biosynthesis MoeB protein	Molybdopterin biosynthesis protein MoeB
insecticidal protein	mosquito toxic crystal protein-like protein mosquitocidal protein MOSQUITOCIDAL TOXIN PROTEIN Mosquitocidal toxin
Multimodular transpeptidase-transglycosylase	Multimodular transpeptidase-transglycosylase / Penicillin-binding protein 1A/1B (PBP1)
N-acetylmuramoyl-L-alanine amidase / S-layer protein(EC:3.5.1.28)	N-acetylmuramoyl-L-alanine amidase
nlpC/P60 family protein	NLP/P60 family protein
Non-ribosomal peptide synthase:Amino acid adenylation	Non-ribosomal peptide synthase
Nucleoside-diphosphate-sugar epimerase	Nucleoside-diphosphate-sugar epimerases
Ornithine cyclodeaminase / Siderophore staphylobactin biosynthesis protein SbnB	Ornithine cyclodeaminase
PBS lyase HEAT-like repeat	PBS lyase HEAT-like repeat domain protein
peptidase S8 and S53 subtilisin kexin sedolisin	peptidase S8 and S53, subtilisin, kexin, sedolisin
Peptidase, M23/M37 family	peptidase, M23/M37 family protein
Peptide chain release factor 2; programmed frameshift-containing	peptide chain release factor 2
Phage protein	phage protein-related protein phage related protein Phage related protein, YorS B.subtilis homolog
Phage T7 exclusion protein	Phage T7 exclusion protein associated hypothetical protein
Phage tail fiber protein	Phage tail fibers
Phage terminase large subunit	Phage terminase, large subunit Phage terminase, large subunit ,
Phage terminase small subunit	Phage terminase, small subunit
Phage transcriptional activator ArpU	phage transcriptional regulator, ArpU family Phage transcriptional regulator, ArpU family subfamily
Phage-related protein	phage-like fragment phage-like protein
Phenylalanyl-tRNA synthetase alpha subunit	Phenylalanyl-tRNA synthetase alpha chain

Continuación Tabla III

Anotación Utilizada	Anotaciones con errores
Phosphohydrolase, MutT/Nudix family	Phosphohydrolase (MutT/nudix family protein)
Prolipoprotein diacylglyceryl transferase	prolipoprotein diacylglyceryl transferase family protein
prophage LambdaBa04, DNA binding protein	prophage LambdaBa04, DNA binding protein, putative prophage LambdaBa04, DNA-binding protein
Pyridine nucleotide-disulfide oxidoreductase; NADH dehydrogenase	pyridine nucleotide-disulphide oxidoreductase family protein
Radical SAM domain heme biosynthesis protein	radical SAM domain protein
recombination protein U (penicillin-binding protein-related factor A)	Recombination protein U
Replication termination protein	replication terminator protein
reverse transcriptase	Retron-type reverse transcriptase
Riboflavin kinase	Riboflavin kinase / FMN adenylyltransferase
Ribose-phosphate pyrophosphokinase(EC:2.7.6.1)	Ribose-phosphate pyrophosphokinase
Ribosomal RNA large subunit methyltransferase N ## LSU rRNA m2A2503	Ribosomal RNA large subunit methyltransferase N
RNA polymerase sigma factor (sigma-A) (sigma-43)	RNA polymerase sigma factor
RNA polymerase sigma-70 factor	RNA polymerase sigma-70 factor, ECF subfamily
Rubredoxin-NAD(+) reductase(EC:1.18.1.1)	Rubredoxin
Saccharopine dehydrogenase [NADP+, L-lysine forming]	Saccharopine dehydrogenase
sensory box sigma-54 dependent DNA-binding response regulator	sensory box sigma-54 dependent DNA-binding response regulator, in GABA cluster
Serine/threonine protein kinase(EC:2.7.1.37)	serine/threonine kinase Serine/threonine protein kinase Serine/threonine protein kinases
sodium/hydrogen exchanger family protein/TrkA domain protein	Sodium/hydrogen exchanger
Soluble lytic murein transglycosylase / N-acetylmuramoyl-L-alanine amidase(EC:3.2.1.-,EC:3.5.1.28)	soluble lytic murein transglycosylase
Sporulation protein YtfJ	Sporulation protein, YTFJ Bacillus subtilis ortholog
thioesterase(EC:3.1.2.-)	Thioesterase Thioesterase domains of type I polyketide synthases or non-ribosomal peptide synthetases thioesterase family protein
Transcriptional activator of acetoin dehydrogenase operon AcoR	transcriptional activator
Transcriptional pleiotropic regulator	transcriptional pleiotropic regulator of transition state genes
transcriptional regulator, AraC family(EC:2.1.1.63)	Transcriptional regulator, AraC family Transcriptional regulator, AraC/XylS family
Transcriptional regulator, GntR family domain / Aspartate aminotransferase	Transcriptional regulator, GntR family

Continuación Tabla III

Anotación Utilizada	Anotaciones con errores
Transcriptional regulator, MarR family; Cinnamoyl ester hydrolase	Transcriptional regulator, MarR family
Transcriptional repressor pagR	Transcriptional repressor PagR,
Transposase, IS204/IS1001/IS1096/IS1165	transposase, IS204/IS1001/IS1096/IS1165 family protein
Transposase, IS605 OrfB	transposase, IS605 family, OrfB
Two component system histidine kinase(EC:2.7.3.-)	Two component sensor histidine kinase Two component system histidine kinase Two component system sensor histidine kinase CiaH
Two-component response regulator	two-component response regulator / ;
Zn-dependent hydroxyacylglutathione hydrolase / Polysulfide binding protein	Zn-dependent hydroxyacylglutathione hydrolase
Wall-associated protein	Wall-associated protein precursor
RNA-directed DNA polymerase (Reverse transcriptase)	Retron-type reverse transcriptase reverse transcriptase
subtilosin A	Subtilisin precursor

Tabla IV Ejemplos de Anotaciones Incongruentes

Anotaciones Incongruentes o Carentes de Significado Biológico.	
Hypothetical	Uncharacterized
Unknown	Unnamed
Putative	Similarity with
putative	Similar to
Predicted	Probably
Probable	Potential
Possibly	Possible
Not a proline racemase	No significant homology
NA	Function not yet clear
Distant similarity	Complete genome
Related ?	Domain-containing
Conserved domain	Homolog
Ortholog	like
till included	

Tabla V Algunas de las Anotaciones Actualizadas

Nombre Actual	Nombre Anterior
Amino acid permease	Amino acid permease family protein
Aminoglycoside N3'-acetyltransferase	Aminoglycoside N3-acetyltransferase
ATP synthase alpha chain	ATP synthase A chain
ATP synthase beta chain	ATP synthase B chain
Phage-like element PBSX protein xkdD	**xkdD** ykxD
Uncharacterized protein yaaB	**yaaB**
RutC family protein yabJ	YabJ, a purine regulatory protein and member of the highly conserved YjgF family
Uncharacterized protein ybeF	**ybeF**
Uncharacterized protein ybfF	**ybfF**
Uncharacterized protein ybfJ	**ybfJ**
Uncharacterized membrane protein ybgB	**ybgB**
Uncharacterized protein ybyB	**ybyB**
Uncharacterized ABC transporter ATP-binding protein YcbN	**ycbN**
Uncharacterized protein yckD	**yckD**
Spore germination lipase lipC	YcsK **lipC**
Uncharacterized protein yczC	YczC
Uncharacterized protein ydaL	YdaL protein
Uncharacterized protein ydbA	YdbA
Uncharacterized protein yddB	YDDB protein
UPF0702 transmembrane protein ydfR	YDFR protein
UPF0702 transmembrane protein ydfS	YdfS
Uncharacterized protein ydjM	YdjM PSPA13 yzvA
Uncharacterized protein ydjN	YdjN
Uncharacterized protein ydjO	YdjO
Uncharacterized membrane protein ydzJ	ydzJ
Uncharacterized transporter yeaB	YeaB (Cation efflux protein) yeaB ydxT
UPF0720 protein yeeF	YeeF

Continuación Tabla V

Nombre Actual	Nombre Anterior
Uncharacterized protein yesL	YESL protein **yesL**
Uncharacterized protein yesV	YESV protein yesV
Uncharacterized protein yfaA	YfaA
Uncharacterized protein yfhS	YfhS protein
Putative membrane-bound acyltransferase YfiQ	YfiQ
Phosphatidylglycerol lysyltransferase EC 2.3.2.3	mprF yfiW YfiX Lysylphosphatidylglycerol synthase
General stress protein 17M	YflT
Uncharacterized protein yfmB	YfmB
Uncharacterized N-acetyltransferase YfmK	YfmK
Uncharacterized protein yhaZ	YhaZ
Uncharacterized HTH-type transcriptional regulator yhbI	YhbI
Probable anti-sigma-M factor yhdK	YhdK
Uncharacterized protein yhjA	YhjA
Uncharacterized membrane protein yhjC	YhjC
Uncharacterized protein yhjD	YhjD
Uncharacterized HTH-type transcriptional regulator yhjH	YhjH
Uncharacterized MFS-type transporter yhjO	YhjO
Uncharacterized protein yjfB	YjfB
Uncharacterized membrane protein ykoI	YKOI
Uncharacterized protein ykoJ	YKOJ
Uncharacterized N-acetyltransferase YkwB	YkwB
Uncharacterized protein ykzH	YkzH
UPF0749 protein ylxX	YLXX protein
Uncharacterized protein ymaD	YMAD protein ymaD
Uncharacterized protein ymcA	YMCA protein ymcA
UPF0714 protein yndL	YndL

Continuación Tabla V

Nombre Actual	Nombre Anterior
Glucuronoxylanase xynC EC 3.2.1.136	YnfF xynC Endoxylanase xynC Glucuronoxylan xylanohydrolase
Uncharacterized membrane protein yngA	YngA
UPF0713 protein yngL	YngL
Uncharacterized protein yoaF	YoaF
Uncharacterized protein yoaO	YoaO
Uncharacterized protein yoaW	YoaW
Uncharacterized protein yobA	YobA
General stress protein 16O	YocK
Uncharacterized protein yoeB	YoeB
SPBc2 prophage-derived uncharacterized protein yokH	YokH
Uncharacterized protein yokU	YokU
SPBc2 prophage-derived uncharacterized protein yolD	YolD
Aldose 1-epimerase EC 5.1.3.3	YoxA galM Galactose mutarotase
Uncharacterized protein yozN	YozN
Uncharacterized protein yozQ	YozQ
Uncharacterized protein yptA	YptA
Spore germination protein-like protein ypzD	YpzD
Uncharacterized protein yqbG	YqbG
Uncharacterized protein yqeD	YqeD
Uncharacterized protein yqgW	YqgW
Uncharacterized protein YqgZ	YqgZ
Uncharacterized protein yqjF	YqjF
Uncharacterized protein yqjY	YqjY
Uncharacterized protein yqkB	YqkB
Uncharacterized protein yqzC	YQZC protein
Uncharacterized protein yqzG	YqzG
Uncharacterized protein yqzJ	YqzJ
Spore coat protein F-like protein YraD	YraD
Uncharacterized protein yrhK	YrhK

Continuación Tabla V

Nombre Actual	Nombre Anterior
Putative serine/threonine-protein kinase yrzF EC 2.7.11.1	YrzF
Uncharacterized protein yrzI	YrzI
Uncharacterized protein yrzK	YrzK
Uncharacterized protein yteU	YTEU
Putative potassium channel protein yugO	YugO
Uncharacterized protein yurZ	YURZ protein
Putative disulfide oxidoreductase yuzD	YuzD-like protein
Uncharacterized FAD-linked oxidoreductase yvdP	YvdP
Uncharacterized protein yvkN	YvkN
Uncharacterized protein yvmC	YVMC
Uncharacterized protein ywaF	YwaF
Uncharacterized protein ywbO	YwbO
Uncharacterized protein ywhL	YwhL
Uncharacterized beta-barrel protein ywiB	YwiB
Uncharacterized protein ywjC	YwjC
Uncharacterized membrane protein ywnJ	YwnJ
Uncharacterized protein ywqH	YwqH
Cell wall-binding protein ywsB	YwsB
Uncharacterized protein yxaI	YxaI
Uncharacterized protein yxeG	YxeG IP1A
Uncharacterized protein YxeH	YxeH IP1B
Uncharacterized protein yxiC	YxiC J3C
Uncharacterized lipoprotein yybP	YybP
Uncharacterized protein yycE	YycE
Two-component system yycF/yycG regulatory protein yycH	YycH protein
Uncharacterized N-acetyltransferase YycN	YycN
Zwittermicin A resistance protein ZmaR	zwittermicin A-resistance protein

7.4 Codificación de CDS de cada secuencia genómica para realizar un análisis numérico

En este estudio, codificamos nuestros datos en los dos sistemas anteriormente mencionados en el apartado de método.

7.5 Matrices de Datos

Cada matriz de datos contienen 3877 loci por 19 genomas, la primera matriz de datos, se generó transformando los CDS como caracteres binarios; en el caso de la presencia del CDS en una secuencia particular se asigno el valor de (1), en caso de ausencia se asigno (0). La segunda matriz, adicionalmente tomo en cuenta el contenido del total de copias de cada CDS presentes en la secuencia dada, estos fueron transformados en frecuencias relativas para su análisis y presentados en una matriz multiestado.

7.6 Variabilidad genómica

Medimos la variabilidad genómica en diferentes formas, el contenido de las proteínas codificadas en el genoma se representa en la figura 4.

7.6.1 Contenido de CDS y tamaño del genoma.

La correlación del contenido de proteínas y tamaño del genoma se observa en la figura 4. En esta figura se pueden observar tres grupos, el primero contiene los genomas más densos IBL-200, IBL-4222, T04001, BGSC-4AJ1 y ATCC-35646 contienen 580 a 720 más proteínas. EL grupo intermedio contienen en promedio ~6000 proteínas. El grupo de los genomas menos densos se observa en las cepas BGSC-4AW1, T03a001, BMB171, BGSC-4Y1 97-27 y Al Hakam.

7.6.2 Matriz de datos Binarios.

El contenido de loci variables vs los loci exclusivos es apreciado en la figura 5, la relación entre estas dos características establece una tendencia entre los genomas que contienen mayor cantidad de loci variables también poseen mayor cantidad de loci exclusivos, con excepción de las cepas T01001 y ATCC-10792 que presentan altos contenidos de loci variable y casi ningún loci exclusivo, adicionalmente se agrupan con posiciones cercanas en la gráfica.

7.6.3 Matriz de datos Multiestado.

La relación entre la Heterozigocidad vs loci polimórficos se observa en la figura 6, en esta no se observa ninguna correlación entre estos dos atributos, sin embargo, se observa una desviación a la izquierda.

7.6.4 Roles funcionales.

En cuanto a los roles funcionales, los genes que codifican para proteínas que censan y responden a los cambios ambientales fueron las que mostraron mayor variación. El segundo grupo se conformó con los genes que portan la información para los elementos móviles, que presentan una alta variabilidad todo ello se observa en la tabla VI

7.6.5 Índice de Shannon.

Los datos codificados como variables binarias presentó los valores mas altos de variabilidad (promedio 7.76 ±0.03) en comparación de los obtenidos con la matriz multiestado (promedio 6.71 ±0.02). La distribución de la variabilidad presenta una distribución normal, mientras que en los datos mutiestado se observa una distribución tri-modal que se observan en la figura 7.

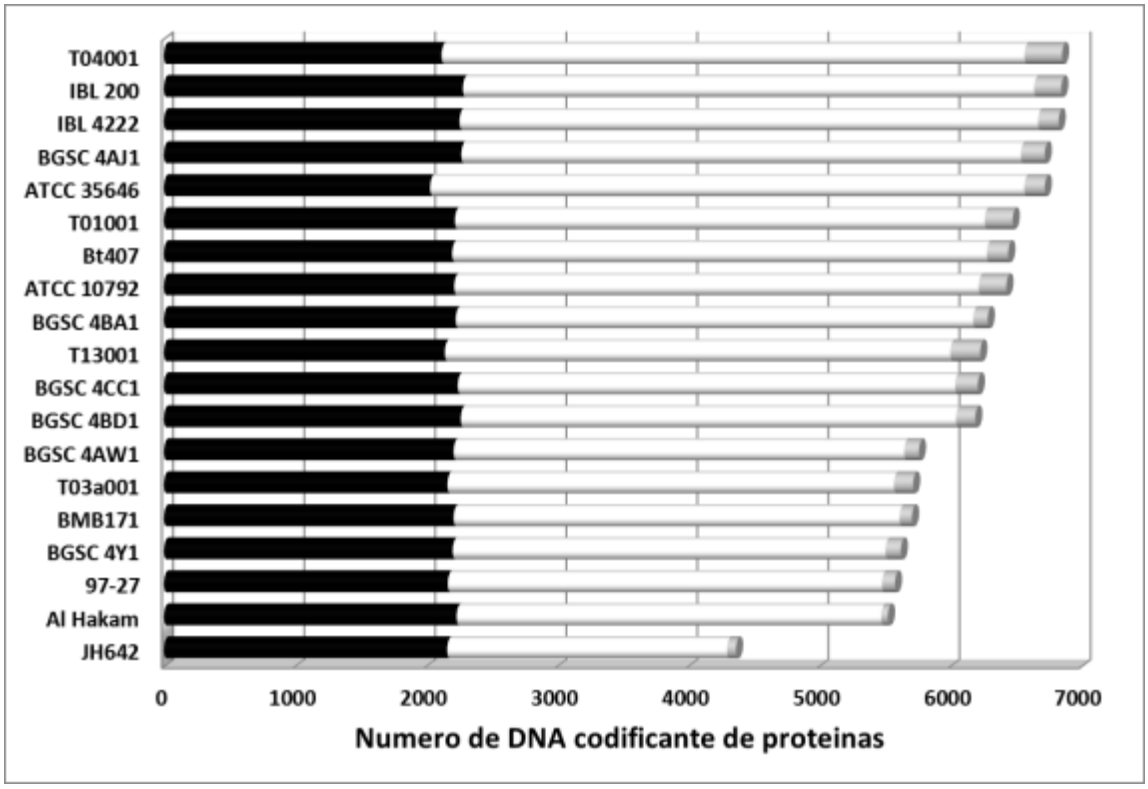


Figura 4 Variación en el contenido de proteínas codificadas en los genomas de *Bacillus thuringiensis* y Correlación en el contenido de CDS y tamaño del genoma. En negro se observan los CDS utilizados en los posteriores análisis (~36%). En blanco los CDS eliminados y en gris los elementos identificados con origen exógeno.

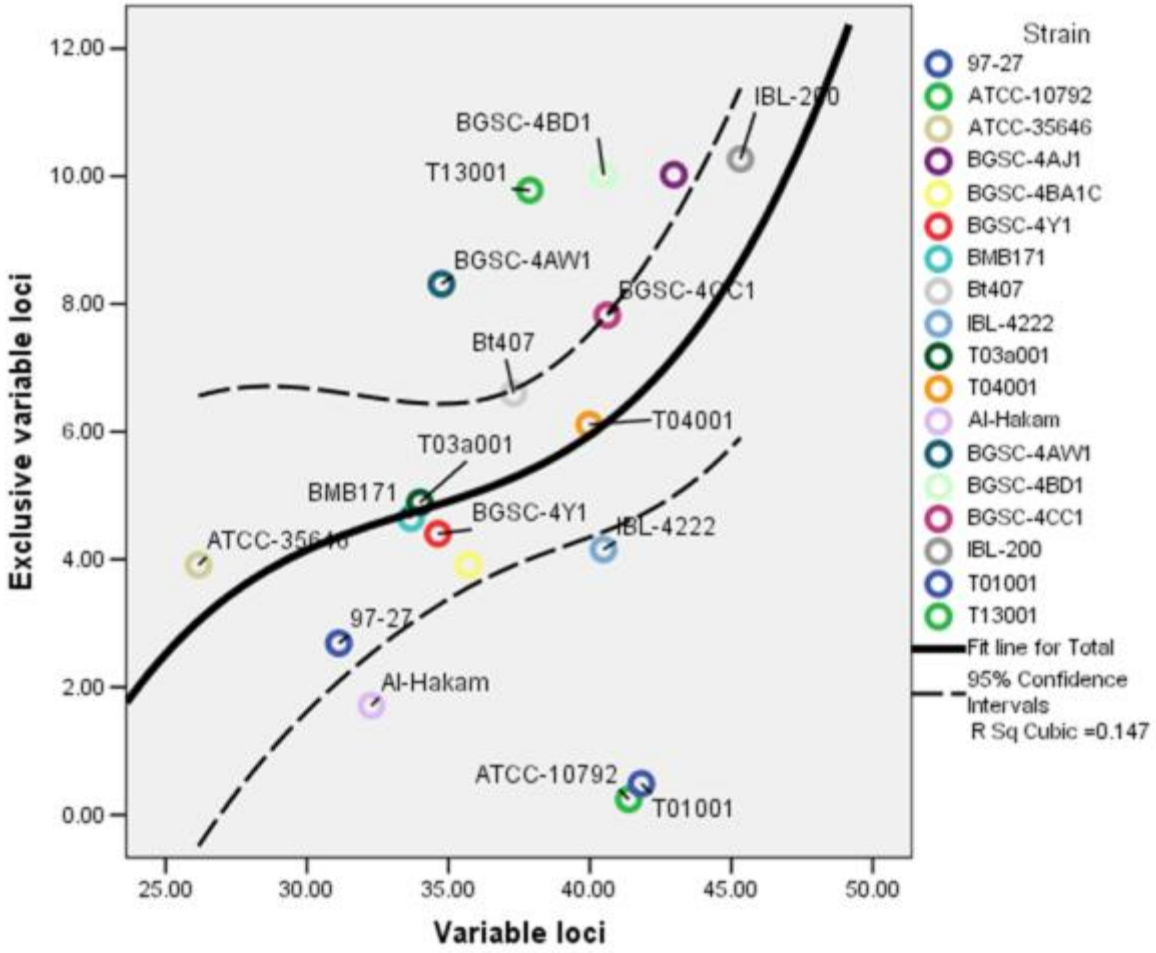


Figura 5 Correlación entre el numero de loci variables frente a la cantidad de loci exclusivos por cepa de *Bacillus thuringiensis*

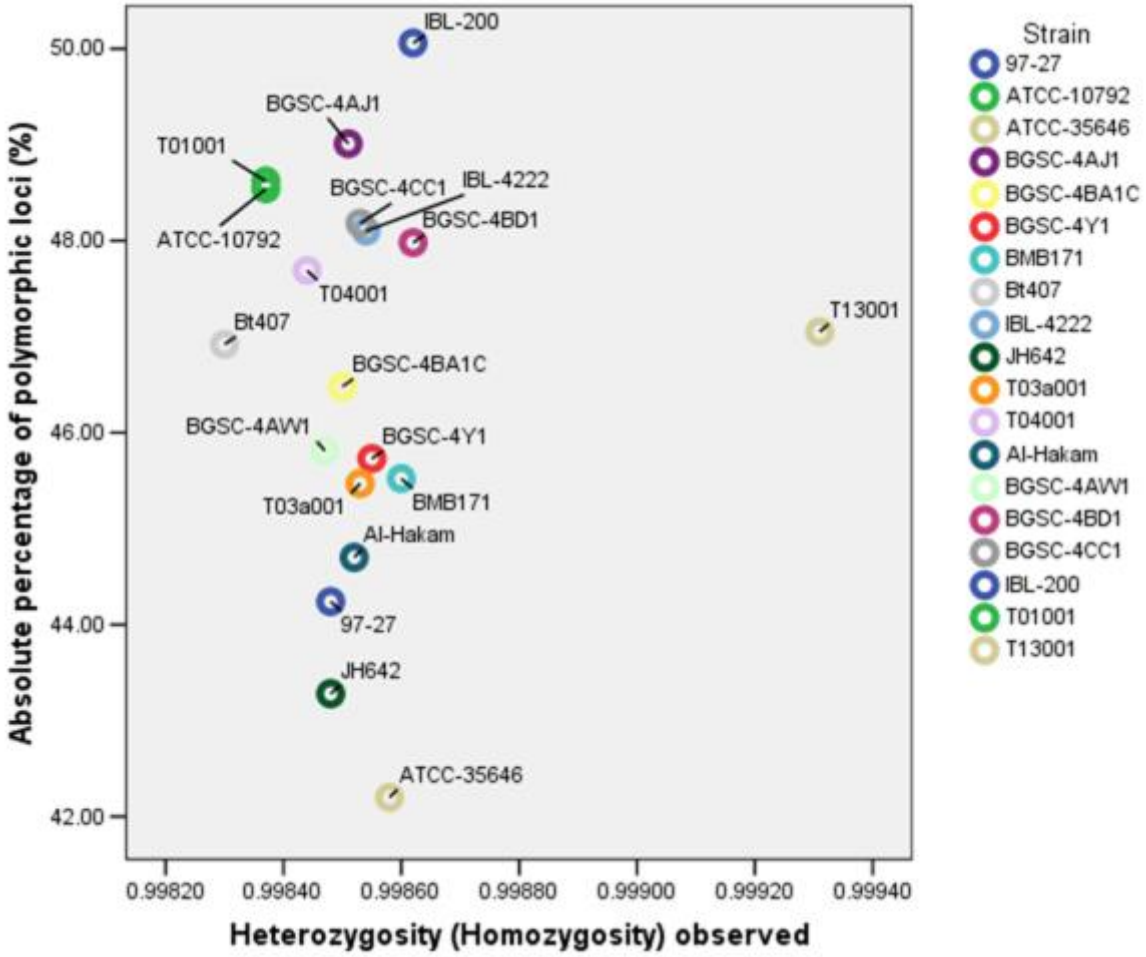


Figura 6 Correlación entre la heterozigocidad observada y el numero absoluto de porcentaje de polimorfismo entre loci de las diferencias variantes de *Bacillus thuringiensis*

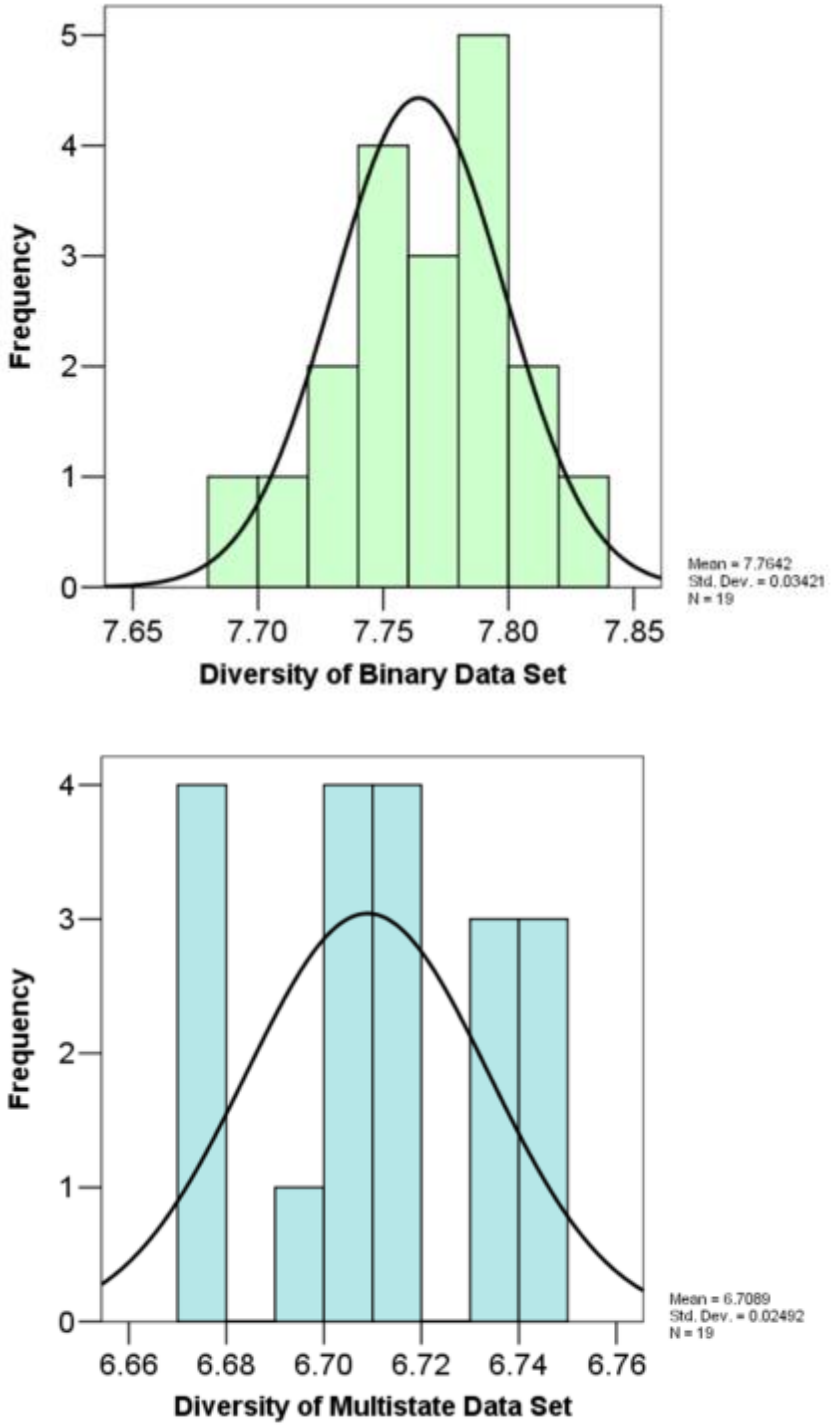


Figura 7 Diversidad de las cepas de *Bacillus thuringiensis*, calculadas con el Índice de Shannon. En verde los datos codificados como binarios, en azul cuando se codifican como datos multiestado

7.8 Estructura genómica poblacional

Probamos la correlación entre dos grupos independientes de datos medidos como distancias pareadas y similitud entre las cepas.

En ambas agrupaciones se observó una clara discriminación de *B subtilis* que fue utilizado como grupo externo y una agregación monofilética (figuras 8 y 9). Utilizamos un punto de corte arbitrario para poder visualizar la presencia de tres grupos denominados A, B1 y B2.

El grupo A fue consistentemente formado por las cepas IBL-4222 y ATCC-3564, serológicamente identificadas como *israelensis*.

El grupo B contiene la mayoría de las cepas analizadas. En la agrupación por Jaccard, el dendograma presenta dos grupos con un número equilibrado de siete y nueve cepas. Por otro lado la representación generada por Manhattan resuelve el grupo B1 con las cepas IBL-200 and T04001, mientras que el grupo B2 contienen las restantes 14 cepas.

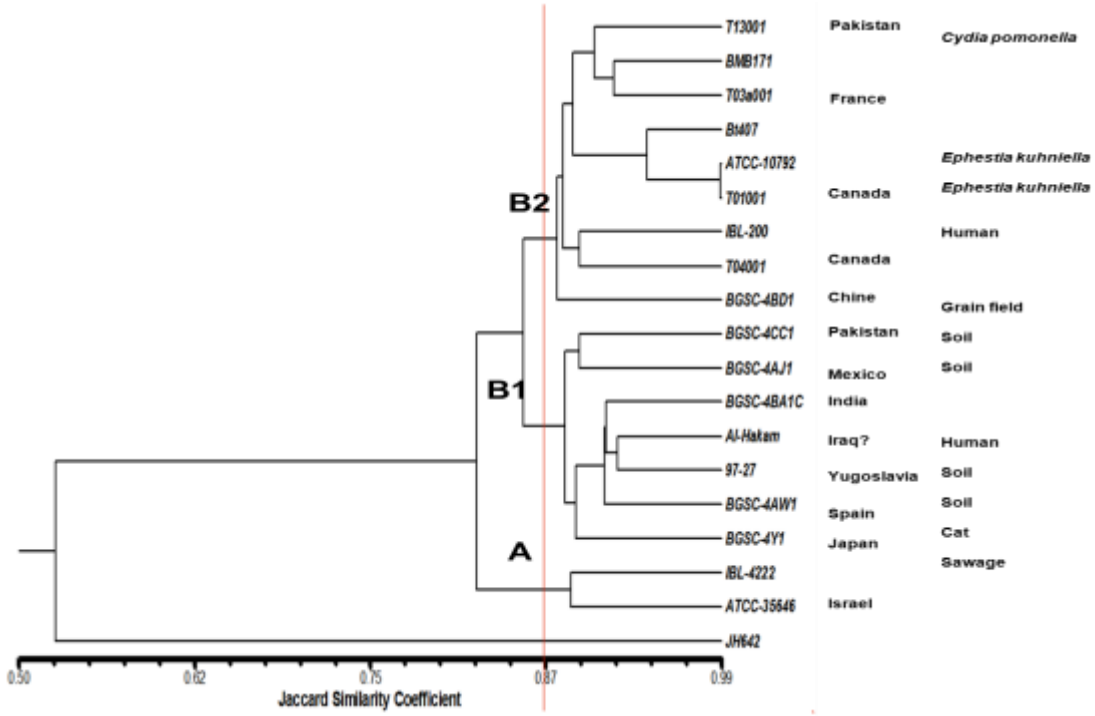


Figura 8 Dendograma producido utilizando el coeficiente de Similitud de Jaccard, agrupado utilizando UPGMA basado en 3877 diferentes loci codificados como datos binarios. *Bacillus subtillis* JH642 se utilizó como grupo externo. La línea clara representa un punto de corte arbitrario.

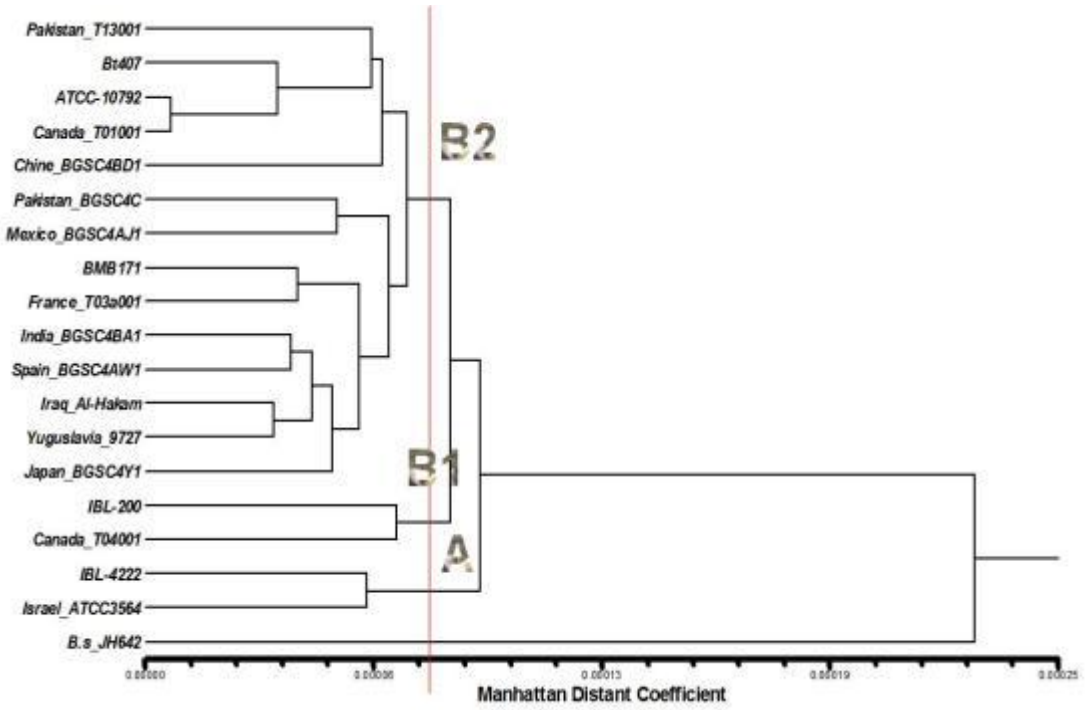


Figura 9 Dendograma producido utilizando el coeficiente de Manhattan, agrupado utilizando UPGMA basado en 3877 diferentes loci codificados como datos multiestado. *Bacillus subtillis* JH642 se utilizó como grupo externo. La línea clara representa un punto de corte arbitrario

8 DISCUSION

Para obtener información de interés biotecnológico, realizamos una comparación sistemática de alto desempeño utilizando los perfiles completos de proteínas codificadas en los cromosomas de 3 genomas completos y 15 genomas parciales (contig) de las diferentes variedades de *B thuringiensis*. Desarrollamos un análisis fenético de la relación existente en una gran cantidad de datos multiestado (3877 x 19) y los comparamos con aquellos que se generan en una codificación como datos binarios (3877 x 19). Adicionalmente, probamos si la forma de codificación de los datos influye en la cuantificación de la diversidad y la representación de la estructura genética poblacional de la especie.

Esta aproximación utilizando el contenido completo de información del cromosomas de Bacillus *thuringiensis* es una herramienta muy poderosa para entender la variación global en la especie. En particular, este análisis presenta la pérdida de algunas proteínas entre las cepas y la variabilidad en el número de copias de genes identificados y anotados.

Como es conocido, el proyecto de secuenciación de las variedades de *B thuringiensis* esta revelando el complejo mundo de esta bacteria, Algunas veces catalogada como entomopatógenica, otras como simbionte de plantas e invertebrados y que ocasionalmente se convierte en patógeno para el humano. Ha visto mermada su historia, enfocada a ciertos fenotipos que pueden no ser parte de su historia natural, como es sabido el fenotipo entomopatogeno es generado por la aportación de algunos plásmidos y poca o nula importancia se ha dado al rol que juegan los genes cromosomales en la adaptación a hospederos o en la patogénesis y menos relevante ha resultado el sentido ecológico-evolutivo de la historia adaptativa a un nicho biótico. En nuestro parecer, los constituyentes genético-cromosomales de *B. thuringiensis* puede auxiliar a responder varias preguntas en relación a los procesos de enfermedad, así como también, pensamos que se pueden vislumbrar proteínas con un alto potencial a la biotecnología, entre sus mas de 3000 proteínas codificadas y anotadas, pueden existir actividades diferentes o novicias que le permiten su compleja existencia.

8.1 Obtención de Secuencias

La obtención de las secuencias, parte principal de este trabajo, también fue la parte limitante del mismo. Realizamos la obtención de las secuencias principalmente en dos fechas, una al inicio de este trabajo, a finales de 2007, en donde verificamos la disponibilidad y el potencial de éxito de obtener una mayor cantidad de secuencias con la presencia de diferentes proyectos de secuenciación con un interesante numero de variedades de *B. thuringiensis*. El segundo evento, tuvo lugar en el 2010, cuando el Naval Medical Research Center liberó las secuencias en fase de *contig* de las ofrecidas variedades. Por otro lado, la secuencia completa de la variedad BMB171 también fue liberada y reportada (He *et al.* 2010).

En total trabajamos con 18 secuencias de las cuales solo tres están completas al momento de escribir este documento, las restantes actualmente siguen en fase de borrador. Incluimos la secuencia de *B. subtilis* como grupo externo, debido que es la especie mejor anotada con un total de 53% de genes codificantes de proteínas identificadas (Kunst *et al.* 1997). Siendo esta un área con gran interés y desarrollo, en la ultima revisión de las bases de datos encontramos dos nuevas secuencias genómicas completas que pertenecen a las variantes *chinensis* CT43 (Jin *et al.* 2011) y *finitimis* YBT-02 (Zhu *et al.* 2011) que por su tardía liberación y la poca información sobre su secuenciación, ya no pudieron ser incluidas.

No es de sorprender que en los próximos meses o semanas el número de secuencias genómicas completas se incremente de forma impresionante, contribuyendo de forma intermitente al conocimiento de *B. thuringiensis* y al grupo Bacillus en general. Por otra parte, seria de mucha utilidad conocer el origen geográfico de las variantes, dato que falta en la mayoría de las cepas. La carencia de esta información evita la oportunidad de probar hipótesis de corte ecológico en la especie.

8.2 Análisis genómico estructural

La función de la biología comparativa es analizar y capturar los patrones bióticos y elaborar una teoría del proceso que pueda explicar el patrón. El patrón son los aspectos ordinales de la vida, mientras que los procesos se entienden como los mecanismos que generan esos patrones (Eldredge *et al.* 1980). Aproximaciones genómico comparativas que son basadas en la secuencia genómica completa proporciona datos más comprensivos de la diversidad de la especie y evita la desviaciones de los métodos tales como MLTS que se realizan solo en pequeñas porciones del genoma (Read *et al.* 2002), Sin embargo, se requiere de una buena calidad en las secuencias para obtener un resultado aceptable. Utilizando el análisis comparativo de secuencias genómicas, varias características generales han sido definidas en otras especies, de las cuales nuestro modelo parece seguir el mismo comportamiento en algunas de ellas, como se describe a continuación:

8.2.1 Tamaño del genoma

El tamaño promedio de 59.6 Mb de las diferentes variedades de *B. thuringienesis* no presenta ninguna diferencia estadísticamente significativa entre ellas (Tabla 1). En comparación con *B. cereus* 5.4 Mb, *B. anthracis* 5.3 Mb y *B. subtilis* 4.2 Mb (Anderson *et al.* 2005) se presentan como los tamaños mas grandes reportados en este grupo en promedio.

La variación en el tamaño genómico es común entre las diferentes poblaciones o variantes bacterianas. En el análisis de 61 genomas de *Escherichia coli* y *Shigella* se ejemplifico esta variabilidad con algunos genomas con apenas 4.56 Mb hasta los que contienen 5.70 Mb, sin embargo en nuestro análisis, el genoma con mayor tamaño es una variante que se analizó en fase de *contig*, por lo que se presume que su tamaño puede estar sobrestimado (Lukjancenko *et al.* 2010). Sin que esto menoscabe la variación en el tamaño de esta comunidad bacteriana.

Actualmente, es ampliamente aceptado que el tamaño de los genomas bacterianos depende del nicho biótico de la especie (Giovannoni *et al.* 2005). En general, los genomas bacterianos pequeños tiende a ser de organismos con nichos estrictos y estables como los que están en asociación con un hospedero, es decir en simbiosis. Por otro lado se piensa que los genomas bacterianos grandes tienden a ocupar nichos complejos y medios variables como el suelo (Bentley *et al.* 2004). Muestras de estos eventos pueden verse claramente en los genomas de *Rickettsia sps*. que presenta un tamaño de apenas 1.1 Mb (Andersson *et al.* 1999) o *Helicobacter pylori* de ~1.6 Mb (Tomb *et al.* 1997) ambas especies son simbiontes estrictos de difícil propagación en medios artificiales lo que marca una clara codependencia evolutiva al hospedero. Mientras que en el otro extremo se pueden encontrar especies tales como *Agrobacterium tumefaciens* C58 que tiene un tamaño genómico de 5.76 Mb (Wood *et al.* 2001) o en otros habitantes regulares de suelo como *Streptomyces coelicolor* con ~8.7 Mb en el cromosoma (Bentley *et al.* 2002).

Es claro que la evolución de los genomas es ambiente dependiente, con una tendencia a la reducción en sistemas poco cambiantes, por el contrario, los habitantes de ambientes en continuo cambio, requieren de un número de genes que les permita sobrevivir y colonizar estos nichos. De manera similar se acepta que el tamaño genómico de las especies es la suma de diferentes eventos histórico evolutivos, como es la duplicación genética, la adquisición de nuevos genes y las perdidas de genes linaje-específicos (Bentley *et al.* 2004).

8.2.2 Contenido de nucleótidos

Las variante de *B. thuringiensis* no presentaron variaciones estadísticamente significativas en el contenido promedio de bases nucleotídicas que conforman el genoma cromosomal. Algunos de los genomas contienen una cantidad de bases sin definir (N), producto de la fase de secuenciación en que se encontraron al momento de obtener las secuencias o a imprecisiones de la técnica. Como se puede observar en la apéndice del contenido de nucleótidos, de forma similar se calculó el contenido de nucleótidos en las especies de *B. anthracis, B. cereus, B. weihenstephanensis* y *B. subtilis* que se presentan en la figura 1, al final de esta, se presenta la representación del promedio de los contenidos de bases nucleotídicas en los genomas estudiados. En esta figura es claro que las especies de *B. subtilis* son diferentes en el contenido de citosina y guanina, con respecto a las especies del grupo *B. cereus* analizadas. Mientras que entre estas últimas,

las variantes de *B. thuringiensis* tienen el menor contenido en promedio de citosinas con respecto al grupo *B. cereus*.

8.2.3 Porcentaje de Adenina-Timina

De acuerdo a los resultados presentes en la tabla 2, el contenido promedio de AT en las variantes de *B thuringiensis* es de 64.77%, lo cual representa un valor alto en estas bases. Es conocido que las mutaciones al azar en donde se cambia una C por T, por la deaminacion de la citosina para formar uracilo, el cual es subsecuentemente replicado como timidina es una forma muy frecuente de mutación, esto es en usencia de reparación del DNA o en circunstancias que rebasan este proceso. Funcionalmente, un alto porcentaje de contenido de AT correlaciona con estructuras distintivas en el DNA, como son las regiones rígidas que frecuentemente son separadas y leídas mas rápidamente, también pueden ser compactadas fácilmente (Pedersen *et al.* 2000). Altos niveles de sesgo de AT se presenta en genes con ortólogos en el mismo genoma por lo que se estiman como no esenciales, reflejando la presión opuesta de los sesgos bajos de AT, que se explican como importantes para preservar la función de un gen indispensable.

De acuerdo al comportamiento en la composición de bases que generalmente muestra la mayoría de las especies estudiadas a nivel global, existe una correlación entre el tamaño del genoma y su contenido de GC, así genomas grandes de habitantes de suelo tienden a contener un alto contenido de GC, mientras que los genomas reducidos de simbiontes obligados tienen a un contenido alto de AT. Un contenido medio de GC se ha encontrado en genomas que presumiblemente poseen un ciclo biológico mas complejo, entre un nicho simbiótico y uno de vida libre, bien denominado un sistema mixto (Bentley *et al.* 2004).

8.2.4 Porcentaje de Guanina-Citosina

En el caso de B. *thuringiensis*, el contenido de GC puede es considerado bajo ~35% (promedio = 34.80%, tabla 1 y 2), mientras que para el habitante común de suelo *B. subtilis* es de 44%. *B anthracis*, *B cereus* y *B weihenstephanensis* que completan el grupo *B. cereus*, contienen 35.38%, 35.44% y 35.56% respectivamente. Las diferencias en el porcentaje de GC entre ellas no son estadísticamente diferentes. Los valores de GC que encontramos asemejan a especies simbióticas como *Ureaplasma parvum* o *Mycoplasma pneumoniae* que contienen en promedio 25 y 32% de GC respectivamente, estas ultimas especies adicionalmente tienen tamaño genómico reducido y una tasa evolutiva rápida (Ochman 1999).

Esto sugiere que por el alto contenido de AT y por consiguiente bajo contenido de GC, *B. thuringiensis* podría tratarse de una especie simbiótica y que su aislamiento en ambientes libres, es solo parte de su ciclo natural, en donde se encuentran las forma de resistencia, que no necesariamente demuestra que estas formas puedan germinar en este ambiente. Mucho se ha estudiado de la resistencia a diferentes condiciones climáticas que las esporas de *B thuringiensis* presenta (Schmidt 1955, ; Myasnik *et al.* 2001, Nicholson, 2002 #785; Saxena *et al.* 2002; Nicholson *et al.* 2005) argumentos que permiten suponer que su sobrevivencia y prevalencia en ambientes naturales, puede ser debido a esta particular habilidad de resistir los cambios ambientales. Por otro lado, el análisis de nucleótidos de los genomas, refleja que estas variedades se asemejan más a las especies simbióticas, que aquellas especies de vida libre o de nichos complejos.

8.2.5 Sintenía

El concepto de sintenía se refiere a regiones de multigenes donde la secuencia de DNA y el orden de los genes esta conservada entre los genomas. El análisis detallado de los mapas genómicos de *Escherichia coli* (Perna *et al.* 2001; Welch *et al.* 2002; Binnewies *et al.* 2006; Brzuszkiewicz *et al.* 2006) (Sims *et al.* 2011)y *Bacillus subtilis* (Kunst *et al.* 1997; Moszer 1998; Istock *et al.* 2001)indican que los genes no necesariamente ocurren a una posición relativa del genoma en todas las especies, pero se pueden detectar ciertos grupos de genes con sintenia, como los regulones y operones. Es ampliamente aceptado que los genomas bacterianos esta ordenados principalmente en operones biosintéticos y catabólicos que facilitan su regulación y expresión en diferentes estadios de la célula (Lawrence 2002).

Estudios anteriores en el grupo *Bacillus cereus* en búsca de sintenia fueron reportados por Rasko y col. (2005), quienes encontraron un alto nivel de ordenación y similitud a nivel de proteínas. El análisis se realizo con datos de *B. anthracis* y *B. cereus*. Por otro lado, Anderson y col. (2005), reportan haber encontrado una serie de rearreglos cromosomales, secuencias remantes de fagos, además de la perdida en el orden de los genes y cambios en los nucleótidos (Anderson *et al.* 2005). Similar a lo que encontramos en nuestro análisis. Como se ha comentado anteriormente, las variantes de *B. anthracis* se presentan como un grupo compacto con muy poca o escasa variabilidad en la mayoría de los estudios (Harrell *et al.* 1995; Read *et al.* 2002) por lo que es de esperar que también conserve el orden en sus genes. Mientras que la historia que presentan las variantes de *B thuringiensis* son muy diferentes a aquellas descritas por *B. anthracis* y ha sido ampliamente documentada por diferentes investigadores (Harrell *et al.* 1995; Read *et al.* 2002).

La búsqueda de sintenia, también fue objeto de estudio en 61 variantes de *Escherichia coli* y *Shigella sp.* En donde el orden en los genes tampoco fue conservado, adicionalmente, se reporta que la localización génica depende de los genomas que se utilicen en el análisis. La gráfica de la posición relativa de los genes mostró que estos se distribuyen sobre varia islas con respecto a un genoma de referencia (Lukjancenko *et al.* 2010; Sims *et al.* 2011). Lo que recuerda la presencia de operones como unidad funcional en los genomas bacterianos. Este hecho tampoco pudo ser puesto de manifiesto en nuestro análisis.

En nuestro análisis, la presunción de sintenia en genomas de la misma especie presentado en la figura 3, no pudo ser observada en todas las variantes de *B. thuringiensis*. De hecho, el supuesto de conservación en el orden y posición relativa de

los genes no fue posible detectarla en la mayoría de las variantes analizadas, bien debido a la baja calidad de las secuencias, a que se trate de una especie con un alto porcentaje de recombinación o que las poblaciones de *B. thuringiensis* estén estructuradas. Para probar la última posibilidad, y saber si existe una estructura poblacional de tipo panmitico en la especie, realizamos la caracterización de la estructura poblacional, con la codificación de las secuencias, la cuantificación de la variabilidad genómica y por último la identificación de la estructura poblacional en la especie.

8.3 Edición de secuencias

El descubrimiento de homología mediante la evaluación de la similitud entre secuencias se ha vuelto rutinario con el desarrollo de métodos más rápidos para la comparación y búsquedas en bases de datos. Por otro lado, las comparaciones inter-genómicas son mas eficiente cuando los genomas están completamente terminados, que cuando se trabaja con ensambles desordenados típicos de un proyecto en progreso (Read *et al.* 2002). En nuestro planteamiento original, la presencia de genomas en proceso fue el paso limitante, por el cual se decidió hacer una edición y codificación de las secuencias disponibles. La edición de secuencias y una supervisión humana son partes fundamentales de cualquier análisis genómico (Brent 2005), debido a los potenciales errores sistemáticos que pueden ser encontrados en la asignación de una función particular a una secuencia dada (Galperin *et al.* 1998; Brenner 1999; Jones *et al.* 2007; Medigue *et al.* 2007; Schnoes *et al.* 2009; Koser *et al.* 2011). En este caso, trabajamos con el 36% de las secuencia, cambiando el objetivo inicial de este proyecto, que era precisamente trabajar con las regiones no categorizadas, sin embargo la baja calidad de las secuencia dificultó la obtención de resultados confiables. Adicional a la edición de secuencias, varios CDS fueron actualizados conforme a lo reportado en los sitios referentes anteriormente señalados.

8.4 Codificación de CDS de cada secuencia genómica para realizar un análisis numérico

Las matrices de datos de diferentes marcadores o perfiles que presenta un sistema particular, son tradicionalmente codificadas, para facilitar o realizar su análisis y la obtención de información. Los conocimientos más tradicionales en esta área, son las bases matemáticas de la biología, en donde un proceso biológico se transforma en una ecuación matemática que pretende describir y capturar la esencia del proceso.

Al inicio de la ciencia, estos elementos de carácter fenotípico, fueron codificados para simplificar y clasificar diferentes procesos y organismos, sobre todo aquellos que nos cualitativos. Así, se puede recordar las diferentes escalas de color o la asignación de un número a un determinado color. Con la incorporación de los resultados producidos por técnicas moleculares, los procesos de codificación fueron modificados con fin de incluir estos resultados en un análisis objetivo. La gran revolución fue vivida con la llegada de las tecnologías de investigación masiva, como los microarreglos, que generan una cantidad nunca antes vista de datos que requirieron de ser interpretados. A través de este portal, el desarrollo de herramientas computacionales que facilitarán esta tareas fueron implementadas, sin embargo los principios e índices matemáticos, permanecieron inmutables en la mayoría de los casos.

En la era de la post-genómica, en donde un sin número de secuencias nucleotídicas y de amino ácidos son generados con una velocidad similar a la de la luz. Están al orden del día las publicaciones que reportan la secuencia completa de una especie o modelo biológico especifico. Lo que anteriormente por su relevancia se publicaba en Science o Nature (Tomb *et al.* 1997; Cole *et al.* 1998) (Hopwood 1967; Cole *et al.* 1998; Initiative 2000; Goodner *et al.* 2001; Venter *et al.* 2001; Wood *et al.* 2001; Bentley *et al.* 2002; Collins *et al.* 2003) (Tomb *et al.* 1997; Wood *et al.* 2002) entre las que se incluyó las secuencias *de B cereus* (Ivanova *et al.* 2003) y *B subtilis* (Harwood *et al.* 1996; Kunst *et al.* 1997); actualmente han sido relegadas a una parte de muchas revista de muy menor impacto y con un mayor nivel de especialización. Por otro lado, los requisitos a cumplir para el análisis y la comparación, como son los reportes encontrados en *Streptococcus* en Genome Biology and Evolution (Suzuki *et al.* 2011), o en Genome Biology (Lefebure *et al.* 2007) son cada vez mas estrictos, poniendo de manifiesto que lo que ahora es prioritario, es obtener información a partir de los ya generados datos. Una muy buena revisión de lo que ha pasado en la secuenciación de genomas bacterianos en los últimos anos puede encontrarse en los reportes de (Binnewies *et al.* 2006) (Alcaraz *et al.* 2010).

Por otro lado, las recién creadas tecnologías de secuenciación masiva, abrieron la puerta a proyectos que no solo incluyen una especie, sino que actualmente, podemos hacer la secuenciación de una comunidad bacteriana de cualquier nicho ecológico, sin necesidad de tener un cultivo o cepa en el laboratorio, todo esto en un tiempo record y con recursos monetarios cada vez menos demandantes, son ejemplo de esto los genomas que colonizan diferentes áreas del cuerpo humano (Costello *et al.* 2009), en el intestino (Qin *et al.* 2010), en piel (Grice *et al.* 2009), biofilm acidofilicos (Tyson *et al.* 2004), de ambientes acuáticos (Rodriguez-Brito *et al.* 2010), suelo, de cepas no cultivable (Schloss *et al.* 2005) incluso de virus (Rodriguez-Brito *et al.* 2010) y microRNAs

Actualmente ya existe una base de datos que contiene los distintos proyectos y metagenomas GOLD que al 12 de marzo de 2012 contenía un total de 15902 genomas, de los cuales 3173 están terminados (Liolios *et al.* 2010).

La otra cara de la moneda, es entonces el ensamblaje, la anotación y obtención de información aplicable a la vida cotidiana. Es entonces cuando regresamos la mirada al origen y retomamos las tan cotizadas notaciones matemáticas y la simplificación de los modelos con fin de recrear un panorama de la complicada e intricada vida de los microrganismos.

No es de extrañar entonces, que nosotros optásemos por esta vía, al igual que muchos otros autores (Burke *et al.* 2004) (Earl *et al.* 2007) (Dworzanski *et al.* 2009) (Lukjancenko *et al.* 2010). Este camino nos permitió estar en posición de hacer una contribución a la historia de vida de *B. thuringiensis* realizando por primera vez una aproximación sistemática fenética basada en datos genómicos. Como se ha descrito en el apartado de método, utilizamos dos sistemas de codificación, uno que hace un simplificación del sistema a un código con solo dos posibles opciones, el denominado sistema binario, que se traduce a la presencia o ausencia de un CDS en nuestro caso, similar a lo hecho por Daubin y col. (2007), en donde además se adhiere el componente mutacional evolutivo al sistema, debido a que si se considera que los eventos de ganar o perder un gen es relativamente raro, entonces la presencia o ausencia de un gen en un genoma puede considerarse como un carácter informativo binario (Daubin *et al.* 2002).

La parte de innovación de este estudio es la codificación en un sistema continúo que representa la frecuencia relativa de la cantidad de copias génicas de cada una de las potenciales proteínas codificadas en cada variante genómica de *B. thuringiensis*. Esta aproximación es menos frecuente utilizada, debido a su mayor grado de dificultad de manejo e interpretación de resultados. Sin embargo, la gran mayoría de los índices matemáticos de la genética poblacional manejan preferentemente esta opción (Hartl *et al.* 1989).

A la fecha muchas técnicas moleculares particularmente usadas en genómica y proteomica, toman uno de las dos sistemas de codificación anteriores, para poder hacer un análisis numérico de los datos, como por ejemplo, los AFLP son codificados como datos binarios (Burke *et al.* 2004), al igual que en la hibridación comparativa por microarreglos (Earl *et al.* 2007), la cromatografía liquida – de inozación por electrospray- acoplado a espectrometría de masas en tándem (LC-MS-MS) (Dworzanski

et al. 2009), incluso en estudio de proteomas predichos a partir de los genomas completos (Lukjancenko *et al.* 2010).

8.5 Variabilidad genómica

Como se ha explicado anteriormente, existen diferentes formas de referir la variabilidad en una especie. Muchas de estas son de tipo cualitativo, sin embargo en este trabajo desarrollamos varias aproximaciones cuantitativas. Como se muestra a continuación:

8.5.1 Contenido de CDS y tamaño del genoma

Correlacionamos el contenido total de CDS y el tamaño del genoma que los contienen. Esta gráfica puede ser observada en la figura 5. En el caso de *B. thuringiensis*, en relación al tamaño y densidad de genes codificantes de proteínas encontramos tres grupos, si bien hay una discusión abierta del potencial nicho ecológico de la especies (Meadows *et al.* 1992), estos hallazgos permiten proponer que pueden coexistir las tres posibilidades ecológicas que se han postulado con anterioridad.

En general la densidad en el contenido de los genomas bacterianos, no varía mucho más de aproximadamente un gen por kilobase de DNA. Claramente, los tamaños genómicos son directamente proporcionales al número de genes que contienen. Genomas con un gran número de genes pueden ser debido a dos posibilidades: 1) poseer un incremento en el número de familias de genes o por 2) un incremento en el número de miembros que codifican en cada familia. Adicionalmente estos genomas grandes, requieren de un sistema de regulación más complejo que coordine la expresión de genes, con el fin de utilizar eficientemente la energía que la célula produce.

La variación en el numero de genes, es algo que parece mas una constante que una excepción, por ejemplo, el genoma más grande reportado de *E coli* contienen 1,158 más genes predichos que el genomas más pequeño de la misma especie (5,315 genes de la cepa EC4115 con el patotipo/serotipo EHEC/OH157:H7 y 4,157 genes en la cepa de

vida comensal BL21). Por otro lado se ha documentado, que la densidad de genes en los genomas de *E coli* y *Shigella* fueron constantes de 0.911± 0.04 genes por 1000 pb (Lukjancenko *et al.* 2010).

En este trabajo se encontró la presencia de tres grupos con diferencias en el contenido de genes codificantes de proteínas (figura 5). De forma interesante el grupo de menor tamaño contiene a las cepas que presentaron una actividad patogénica en humanos. Si bien este aislamiento o conducta, puedo haber sido accidental, resulta poco probable que estas cepas contengan el mismo número aproximado de genes y que tengan potencial patogénico. Por otro lado, el grupo que contienen la cantidad mayor de elementos codificantes fueron, curiosamente aislados de mamíferos.

8.5.2 Matriz de datos binarios

La correlación entre el numero de loci variables frente a la cantidad de loci exclusivos por cepa presentes en la figura 6 se realizo utilizando la matriz de datos binarios. En esta gráfica es posible ver que la cantidad de loci variables se incrementa en función del incremento de los loci exclusivos, los que suponemos con un origen endógeno. Es decir la transferencia horizontal de genes esta participando activamente en la generación de variabilidad a este nivel. Un caso por demás interesante, se encontró en las cepas T01001 y ATCC-10792 que presentan altos contenidos de loci variables que parecen compartir entre ellas, motivo por el cual se agrupan muy cercanas en la gráfica, sin embargo poseen escasos loci exclusivos. La presencia de HGT es un proceso común de variabilidad en los genomas bacterianos (Charkowski 2004) que les permite adecuarse a ambiente extremos (Earl *et al.* 2008), dentro de los que se incluyen los marinos(Alcaraz *et al.* 2008).

8.5.3 Matriz de datos Multiestado

Tomando como base los datos contenidos en la matriz de datos multiestado, calculamos la heterozigocidad como una medida de la pérdida de la clonalidad entre las cepas y de la diversidad genética en la especie, la correlacionamos con el número de loci polimórficos según se visualiza en la figura 7. Utilizamos esta medida por que es ampliamente utilizada en especies sexuales debido a que permite hacer una comparación numérica directa entre diferentes especies (Whittam *et al.* 1983; Selander *et al.* 1985; Selander *et al.* 1986), además que puede reflejar parte de la estructura poblacional (Hartl *et al.* 1989; Ward *et al.* 1992). Los altos valores obtenidos en la heterozigocidad no permiten establecer la perdida de la clonalidad reflejada en este atributo, de hecho, en esta análisis se reportan los valores mas altos de diversidad genética de entre 0,9985 a 0,9984. Sin embargo se puede apreciar una desviación a la izquierda de este valor, es decir, puede estar disminuyendo la homocigocidad de las variantes analizadas. El alto grado de polimorfismo encontrado y el alto contenido de diversidad genética, permite suponer que se trata de una comunidad con constante e intermitente transferencia o procesos de recombinación entre las poblaciones, que no permite hacer una clara separación de pequeños grupos de clones o linajes en la especie, con la excepción hecha en las serovariedades *israelensis*, que tienen un comportamiento clonal.

8.5.4 Roles Funcionales

La versatilidad y plasticidad metabólica del género no tiene comparativo con especies simbióticas, en este orden se asemejan más a especies de vida libre. Por presentar un ejemplo, los cambios en vías metabólicas, como se ha documentado en la carencia de la alfa-cetoglutarato deshidrogenasa en el ciclo de los ácidos tricarboxilicos, que es remplazada por la acción de una descaboxilasa transaminasa, es sin duda una muestra de la adaptación a un ambiente. (Aronson *et al.* 1975). En la tabla VI se presentan algunos ejemplos que representan de buena manera las anteriores aseveraciones. Por un lado las proteínas que responden a cambios ambientales y controlan los procesos metabólicos presentan altos números de copias y variación en los genomas, esto ya se publicado con anteriormente (Pérez-Rueda *et al.* 2001; Martínez-Antonio *et al.* 2003; Earl *et al.* 2007; Earl *et al.* 2008; van Hijum *et al.* 2009). Además del número de genes o proteínas, los cambios en la expresión de estos también se ha documentado utilizando M-CGH en donde cerca del 28% de 3722 genes presentaron variación en su perfil de expresión con respecto a cepas de *B. subtilis* (Earl *et al.* 2007).

De forma adicional, los genes que codifican elementos móviles y transferibles también tienen un alto contenido numérico y variabilidad entre las cepas analizadas en este estudio.

8.5.5 Índice de Shannon

Como se ha apreciado, las métricas utilizadas en los ejemplos anteriores, solo nos permitieron realizar un cálculo con uno de los dos sistemas utilizados para la codificación de los datos de las secuencias genómicas de *B. thuringiensis*. Por ello, utilizamos ahora el Índice de Shannon que nos permitió hacer una comparación de la variabilidad entre los dos sistemas como puede observarse en la figura 8. Los datos de la matriz binaria mostraron valores más altos en comparación con los obtenidos con los datos de la matriz multiestado. De forma interesante se presentan los valores en distribuciones de variación distintas, mientras que los datos codificados como caracteres binarios presenta una distribución normal, similar a lo que representaría un alelo en Equilibrio de HW, los datos codificados como caracteres multiestado presenta una distribución tri-modal, similar a lo que se ha encontrado en utilizando otros índices. De esta forma comprobamos la premisa dada por Smith, quién acertadamente describe que la representación de la variabilidad depende de gran medida en la forma de cómo se codifiquen y analicen los datos (Smith *et al.* 1993).

8.6 Estructura genómica poblacional de *Bacillus thuringiensis*

Como se ha descrito anteriormente, en especies microbianas con reproducción asexual se han descrito tres posibles estructuras poblacionales, que van desde la clásica estructura clonal, una estructura epidémica hasta lo que se considera en panmixia. Esta estructuración de las poblaciones obedece a sistemas de reproducción de tipo sexual, bien generados por los sistemas de intercambio inespecífico que poseen las bacterias, como son la transformación, la conjugación, la transducción y en general la transferencia horizontal de genes por cualquiera que sea el mecanismo que lleve a ello, el fallo o no de los sistemas SOS o de un incesante sistema de recombinación génica. El flujo constante de material genético de diferentes fuentes, genera una pérdida de la clonalidad en diferentes niveles, este hecho puede verse reflejado en la estructura de la población, desde hace varios años el dogma de la clonalidad en especies bacterianas con reproducción asexual, ha venido a menos con el descubrimiento de diferentes especies que rompen el mencionado comportamiento (Hopwood 1967; Feldman *et al.* 1996; Haubold *et al.* 1998; Spratt *et al.* 1999; Achtman 2004; Tettelin *et al.* 2005; Binnewies

et al. 2006). Resaltan los estudios realizados en *B subtilis*, quien precisamente, rompe el paradigma de la clonalidad poblacional (Istock *et al.* 1992)

En nuestro caso, es fácil distinguir los taxones que mantienen una cohesión estrecha en la variedad *isralensis*. A las que pudiéramos designar como un linaje que representa una fracción de la totalidad de la muestra, su presencia recuerda un originen clonal y muy posiblemente su cercanía con *B anthracis*, que presenta un comportamiento poblacionales similar (Ankarloo *et al.* 2000).

Por otro lado, la mayoría de las variedades estudiadas no presentan el mismo patrón de clonalidad, mas aun presentan toda una gama de cambios entre ellos, una variedad semejante a mosaicos, en donde no se localiza fácilmente un patrón determinado. Este hecho puesto de manifiesto en las diferentes agrupaciones obtenidas con los dendrogramas previamente mostrados en las figuras 8 y 9, que nos permite hacer esta aseveración. Lo que denominamos arbitrariamente como grupo B, no se conserva en distribución en las ramas de los arboles. Indicando que estas agrupación pueden no ser necesariamente el reflejo de una origen filogenético de las cepas.

Por otro lado, la distribución de las cepas en las diferentes ramas y la presencia de al menos un linaje conservado en las dos aproximaciones utilizadas para la reconstrucción de los dendrogramas, permite la conclusión de que estas cepas presentan una estructural poblacional de tipo epidémica. Porque cumple con todos los requisitos descritos para este modelo. Es decir, se observa la presencia de linajes conservados en espacio y tiempo, y linajes que no presentan ninguna relación aparente, con una distribución geográfica, de huésped de aislamiento o tipo comportamiento ecológico desarrollado.

Similar a lo encontrado por Ankarloo y col, en cepas aisladas de suelo de *B. thuringiensis* var *israelensis* H14, en nuestro estudio las serovariedades de *isralensis* también presentan un linaje clonal. Por otro lado, la totalidad de las cepas no presenta ese patrón. Por el contrario son más variables en varios niveles. A la luz de estos resultados, recomendamos que los estudios reportados en mezclas de especies del grupo *Bacillus cereus,* como se presenta en: (Ko *et al.* 2004; Priest *et al.* 2004, Helgason, 1998 ; Ehling-Schulz *et al.* 2005), no pueden ser tomada como referentes, dado que el supuesto de que se trata de una sola especie utiliza técnicas que no poseen criterio taxonómico aceptado. Además de poder tratarse de una mezcla de estructuras poblacionales.

9 CONCLUSIONES

- Se obtuvieron con éxito 18 secuencias cromosómicas de igual numero de variantes de *Bacillus thuringiensis* con diferente grado y calidad de secuenciación.

- Al momento del análisis, solo 3 de estas secuencias, se presentaron como proyectos terminados

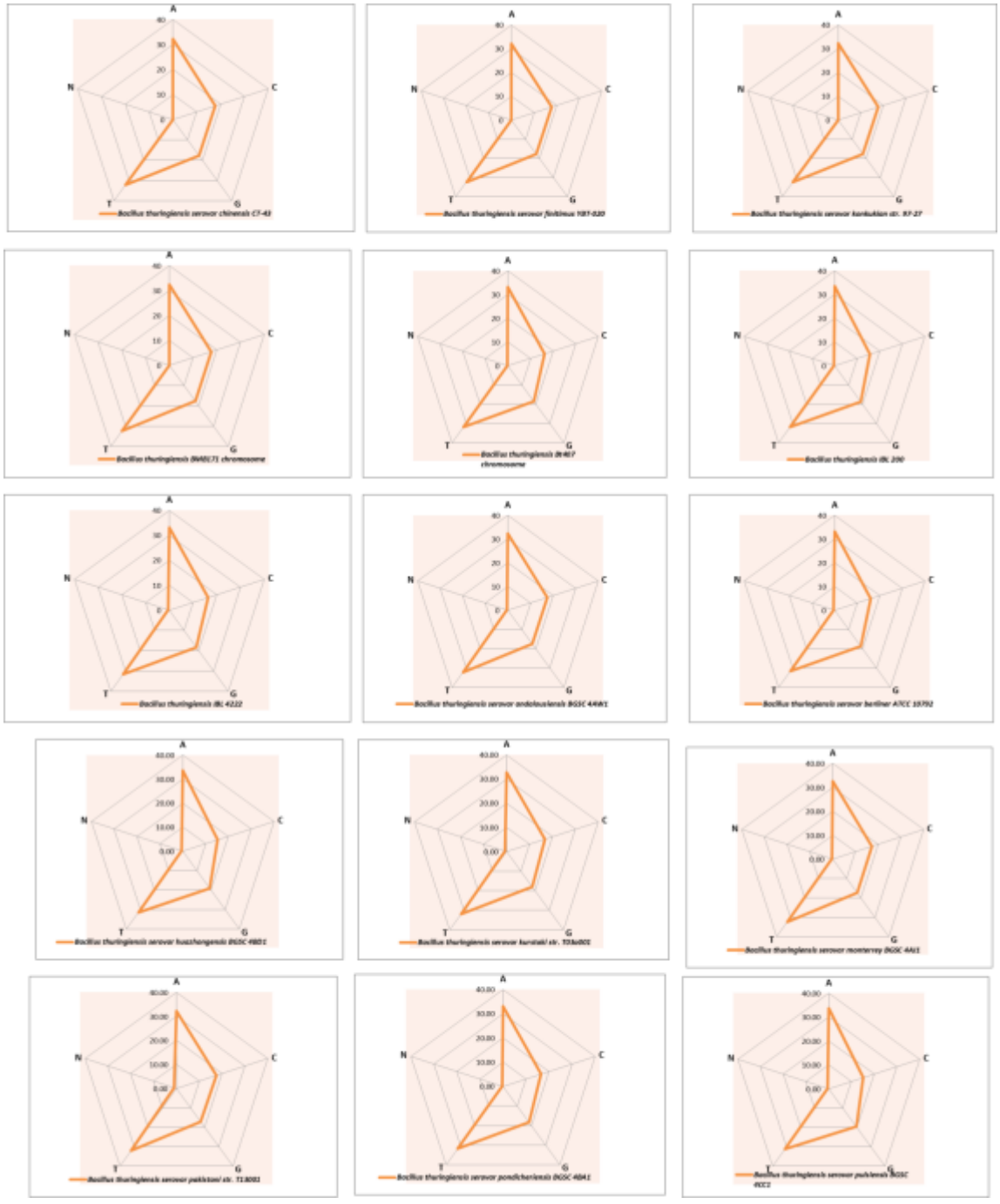
- La cuantificación de la variabilidad genómica, se realizo con diferentes métricas. En donde los valores calculados son los más altos reportados en la especie hasta el momento.

- La presencia constante de tres grupos de cepas estableció una estructura población tipo epidémica en las cepas analizadas de *Bacillus thuringiensis*

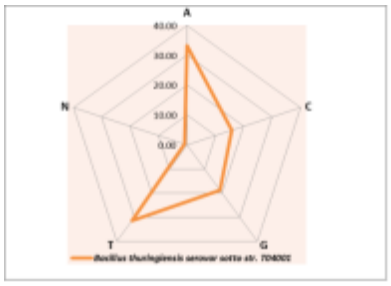
APENDICES

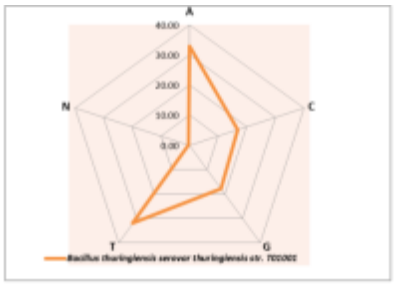
CONTENIDO PROMEDIO DE NUCLEOTIDO DE LAS ESPECIES DEL GRUPO *Bacillus cereus sensu lato*

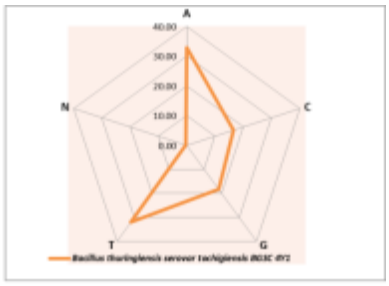
Contenido de Nucleótidos en las cepas de *Bacillus thuringiensis*

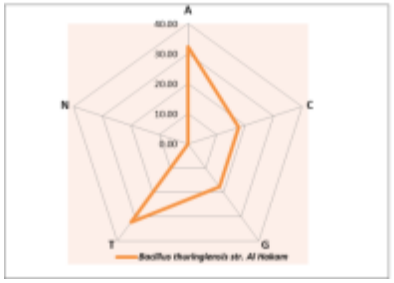


Contenido de Nucleótidos en las cepas de *Bacillus thuringiensis*

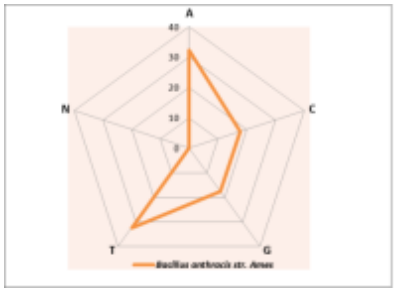


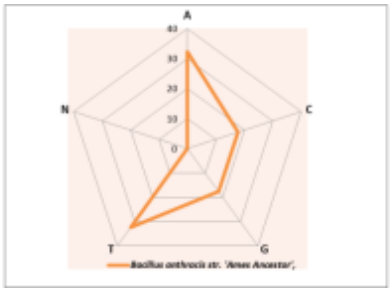


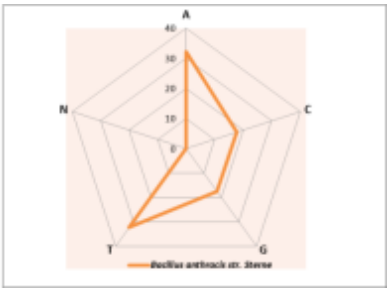




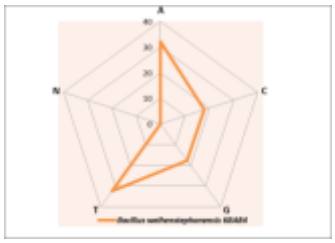
Contenido de Nucleótidos en las cepas de *Bacillus anthracis*



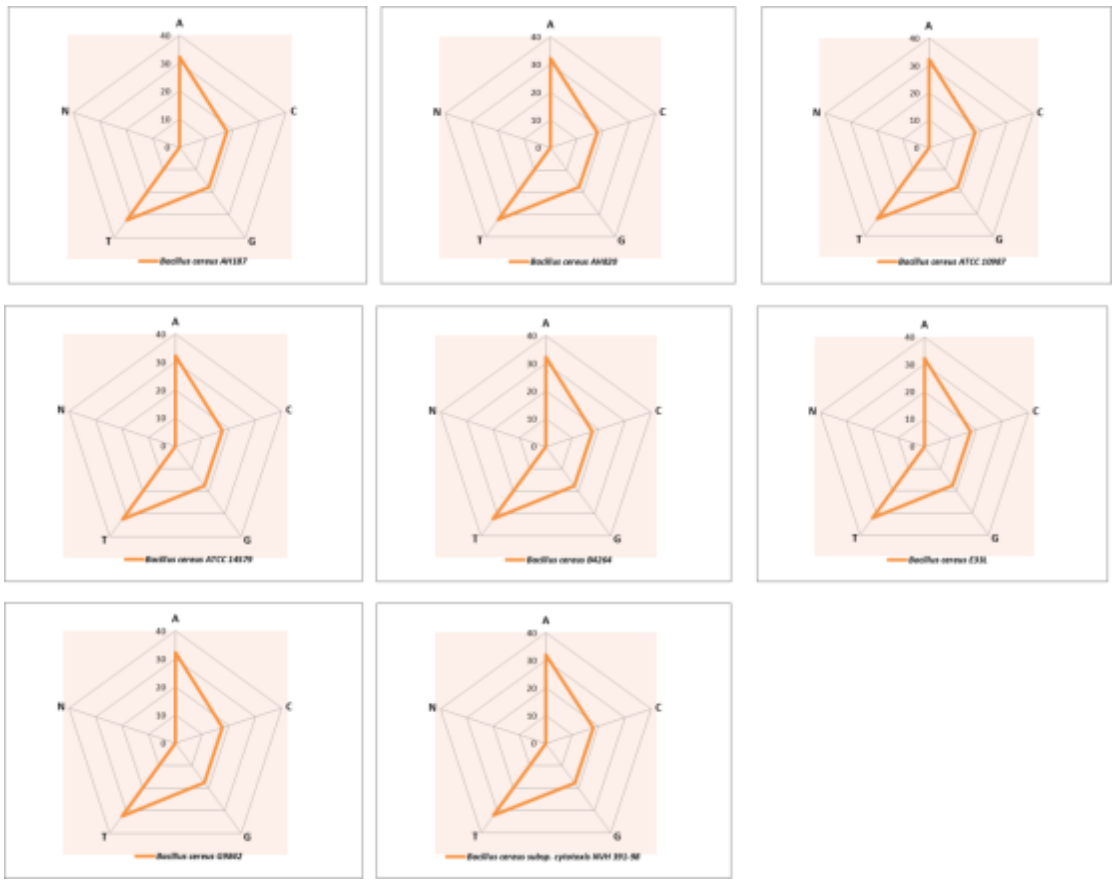




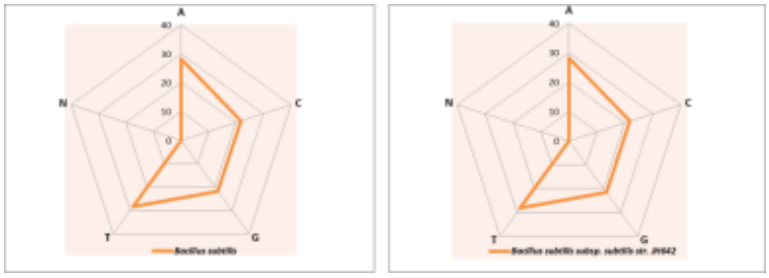
Contenido de Nucleótidos en las cepas de *Bacillus weihenstephanensis*



Contenido de Nucleótidos en las cepas de *Bacillus cereus*



Contenido de Nucleótidos en las cepas de *Bacillus subtilis*



ACRONIMOS Y PORTALES *on line*

BRC Bioinformatics Resource Center http://www.niaid.nih.gov/labsandresources/resources/brc/Pages/default.aspx

JCVI J. Crain Venter Institute and the TIGR Comprehensive Microbial Resource http://www.jcvi.org/, http://gsc.jcvi.org/projects/msc/*Bacillus*/

NIAID National Institute of Allergy and Infectious Diseases http://www.niaid.nih.gov/labsandresources/Pages/Default.aspx

NCBI National Center for Biotechnology Information website http://www.ncbi.nlm.nih.gov/

Pathema-*Bacillus* http://pathema.jcvi.org/cgi-bin/*Bacillus*/PathemaHomePage.cgi

Genome Atlas http://www.cbs.dtu.dk/services/GenomeAtlas/

TIGR. Institute for Genomic Research was a non-profit genomics research institute founded in 1992 by Craig Venter in Rockville, Maryland, United States. It is now a part of the J. Craig Venter Institute. http://www.tigr.org/db.shtml

CMR Comprehensive Microbial Resource http://cmr.jcvi.org/tigr-scripts/CMR/CmrHomePage.cgi

PATRIC Pathosystems Resource Integration Center http://patric.vbi.vt.edu/

TGI .The Gene Index Databases http://compbio.dfci.harvard.edu/tgi/

DFCI Dana Farber Cancer Institute. http://www.danafarber.org

RAST and MG-RAST Rapid Annotation using Subsystem Technology and Meta Rapid Annotation using Subsystem Technology http://metagenomics.nmpdr.org/

EMBL-EBI European Bioinformatics Institute at European Molecular Biology Laboratory, http://www.ebi.ac.uk/

Ensembl The goal of Ensembl was therefore to automatically annotate the genome, integrate this annotation with other available biological data and make all this publicly available via the web. Ensembl is a joint project between EBI, an outstation of the EMBL, and the Wellcome Trust Sanger Institute (WTSI). Both institutes are located on the WTSI in Hinxton, south of the city of Cambridge, United Kingdom. http://www.ensembl.org/index.html

Genome Atlas 3.0 DNA structural atlases for complete microbial Genomes http://www.cbs.dtu.dk/services/GenomeAtlas/

Microgen, the laboratory for genomics and bioinformatics. University of Oklahoma Health Sciences Center. http://microgen.ouhsc.edu/index.html

GOLD, Genomes OnLine Database http://www.genomesonline.org/

UniProt , Universal Protein Resource http://www.uniprot.org/

HMM, Hidden Markov Models

RAST. Rapid Annotation using Subsystem Technology

The National Centre for Text Mining (NaCTeM) http://www.ddbj.nig.ac.jp/FT/full_index.html

The Taxonomicon & Systema Naturae 2000 http://sn2000.taxonomy.nl/main/classification/252.htm

LITERATURA CITADA

Achtman, M. (2004). Population structure of pathogenic bacteria revisited. International Journal of Medical Microbiology 294: 67-73.

Addison, J. A. (1993). Persistance and non-target effects of *Bacillus thuringiensis* in soil: a review. Canadian Journal of Forest Research 23: 2329-2342.

Alcaraz, L. D., G. Moreno-Hagelsieb, L. E. Eguiarte, V. Souza, L. Herrera-Estrella and G. Olmedo (2010). Understanding the evolutionary relationships and major traits of *Bacillus* through comparative genomics. BMC Genomics 11: 332.

Alcaraz, L. D., G. Olmedo, G. n. Bonilla, R. Cerritos, G. Hernadez, et al. (2008). The genome of Bacillus *coahuilensis* reveals adaptations essential for survival in the relic of an ancient marine environment. Proceedings of the National Academy of Sciences 105(15): 5803-5808.

Anderson, I., A. Sorokin, V. Kapatral, G. Reznik, A. Bhattacharya, et al. (2005). Comparative genome analysis of *Bacillus cereus* group genomes with Bacillus subtilis. FEMS Microbiology Letters 205: 175-184.

Andersson, J. O. and S. G. Andersson (1999). Genome degradation is an ongoing process in *Rickettsia*. Molecular Biology and Evolution 16(9): 1178-1191.

Andorf, C., D. Dobbs and V. Honavar (2007). Exploring inconsistencies in genome-wide protein function annotations: a machine learning approach. BMC Bioinformatics 8(1): 284.

Ankarloo, J., D. A. Caugant, B. M. Hansen, A. Berg, A.-B. Kolsto and A. Lovgren (2000). Genome stability of *Bacillus thuringiensis* subsp israelenesis isolates. Current Microbiology 40(1): 51-56.

Apaydin, O. (2004). Isolation and Characterization of *Bacillus thuringiensis* strains from different grain habitats. Biotechnology and Bioengineering. Turkey, Izmir Institute of Technology. Master.

Aronson, J. N., D. P. Borris, J. F. Doerner and E. Akers (1975). Gamma-Aminobutyric acid pathway and modified tricarboxylic acid cycle activity during growth and sporulation of *Bacillus thuringiensi*s. Appl Microbiol. 30(3): 489-492.

Arrieta, G., A. Hernandez and A. M. Espinoza (2004). Diversity of Bacillus thuringiensis strains isolated from coffe plantations infested with the coffe berry borer *Hypothenemus hampei*. Rev. Biol. Trop. 52(3): 757-764.

Ash, C., J. A. E. Farrow, S. Wallbanks and M. D. Collins (1991). Phylogenetic heterogeneity of the genus *Bacillus* revealed by comparative analysis of small-subunit-ribosomal RNA sequences. Lett Appl Microbiol 13: 202 - 206.

Bel, Y., F. Granero, T. M. Alberola, Martinez-Sebastian and J. Ferre (1997). Distribution, frequency and diversity of *Bacillus thuringiensis* in olive three invironments in Spain. Systematic and Applied Microbiology 20: 652-658.

Bentley, S. D., K. F. Chater, A. M. Cerdeno-Tarraga, G. L. Challis, N. R. Thomson, et al. (2002). Complete genome sequence of the model actinomycete *Streptomyces coelicolor* A3(2). Nature 417(6885): 141-147.

Bentley, S. D. and J. Parkhill (2004). Comparative Genomic Structure of Prokaryotes. Annual Review of Genetics 38(1): 771-791.

Binnewies, T., Y. Motro, P. Hallin, O. Lund, D. Dunn, et al. (2006). Ten years of bacterial genome sequencing: comparative-genomics-based discoveries. Functional & Integrative Genomics 6(3): 165-185.

Blackwood, K. S., C. Y. Turenne, D. Harmsen and A. M. Kabani (2004). Reassessment of Sequence-Based Targets for Identification of Bacillus Species. Journal of Clinical Microbiology 42(4): 1626-1630.

Brands, S. (2004-2011). The Taxonomicon & Systema Naturae 2000. Retrieved 19.12.2011, from http://www.taxonomy.nl/taxonomicon/TaxonTree.aspx?id=71320.

Brenner, S. E. (1999). Errors in genome annotation. Trends Genet 15(4): 132-133.

Brent, M. R. (2005). Genome annotation past, present, and future: How to define an ORF at each locus. Genome Research 15(12): 1777-1786.

Brzuszkiewicz, E., H. Brggemann, H. Liesegang, M. Emmerth, T. Ã–lschlger, et al. (2006). How to become a uropathogen: Comparative genomic analysis of extraintestinal pathogenic *Escherichia coli strains*. Proceedings of the National Academy of Sciences 103(34): 12879-12884.

Burke, S. A., J. D. Wright, M. K. Robinson, B. V. Bronk and R. L. Warren (2004). Detection of Molecular Diversity in *Bacillus atrophaeus* by Amplified Fragment Length Polymorphism Analysis. Appl. Environ. Microbiol. 70(5): 2786-2790.

Camin, J. H. and R. R. Sokal (1965). A method for deducing braching sequences in phylogeny. Evolution 19: 311-326.

Carlson, C. R., D. A. Caugant and A.-B. Kolsto (1994). Genotypic Diversity among *Bacillus cereus and Bacillus thuringiensis* Strains. Appl. Environ. Microbiol. 60(6): 1719-1725.

Cavalli-Sforza, L. L. and A. W. F. Edwards (1967). Phylogenetic analysis: Models and estimation procedures. Evolution 21: 550-570.

Challacombe, J. F., M. R. Altherr, G. Xie, S. S. Bhotika, N. Brown, et al. (2007). The Complete Genome Sequence of *Bacillus thuringiensis* Al Hakam. J. Bacteriol. 189(9): 3680-3681.

Charkowski, A. O. (2004). Making sense of an alphabet soup: the use of a new bioinformatics tool for identification of novel gene islands. Focus on Identification of genomic islands in the genome of *Bacillus cereus* by comparative analysis with *Bacillus anthracis*. Physiological Genomics 16(2): 180-181.

CIBIOGEM-CONABIO, P. G.-. Sistema de Información de Organismos Vivos Modificados (SIOVM). Retrieved 10.01.2012, from http://www.conabio.gob.mx/conocimiento/bioseguridad/doctos/consulta_SIOVM.html.

Cole, S. T., R. Brosch, J. Parkhill, T. Garnier, C. Churcher, et al. (1998). Deciphering the biology of *Mycobacterium tuberculosis* from the complete genome sequence. Nature 393(6685): 537-544.

Collins, F. S., M. Morgan and A. Patrinos (2003). The Human Genome Project: Lessons from Large-Scale Biology. Science 300(5617): 286-290.

Costello, E. K., C. L. Lauber, M. Hamady, N. Fierer, J. I. Gordon and R. Knight (2009). Bacterial Community Variation in Human Body Habitats Across Space and Time. Science 326(5960): 1694-1697.

Daffonchio, D., A. Cherif and S. Borin (2000). Homoduplex and Heteroduplex Polymorphisms of the Amplified Ribosomal 16S-23S Internal Transcribed Spacers Describe Genetic Relationships in the "*Bacillus cereus Group*". Appl. Environ. Microbiol. 66(12): 5460-5468.

Damgaard, P. H., A. Abdel-Hammed, J. Eilenberg and P. H. Smits (1998). Natural occurence of *Bacillus thuringiensis* on grass foliage. World Journal of Microbiology and Biotechnology 14: 239-242.

Dangar, T. K., Y. K. Babu and J. Das (2010). Population dynamics of soil microbes and diversity of *Bacillus thuringiensis* in agricultural and botanic garden soils of India. African Journal of Microbiology Research 9(4): 495-501.

Darling, A. E., B. Mau and N. T. Perna (2010). Progressive Mauve: Multiple Genome Alignment with Gene Gain, Loss and Rearrangement. PLoS ONE 5(6): e11147.

Darling, A. E., I. Miklos and M. A. Ragan (2008). Dynamics of genome rearrangement in bacterial populations. PLoS genetics 4(7): e1000128.

Das, J., B. Da and T. K. Dangar (2006). Microbial population and *Bacillus thuringiensis* diversity in saline rice field soils of coastal Orissa, India. African Journal of Microbiology Research: 326-331.

Daubin, V., M. Gouy and G. PerriÃ¨re (2002). A Phylogenomic Approach to Bacterial Phylogeny: Evidence of a Core of Genes Sharing a Common History. Genome Research 12(7): 1080-1090.

Drobniewski, F. A. (1993). *Bacillus cereus* and related species. Clinical Microbiology Reviews 6(4): 324-338.

Dworzanski, J. P., D. N. Dickinson, S. V. Deshpande, A. P. Snyder and B. A. Eckenrode (2009). Discrimination and Phylogenomic Classification of *Bacillus anthracis-cereus-thuringiensis* Strains Based on LC-MS/MS Analysis of Whole Cell Protein Digests. Analytical Chemistry 82(1): 145-155.

Earl, A. M., R. Losick and R. Kolter (2007). *Bacillus subtilis* Genome Diversity. J. Bacteriol. 189(3): 1163-1170.

Earl, A. M., R. Losick and R. Kolter (2008). Ecology and genomics of *Bacillus subtilis*. Trends in Microbiology 16(6): 269-275.

Edwards, A. W. F. and L. L. Cavalli-Sforza (1964). Reconstruction of evolutionary tree. Phenetic and phylogenetic classification. V. H. Heywood and J. McNeill. Oxford, Oxford Science Publication, The Systematics Association Publication. 6: 67-76.

Edwards, J. D. (2007). Oracle Oracle Corporation.

Ehling-Schulz, M., B. Svensson, M.-H. Guinebretiere, T. Lindback, M. Andersson, et al. (2005). Emetic toxin formation of *Bacillus cereus* is restricted to a single evolutionary lineage of closely related strains. Microbiology 151(1): 183-197.

Eisen, J., J. Heidelberg, O. White and S. Salzberg (2000). Evidence for symmetric chromosomal inversions around the replication origin in bacteria. Genome Biology 1(6): research0011.1 - research0011.9.

Eldredge, N. and J. Cracraft (1980). Phylogenetic patterns and the evolutionary process. New York, Columbia University Press.

Feldman, M. W., S. P. Otto and F. B. Christiansen (1996). Population genetic perspectives on the evolution of recombination. Annual Review of Genetics 30(1): 261-295.

Fitch, W. L. and E. Margoliash (1967). Construction of phylogenetic tree. Science 155: 279-284.

Freitas, D., M. Reis, C. Lima-Bittencourt, P. Costa, P. Assis, E. Chartone-Souza and A. Nascimento (2008). Genotypic and phenotypic diversity of *Bacillus spp*. isolated from steel plant waste. BMC Research Notes 1(1): 92.

Fritze, D. (2004). Taxonomy of the Genus *Bacillus* and Related Genera: The Aerobic Endospore-Forming Bacteria. Phytopathology 94(11): 1245-1248.

Galperin, M. Y. and E. Koonin (1998). Sources of systematic error in functional annotation of genomes: domain rearrangement, non-orthologous gene displacement and operon disruption. In Silico Biol 1(1): 55-67.

Gilks, W. R., B. Audit, D. De Angelis, S. Tsoka and C. A. Ouzounis (2002). Modeling the percolation of annotation errors in a database of protein sequences. Bioinformatics 18(12): 1641-1649.

Giovannoni, S. J., H. J. Tripp, S. Givan, M. Podar, K. L. Vergin, et al. (2005). Genome Streamlining in a Cosmopolitan Oceanic Bacterium. Science 309(5738): 1242-1245.

Goodner, B., G. Hinkle, S. Gattung, N. Miller, M. Blanchard, et al. (2001). Genome Sequence of the Plant Pathogen and Biotechnology Agent *Agrobacterium tumefaciens* C58. Science 294(5550): 2323-2328.

Green, M. L. and P. D. Karp (2005). Genome annotation errors in pathway databases due to semantic ambiguity in partial EC numbers. Nucleic Acids Research 33(13): 4035-4039.

Grice, E. A., H. H. Kong, S. Conlan, C. B. Deming, J. Davis, et al. (2009). Topographical and Temporal Diversity of the Human Skin Microbiome. Science 324(5931): 1190-1192.

Hallin, P. F. and D. W. Ussery (2004). CBS Genome Atlas Database: a dynamic storage for bioinformatic results and sequence data. Bioinformatics 20(18): 3682-3686.

Han, C. S., G. Xie, J. F. Challacombe, M. R. Altherr, S. S. Bhotika, et al. (2006). Pathogenomic Sequence Analysis of *Bacillus cereus* and *Bacillus thuringiensis* Isolates Closely Related to *Bacillus anthracis*. J. Bacteriol. 188(9): 3382-3390.

Harrell, L. J., G. L. Andersen and K. H. Wilson (1995). Genetic variability of *Bacillus anthracis* and related species. J. Clin. Microbiol. 33(7): 1847-1850.

Hartl, D. L. and A. G. Clark (1989). Principles of population genetics. Sunderland, Massachusetts, Sinauer Associates, Inc.

Harwood, C. R. and A. Wipat (1996). Sequencing and functional analysis of the genome of *Bacillus subtilis* strain 168. FEBS Letters 389(1): 84-87.

Haubold, B., M. Travisano, P. B. Rainey and R. R. Hudson (1998). Detecting Linkage Disequilibrium in Bacterial Populations. Genetics 150(4): 1341-1348.

He, J., X. Shao, H. Zheng, M. Li, J. Wang, et al. (2010). Complete genome sequence of *Bacillus thuringiensis* mutant strain BMB171. Journal of Bacteriology 192(15): 4074-4075.

Hedrick, P. (2009). Population Genetics and Ecology.

Helgason, E., D. A. Caugant, M.-M. Lecadet, Y. Chen, J. Mahillon, A. Lövgren, I. Hegna, K. Kvaløy and A.-B. Kolstø (1998). Genetic Diversity of *Bacillus cereus/B. thuringiensis;* Isolates from Natural Sources. Current Microbiology 37(2): 80-87.

Helgason, E., D. A. Caugant, I. Olsen and A.-B. Kolstø (2000a). Genetic Structure of Population of *Bacillus cereus* and *B. thuringiensis* Isolates Associated with Periodontitis and Other Human Infections. J. Clin. Microbiol. 38(4): 1615-1622.

Helgason, E., O. A. Okstad, D. A. Caugant, H. A. Johansen, A. Fouet, M. Mock, I. Hegna and A.-B. Kolstø (2000b). *Bacillus anthracis, Bacillus cereus, and Bacillus thuringiensis* -One Species on the Basis of Genetic Evidence. Appl. Environ. Microbiol. 66(6): 2627-2630.

Hernandez, E., F. Ramisse, J.-P. Ducoureau, T. Cruel and C. Jean-Didier (1998). *Bacillus thuringiensis subsp. konkukian* (Serotype H34) Superinfection: Case Report and Experimental Evidence of Pathogenicity in Immunosuppressed Mice. J Clin Microbiol 36(7): 2138-2139.

Hoffmaster, A. R., K. K. Hill, J. E. Gee, C. K. Marston, B. K. De, et al. (2006). Characterization of *Bacillus cereus* Isolates Associated with Fatal Pneumonias: Strains Are Closely Related to *Bacillus anthracis* and Harbor B. anthracis Virulence Genes. J. Clin. Microbiol. 44(9): 3352-3360.

Hopwood, D. A. (1967). Genetic analysis and genome structure in Streptomyces coelicolor. Bacteriol Rev 31(4): 373-403.

Huson, D. H., D. C. Richter, C. Rausch and R. Rupp (2010). Dendroscope, Uni Tuebingen.

Initiative, T. A. G. (2000). Analysis of the genome sequence of the flowering plant *Arabidopsis thaliana*. Nature 408: 796-815.

Iriarte, J., Y. Bel, M. Ferrandis, R. Andrew, J. Murillo, J. Ferre and P. Caballero (1998). Environmental distribution and diversity of *Bacillus thuringiensis* in Spain. Systematic and Applied Microbiology 21: 97-106.

Istock, C., K. Duncan, N. Ferguson and W. Zhou (1992). Sexuality in a natural population of bacteria *Bacillus subtilis* challenges the clonal paradigm. Molecular Ecology 1(2): 95-103.

Istock, C. A., N. Ferguson, N. L. Istock and K. E. Duncan (2001). Geographical diversity of genomic lineages in *Bacillus subtilis* (Ehrenberg) Cohn sensu lato. Organisms Diversity & Evolution 1(3): 179-191.

Ivanova, N., A. Sorokin, I. Anderson, N. Galleron, B. Candelon, et al. (2003). Genome sequence of *Bacillus cereus* and comparative analysis with *Bacillus anthracis*. Nature 423(6935): 87-91.

Jaccard, P. (1901). Distribution de la flore alpine dans le bassin des Dranses et dans quelques regions voisines. Bulletin de la Societi Vaudoise des Sciences Naturelles 37: 241-272.

JCVI. Bacillus cereus Group Genome Project. from http://gsc.jcvi.org/projects/msc/bacillus/.

Jensen, L. J., C. Friis and D. W. Ussery (1999). Three views of microbial genomes. Research in Microbiology 150(9-10): 773-777.

Jin, H., W. Jieping, Y. Wen, S. Xiaohu, Z. Huajun, et al. (2011). Complete genome sequence of *Bacillus thuringiensis subsp. chinensis* strain CT-43. J Bacteriol 193(13): 3407-8.

Jones, C., A. Brown and U. Baumann (2007). Estimating the annotation error rate of curated GO database sequence annotations. BMC Bioinformatics 8(1): 170.

Ko, K. S., J.-W. Kim, J.-M. Kim, W. Kim, S.-i. Chung, I. J. Kim and Y.-H. Kook (2004). Population Structure of the *Bacillus cereus* Group as Determined by Sequence Analysis of Six Housekeeping Genes and the plcR Gene. Infect. Immun. 72(9): 5253-5261.

Koser, C. U., S. Niemann, D. K. Summers and J. A. C. Archer (2011). Overview of errors in the reference sequence and annotation of *Mycobacterium tuberculosis H37Rv*, and variation amongst its isolates. Infection, Genetics and Evolution(0).

Kuehn, B. M. (2008). 1000 Genomes Project Promises Closer Look at Variation in Human Genome. JAMA: The Journal of the American Medical Association 300(23): 2715.

Kunst, F., N. Ogasawara, I. Moszer, A. M. Albertini, G. Alloni, et al. (1997). The complete genome sequence of the Gram-positive bacterium *Bacillus subtilis*. Nature 390(6657): 249-256.

Lawrence, J. G. (2002). Shared strategies in gene organization among prokaryotes and eukaryotes. Cell 110: 407-413.

Lecadet, M. M., E. Frachon, V. C. Dumanoir, H. Ripouteau, S. Hamon, P. Laurent and I. Thiéry (1999). Updating the H-antigen classification of *Bacillus thuringiensis*. Journal of Applied Microbiology 86(4): 660-672.

Lefebure, T. and M. Stanhope (2007). Evolution of the core and pan-genome of *Streptococcus*: positive selection, recombination, and genome composition. Genome Biology 8(5): R71.

Lenski, R. E. (1993). Assessing the genetic structure of microbial populations. Proc Natl Acad Sci USA 90: 4334-4336.

Liolios, K., I.-M. Chen, K. Mavromatis, N. Tavernarakis, P. Hugenholtz, V. M. Markowitz and N. Kyrpides. (2010, 2012-03-12). The Genomes on line database (GOLD) in 2009: status of genomic and metagenomic projects and their associated metadata. Nucleic Acids Res Retrieved 12.03.2012, 2012, from http://www.genomesonline.org/cgi-bin/GOLD/index.cgi.

Lukjancenko, O., T. Wassenaar and D. Ussery (2010). Comparison of 61 Sequenced Escherichia coli Genomes. Microbial Ecology 60(4): 708-720.

Martínez-Antonio, A. and J. Collado-Vides (2003). Identifying global regulators in transcriptional regulatory networks in bacteria. Current Opinion in Microbiology 6(5): 482-489.

Meadows, M. P., D. J. Ellis, J. Butt, P. Jarrett and H. D. Burges (1992). Distribution, Frequency, and Diversity of *Bacillus thuringiensis* in an Animal Feed Mill. Appl Environ Microbiol. 58(4): 1344-1350.

Meadwos, M. P. (1993). *Bacillus thuringiensis* in the environment: ecology and risk assessment,. Bacillus thuringiensis, an environmental biopesticide: theory and practice. P. F. Entwistle, J. S. Cory, M. J. Bailey and H. S. New York, N.Y. , Wiley,: 193-220.

Medigue, C. and I. Moszer (2007). Annotation, comparison and databases for hundred of bacterial genomes. Research in Microbiology 158(724-736).

Minkowski, H. (1910). Geometrie der zahlen, Leipzing und Berling druck und Verlang von B. G. Teubner.

Mizuki, E., T. Ichimatsu, S. H. Hwang, Y. S. Park, H. Saitoh, K. Higuchi and M. Ohba (1999). Ubiquity of *Bacillus thuringiensis* on phyloplanes of arboreous and herbaceous plants in Japan. Journal of Applied Microbiology 86: 976-984.

Moszer, I. (1998). The complete genome of *Bacillus subtilis*: from sequence annotation to data management and analysis. FEBS Letters 430(12): 28-36.

Myasnik, M., R. Manasherob, E. Ben-Dov, A. Zaritsky, Y. Margalith and Z. e. Barak (2001). Comparative sensitivity to UV-B Radiation of two *Bacillus thuringiensis* subspecies and other *Bacillus sp*. Current Microbiology 43(140-143).

Nicholson, W. L. (2002). Roles of *Bacillus* endospores in the environment. Cellular and Molecular Life Sciences 59(3): 410-416.

Nicholson, W. L., A. C. Schuerger and P. Setlow (2005). The solar UV environment and bacterial spore UV resistance: considerations for Earth-to-Mars transport by natural processes and human space flight. Mutation Research/Fundamental and Molecular Mechanisms of Mutagenesis 571(12): 249-264.

Ohba, M. and Y. Aratake (1994). Comparative study of the frequency and flagellar serotypes flora of *Bacillus thuringiensis* in soils and silkworm-breeding environments. Journal of Applied Bacteriology 76: 203-209.

Oyvind, H., D. A. T. Harper and P. D. Ryan (2001). Paleontological Statistics software package for education and data analysis. Paleontologica Electronica 4(1): 9

Pedersen, A. G., L. J. Jensen, S. Brunak, H.-H. Stærfeldt and D. W. Ussery (2000). A DNA structural atlas for *Escherichia coli*. Journal of Molecular Biology 299(4): 907-930.

Pérez-Rueda, E. and J. Collado-Vides (2001). Common History at the Origin of the Position–Function Correlation in Transcriptional Regulators in Archaea and Bacteria. Journal of Molecular Evolution 53(3): 172-179.

Perna, N. T., G. Plunkett, V. Burland, B. Mau, J. D. Glasner, et al. (2001). Genome sequence of enterohaemorrhagic *Escherichia coli* O157:H7. Nature 409(6819): 529-533.

Priest, F. G., M. Barker, L. W. J. Baillie, E. C. Holmes and M. C. J. Maiden (2004). Population Structure and Evolution of the *Bacillus cereus* Group. Journal of Bacteriology 186(23): 7959-7970.

Qin, J., R. Li, J. Raes, M. Arumugam, K. S. Burgdorf, et al. (2010). A human gut microbial gene catalogue established by metagenomic sequencing. Nature 464(7285): 59-65.

Read, T. D., S. L. Salzberg, M. Pop, M. Shumway, L. Umayam, et al. (2002). Comparative Genome Sequencing for Discovery of Novel Polymorphisms in *Bacillus anthracis*. Science 296(5575): 2028-2033.

Rodriguez-Brito, B., L. Li, L. Wegley, M. Furlan, F. Angly, et al. (2010). Viral and microbial community dynamics in four aquatic environments. ISME J 4(6): 739-751.

Rohlf, E. J. (1993). NTSYS-pc numerical taxonomy and multivariate analysis system version 2.02i Seteuket, N.Y., Exeter Software.

Saxena, D., E. Ben-Dov, R. Manasherob, Z. e. Barak, S. Boussiba and A. Zaritsky (2002). A UV Tolerant Mutant of *Bacillus thuringiensis subsp. kurstaki* Producing Melanin. Current Microbiology 44(1): 25-30.

Schloss, P. and J. Handelsman (2005). Metagenomics for studying unculturable microorganisms: cutting the Gordian knot. Genome Biology 6(8): 229.

Schmidt, C. F. (1955). The Resistance of Bacterial Spores with Reference to Spore Germination and its Inhibition. Annual Review of Microbiology 9(1): 387-400.

Schnepf, E., N. Crickmore, J. Van Rie, D. Lereclus, J. A. Baum, J. Feitelson and D. R. Zeigler (1998). *Bacillus thuringiensis* and its pesticidal crystal proteins. Microbiol Mol Biol Rev 62: 775-806.

Schnoes, A. M., S. D. Brown, I. Dodevski and P. C. Babbitt (2009). Annotation Error in Public Databases: Misannotation of Molecular Function in Enzyme Superfamilies. Plos Comput Biol 5(12): e1000605.

Selander, R. K., D. A. Caugant, H. Ochman, J. M. Musser, M. N. Gilmour and T. S. Whittam (1986). Methods of multilocus enzymes electrophoresis for bacterial populations genetics and systematics. Appl Environ Microbiol 51(5): 873-884.

Selander, R. K., R. M. Mckinney, T. S. Whittam, W. F. Bibb, D. J. Brenner, F. S. Nolte and P. E. Pattison (1985). Genetic structure of populations of *Legionella pneumophilla*. Journal of Bacteriology 163(3): 1021-1037.

Sims, G. E. and S.-H. Kim (2011). Whole-genome phylogeny of *Escherichia coli/Shigella* group by feature frequency profiles (FFPs). Proceedings of the National Academy of Sciences 108(20): 8329-8334.

Smith, J. M., N. H. Smith, M. O'Rourke and B. G. Spratt (1993). How clonal are bacteria? Proceedings of the National Academy of Sciences 90(10): 4384-4388.

Sneath, P. H. A. and R. R. Sokal (1973). Numerical Taxonomy -The principles and practice of numerical classification. San Francisco, W. H. Freeman.

Spratt, B. G. and M. C. J. Maiden (1999). Bacterial population genetics, evolution and epidemiology. Philosophical Transactions of the Royal Society B: Biological Sciences 354: 701-710.

Suzuki, H., T. Lefebure, M. J. Hubisz, P. Pavinski Bitar, P. Lang, A. Siepel and M. J. Stanhope (2011). Comparative Genomic Analysis of the *Streptococcus dysgalactiae* Species Group: Gene Content, Molecular Adaptation, and Promoter Evolution. Genome Biology and Evolution 3: 168-185.

Swiecicka, I., K. Fiedoruk and G. Bednarz (2002). The occurrence and properties of *Bacillus thuringiensis* isolated from free-living animals. Letters in Applied Microbiology 34(3): 194-198.

Tettelin, H., V. Masignani, M. J. Cieslewicz, C. Donati, D. Medini, et al. (2005). Genome analysis of multiple pathogenic isolates of *Streptococcus agalactiae*: Implications for the microbial pan-genome. Proceedings of the National Academy of Sciences of the United States of America 102(39): 13950-13955.

The Appache Software Foundation, C. (1989-2004). SPSS. US.

The MathWorks, I. (1984-2008). MATLAB.

Tibayrenc, M., F. Kjellberg and F. J. Ayala (1990). A clonal theory of parasitic protozoa: The population structures of *Entamoeba, Giardia, Leishmania, Naegleria, Plasmodium,Trichomonas, and Trypanosoma* and their medical and taxonomical consequences. Proc Natl Acad Sci USA 87: 2414-2418.

Ticknor, L. O., A. B. Kolsto, K. K. Hill, P. Keim, M. T. Laker, M. Tonks and P. J. Jackson (2001). Fluorescent Amplified Fragment Length Polymorphis Analysis of Norwegian *Bacillus cereus and Bacillus thuringiensis* soil isolates. Appl Environ Microbiol 67(10): 4863-4873.

Tomb, J.-F., O. White, A. R. Kerlavage, R. A. Clayton, G. G. Sutton, et al. (1997). The complete genome sequence of the gastric pathogen *Helicobacter pylori*. Nature 388(6642): 539-547.

Turnbull, P. C. B., N. M. Sirianni, C. I. LeBron, M. N. Samaan, F. N. Sutton, A. E. Reyes and L. F. Peruski (2003). MICs of selected antibiotics for *Bacillus anthracis, Bacillus cereus, Bacillus thuringiensis,* and *Bacillus mycoides* from a range of clinical and environmental sources as determined by the Etest. J Clin Microbiol 42: 3626 - 3634.

Tyson, G. W., J. Chapman, P. Hugenholtz, E. E. Allen, R. J. Ram, et al. (2004). Community structure and metabolism through reconstruction of microbial genomes from the environment. Nature 428(6978): 37-43.

Ussery, D. W., T. M. Wassenaar and S. Borini (2008). Computing for comparative microbial genomics. Germany, Springer.

van Hijum, S. A. F. T., M. H. Medema and O. P. Kuipers (2009). Mechanisms and Evolution of Control Logic in Prokaryotic Transcriptional Regulation. Microbiol. Mol. Biol. Rev. 73(3): 481-509.

Venter, J. C., M. D. Adams, E. W. Myers, P. W. Li, R. J. Mural, et al. (2001). The Sequence of the Human Genome. Science 291(5507): 1304-1351.

Via, M., C. Gignoux and E. Burchard (2010). The 1000 Genomes Project: new opportunities for research and social challenges. Genome Medicine 2(1): 1-3.

Vilas-Boas, G., A. P. S. Peruca and O. M. N. Arantes (2007). Biology and taxonomy *Bacillus cereus, Bacillus anthracis* and *Bacillus thuringiensis*. Can J Microbiol 53: 673-697.

Wagner, W. H. J. (1961). Problems in the classification of terms. Rec Adv Bot 1: 841-844.

Ward, R. D., D. O. F. Skibinski and M. Woodwark (1992). Protein heterozygosity, protein structure, and taxonomic differentiation. Evolutionary Biology. M. K. Hecht. New York, Plenum Press. 26: 73-147.

Welch, R. A., V. Burland, G. Plunkett, P. Redford, P. Roesch, et al. (2002). Extensive mosaic structure revealed by the complete genome sequence of uropathogenic *Escherichia coli*. Proceedings of the National Academy of Sciences 99(26): 17020-17024.

Whittam, T. S., H. Ochman and R. K. Selander (1983). Multilocus genetic structure in natural populations of *Escherichia coli*. Proceedings of the National Academy of Sciences 80(6): 1751-1755.

Williams, J. G., A. R. Kubelik, K. J. Livak, J. A. Rafalski and S. Tingey, V. (1990). DNA polymorphisms amplified by arbitrary primers are useful as genetic markers. Nucleic Acids Res 18(22): 6531-6535.

Wood, D. W., J. C. Setubal, R. Kaul, D. E. Monks, J. P. Kitajima, et al. (2001). The Genome of the Natural Genetic Engineer *Agrobacterium tumefaciens* C58. Science 294(5550): 2317-2323.

Wood, V., R. Gwilliam, M. A. Rajandream, M. Lyne, R. Lyne, et al. (2002). The genome sequence of *Schizosaccharomyces pombe*. Nature 415(6874): 871-880.

Zhu, Y., H. Shang, Q. Zhu, F. Ji, P. Wang, et al. (2011). Complete genome sequence of *Bacillus thuringiensis serovar finitimus* strain YBT-020. J Bacteriol 193(9): 2379-2380.

RESUMEN BIOGRAFICO

Amada Torres Salazar

Candidato para el Grado de

Doctor en Ciencias con Especialidad en Biotecnología

Tesis: Análisis genómico de *Bacillus thuringiensis*: Estructura Poblacional

Campo de Estudio: Biología

Datos Personales: Nacida en la Capital de la Republica Mexica, hija del Sr. David Torres Anguiano y la Sra. Mercedes Salazar González.

Educación: Egresada de la Escuela Nacional de Ciencias Biológicas del Instituto Politécnico Nacional en la generación 1994-1999, destacada académica, política y deportivamente. Obtuvo el tercer lugar de la generación de Químico Bacteriólogo y Parasitólogo. La postulación a la mejor tesis de Licenciatura a Nivel Nacional. Permaneció por siete años consecutivos en el Programa de Formación de Investigadores, COFFA-IPN, en el mismo periodo se vio favorecida por la Beca a la Excelencia Académica del IPN. Líder de grupo y parte del Consejo Estudiantil por cinco años consecutivos. Miembro de la Asociación de Líderes Politécnicos. Capitana de la Selección Ganadora de la Copa Inter-Politécnica de Básquet-ball Femenil y mejor jugadora del año 95-96. Seleccionada Nacional del Equipo Politécnico 96-97. Líder de la Puesta Deportiva del 20 de Noviembre 96 y 97. Entrenadora de las selecciones de la ENCB en la misma disciplina en ambas ramas 97-99.

Experiencia Profesional: Técnico en Investigación del Instituto Mexicano del Seguro Social 2003-2009. Desarrollo Estancias en el área de Bioinformática y Salud Publica en el Instituto de Salud Carlos III de España. En Genética Cuantitativa Avanzada y Tecnologías de Alto-Desempeño en la Southwest Foundation For Biomedical Research, US. Profesor Invitado en las Catedras de Bioinformática UACJ; Farmacogenómica FCQ-UANL; Variación de la Información Genética U Simón Bolívar y Microbiología Veterinaria ENCB-IPN, además de dictar diversos cursos de actualización en las áreas de Biología Molecular Avanzada, Bioinformática, Genética de Poblaciones y Evolución.

PRODUCCION ACADEMICA

ARTICULOS

Torres Salazar A. *Evolución Molecular (Molecular Evolution)*. Ciencia Conocimiento Tecnología 2009, 90: 53-56 Ed. Consejo Nacional de Ciencia y Tecnología de **Nuevo León.** **http://www.conocimientoenlinea.com/content/view/975/216/**

http://www.scribd.com/doc/19625260/Revista-Conocimiento-90

Torres Salazar A, Cerda-Flores RM. *Frequency of Incompatibility for ABO and Rh (D) blood group systems in 96 Mexican Married Couples with Recurrent Spontaneous Abortions and normal karyotype*. Medicina Torreon 2010, 2(3):11-15, ISSN:1405-5422

E. A. Marroquín-Rodríguez, A. García-Moyeda, J. Luna-Guillermo, Y. Flores-Peña, M. A. Cisneros-Estala, J. A. Villarreal-Garza, **A. Torres-Salazar,** H. Torre-Martínez, R. M. Cerda-Flores. *(Prevalence of Malocclusion in a mixed population of the state of Nuevo Leon, Mexico)* Medicina de Torreon 2010, 2(1): 35-40, ISSN:1405-5422

Pinto, L; Thomas E, **A Torres Salazar,** Hoheisel J, and Youns M. *Berberine selectively inhibits cell growth and mediates caspase-indepent cell death in human pancreatic cancer cells.* Planta Medica, 2010, PLAMED-2010-03-0271-OP.R2

Torres Salazar A; M. Youns, J. Hoheisel, and M Wink. Anti-inflammatory and Anti-cancer activities of Essential oils and their biological constituents Int J Clin Pharmacol Ther. 2011; 49:93-5.

Torres-Salazar A. Cerda-Flores R., Pereyra-Alférez B., Parasporins, a new group of bacilli protein against Cancer. 2012 Rev. Medicina de Torreon 20-24

CAPITULOS DE LIBRO

Pereyra-Alférez B., **Torres-Salazar A.,** Orrantia-Borunda E., Galán-Wong L. J. *Biolixiviacion en Genomas y Proteomas del Siglo XXI*: Biotecnología Ambiental, (Bioleaching in genomics and proteomics of the Century XXI: Environmental Biotechnology) Edit La Universidad Autónoma de Coahuila 2008

Cerda-Flores R M, **Torres Salazar A**., Silva-Martínez L E, Rodríguez-Vela H, Marty-Gonzalez L F., Jin L, Barton S A., Chakraborty R. *mtDNA haplotypes in mexican mestizos whose grandmothers were born in Cuatrocienegas, Coahuila during the cohort 1882-1919.* III Simposio Internacional El Hombre Temprano en America. Editores: J. C. Jiménez et al., México: UNAM, Instituto de Investigaciones Antropológicas; INAH: Museo del Desierto, 45-52, 2010. ISBN: 978-607-02-1947-4
http://www.iia.unam.mx/catalogoPublic/detalles.php?clave=377

Pereyra-Alférez B., Espino Vazquez A.N., **Torres-Salazar A.** ***Bacillus thuringiensis*** Edit. La Universidad Autónoma de Nuevo León 2012(acepted))

Cerda-Flores, R., **Torres Salazar A,** S. A. Barton, R. Chakraborty. *13 STR LOCI IN MEXICAN MESTIZOS WHOSE 4 GRANDPARENTS WERE BORN IN CUATROCIENEGAS, COAHUILA DURING THE COHORT 1882-1919.* Instituto Nacional de Antropología e Historia (acepted)

Rodríguez García A., **Torres-Salazar A., Cáncer de la Cavidad Oral** Edit. La Universidad Autónoma de Nuevo León (acepted)

Bernis, F. 1997. La Clase Aves: un recorrido biológico por la Taxonomía.

Editorial Complutense. España. 193 pp.

Técnicas y métodos son, pues, armas de dos filos. Esto, insistimos, se nota mejor en los novicios y cadetes de la investigación, aunque también en profesionales pertinaces. En lo personal surge el peligro del encasillamiento, de la perdida de la perspectiva y del jibarismo o mentecatismo del presunto investigador. Ves cómo dicha persona, una vez pertrechada de unos pocos métodos y técnicas, no sale ya en toda su vida de ese corsé, y puede conjugar una excelente habilidad y manipulación técnica y metódica con la más enteca y estúpida visión del resto del mundo físico y del mundo humano. Y no rara vez, en esta limitación personal, se confunde algoritmo con realidad, es decir, la cifra, el índice o el estadístico, con el ser o el proceso donde se aplicaron. Porque no conviene olvidar que la misma formula o ecuación, con sus resultados numéricos, puede aplicarse indistintamente a la comercialización del cultivo de tomates o a la manipulación de un proceso molecular