INSTITUT PASTEUR
DE MONTEVIDEO

Tesis de Maestría en Ciencias Biológicas
PEDECIBA Biología
Subárea Genética

Abril 2018

Genómica comparativa de *Mycobacterium bovis*: aproximaciones epidemiológicas y filogenéticas

Lic. Moira Lasserre
Unidad de Biología Molecular
Institut Pasteur de Montevideo

Orientadora - Dra. Luisa Berná
Co-orientador - Dr. Carlos Robello

Índice

ABREVIACIONES

ADN – Ácido Desoxirribonucleico

bTB – Tuberculosis bovina

DR – Repetidos Directos

Indel – Inserción o deleción de nucleótidos.

MIRU – *Mycobacterial Interspersed Repetitive Unit*

MIRU-VNTR – MIRU *Variable Number Tandem Repeat*

MTBC – Complejo *Mycobacterium tuberculosis*

PGRS – *Polymorphic GC Repeat Sequence*

RD – Región de Diferencia

RFLP – *Restriction fragment length polymorphism*

SNP – *Single nucleotide polymorphism*

RESUMEN

La tuberculosis bovina es una enfermedad que produce graves riesgos para la economía y bienestar animal. *Mycobacterium bovis*, su agente etiológico, pertenece a un grupo de organismos genéticamente monomórficos denominado complejo *Mycobacterium tuberculosis* (MTBC) que se caracterizan por presentar una muy alta identidad nucleotídica (99.9%). Este rasgo dificulta la tipificación de cepas y determinación de características únicas dependientes de su diversidad nucleotídica.

Uruguay, como país que ha dependido históricamente de la producción de ganado para su sustento económico, se ve directamente afectado por la prevalencia de *M. bovis* en el país. Su alta tasa de ganado *per capita* y producción intensiva de ganado ayudan a que este sea un ambiente apropiado para evaluar la variabilidad genómica entre diversas cepas. Al mismo tiempo, a nivel mundial hay escasa información sobre los diferentes genotipos reconocidos en este patógeno, y sobre su estructura poblacional. Históricamente se han identificado los complejos clonales European 1, European 2, African 1 y African 2. Pero no existen estudios sobre la estructura poblacional interna de los mismos. Esto representa una limitante para la búsqueda de estrategias de genotipado rápidas y accesibles.

Las dos problemáticas mencionadas en el párrafo anterior fueron estudiadas en este trabajo. Por un lado, comparamos 186 genomas de MTBC de origen mundial y 23 nuevos genomas de *M. bovis* aislados en Uruguay. A pesar de presentar una estructura genómica conservada, observamos que la población global de *M. bovis* es altamente estructurada y se evidenciaron tres grupos existentes entre las cepas uruguayas. A partir de esto, se identificaron las principales fuentes de variabilidad genómica entre estas cepas: regiones de diferencia (RD), genes variables, duplicaciones y genes nuevos. Estos resultados resaltan la existencia de una variabilidad genética intraespecífica mayor a la esperada. Por otro lado, obtuvimos la estructura filogenética resultante de analizar los polimorfismos de sitio único (SNP) de los genomas disponibles en el momento (n = 1.238). Se identificaron dentro de *M. bovis* 5 grandes linajes y correspondientes sub linajes, aportando un grado de resolución no descrito previamente. A partir de esta nueva clasificación nos enfocamos en la búsqueda de un set mínimo de SNPs informativos que reconstruyeran esta diversidad encontrada. Obtuvimos un set final de 56 SNPs que fue corroborado satisfactoriamente durante la clasificación de cepas control, el cual esperamos sea utilizado en el futuro para la caracterización rápida de nuevas cepas.

Estos estudios contribuyen a una mayor comprensión de la estructura poblacional y clasificación de *M. bovis*.

INTRODUCCION

Complejo *Mycobacterium tuberculosis*

El complejo *Mycobacterium tuberculosis* (*Mycobacterium tuberculosis* complex, MTBC) pertenece al phylum *Actinobacteria*, clase *Actinobacteria*, orden *Actinomycetales*, suborden *Corynebacterineae*, familia *Mycobacteriaceae*, género *Mycobacterium* [1]. Este género está compuesto por más de 170 especies. Este complejo se caracteriza por causar tuberculosis en diferentes hospederos mamíferos e incluye varias especies: *M. canetti* [2], *M. tuberculosis* [3], *M. africanum* [4], *M. mungi* [5], el bacilo Dassie [6], *M. suricattae* [7], *M. orygis* [8], *M. microti* [9], *M. pinnipedii* [10], *M. caprae* [11], *M. bovis* [12] y *M. bovis* BCG (bacilo Calmette-Guérin) [13]. Estas especies presentan claras diferencias epidemiológicas que incluyen preferencia de hospedero, fenotipos y patogenicidad [14], pero a su vez se encuentran muy cercanamente relacionadas y ocasionan patologías similares [15,16].

Caracterización molecular del Complejo *Mycobacterium tuberculosis*

Las micobacterias del MTBC se caracterizan por compartir un alto grado de homología nucleotídica, presentando un 99.95% de identidad y una secuencia idéntica del rRNA 16S [15,16]. Se lo considera un grupo genéticamente monomórfico ya que el intercambio de material genético por transferencia horizontal de genes es un evento muy raro entre estas especies [17], y su diversidad actual refleja una evolución principalmente clonal resultado de un reciente cuello de botella evolutivo.

El tipificado de las cepas es imprescindible para poder realizar estudios epidemiológicos, de incidencia y de variabilidad de las diferentes cepas, y permite estudiar los factores que causan la diseminación y distribución geográfica de una enfermedad. Como en todas las enfermedades infecciosas, es clave dilucidar las rutas de transmisión y diseminación en el ambiente.

Tradicionalmente, la identificación de los miembros del MTBC se basaba en el estudio de características fenotípicas evidenciadas en base a condiciones de cultivo o pruebas bioquímicas. Pero estas pruebas tienden a ser lentas, necesitan suficiente crecimiento bacteriano y no proporcionan una distinción de especies precisa [18]. Posteriormente, avances en biología molecular permitieron la introducción de técnicas basadas en el estudio del ADN de estas bacterias [19–22]. La alta conservación entre todos los miembros del MTBC permite que las técnicas desarrolladas para la caracterización molecular de una especie del complejo se puedan extender a las otras.

A pesar de la alta homogeneidad genética mencionada anteriormente, la secuenciación de los genomas de *M. bovis* y *M. tuberculosis* permitió reconocer varias regiones altamente polimórficas debido a variaciones en la cantidad de copias y/o posición en el genoma [19,20,23–25]. Estas regiones polimórficas pueden corresponderse con genes que confieren una ventaja selectiva sobre otras bacterias [26] o localizarse en regiones intergénicas como secuencias de inserción o repetidos de una misma secuencia [20]. Regiones variables tales como elementos genéticos móviles, presencia o ausencia de regiones del genoma (RDs, Regiones de Diferencia) [21], repetidos con un alto contenido en G+C (*polymorphic GC repeat sequence*, PGRS) [27,28], regiones de repetidos directos (DR) [29], repetidos en tándem en número variable (VNTR) [22] y *single nucleotide polymorphisms* (SNPs) [22,30,31] han sido explotadas para el desarrollo de métodos de tipificación más precisos que permitan un mayor entendimiento de la epidemiología, evolución y estructura poblacional de estos organismos [32].

El tipificado de MTBC usualmente se realiza para uno de los siguientes fines: 1) – epidemiología molecular clásica, 2) – estudios evolutivos y filogenéticos y 3) – clasificación de cepas. Pero las técnicas existentes de genotipado no se ajustan de igual manera para todas estas aplicaciones. A continuación, se describen en detalle aquellas más utilizadas en la tipificación de *M. bovis* y que tienen mayor relevancia para esta tesis.

Spoligotyping

El spoligotyping (*spacer oligonucleotide typing*), ha sido uno de los métodos epidemiológicos más populares y estandarizados desarrollados para el genotipado de MTBC [33]. Se basa en el polimorfismo existente en el locus DR. Este pertenece a la familia CRISPR de ADN repetitivo [34], y se compone por una serie de repetidos directos conservados de 36 bp entremezclados con secuencias únicas de 34-41 bp denominadas secuencias espaciadoras (*spacers*). Un repetido directo y su espaciador adyacente se denomina *direct variant repeat* (DVR). Debido a que las cepas varían en el número de DVR, la presencia o ausencia de los espaciadores puede ser utilizada como forma de tipificado. En spoligotyping, se amplifica el locus DR en su totalidad, utilizando primers inversos complementarios a la secuencia de repetidos directos cortos. Los productos de PCR se hibridan a una membrana que contiene 43 oligonucleótidos representativos de los espaciadores que han sido identificados en *M. tuberculosis* H37Rv y *M. bovis* BCG (Figura 1). Hasta el momento, se han identificado 94 espaciadores distintos, pero solo los 43 originales se usan rutinariamente ya que los demás solo aumentan levemente la resolución del método. Las cepas se diferencian entonces por el número de espaciadores que faltan del set completo de 43. El genotipo resultante se obtiene en formato binario (1 = presencia del espaciador, 0 = ausencia de espaciador), lo cual

hace que los resultados puedan ser fácilmente interpretados, computarizados y comparados entre diferentes laboratorios [35,32].

Se ha reconocido que esta técnica permite la identificación de MTBC hasta el nivel de subespecie y es incapaz de distinguir al nivel de cepas. Tampoco es capaz de detectar infecciones mixtas, ya que el patrón obtenido podría corresponder a la sumatoria de espaciadores de las cepas involucradas [36]. Además se ha visto que presenta un menor nivel de discriminación que otros genotipados tradicionales (como IS6110-RFLP) [32,37]. Aun así, el spoligotyping puede ser efectivamente usado para la diferenciación de cepas con bajas copias de la secuencia de inserción IS6110, como es el caso de *M. bovis* y *M. caprae*. Por lo tanto, esta técnica es la más utilizada en laboratorios de caracterización de aislados de M. bovis [38]. En micobacterias no tuberculosas no produce ninguna señal, evidenciando la especificidad que tiene este método para MTBC. Por último, es un método sencillo al basarse en PCR, es costo-efectivo, con resultados rápidos y reproducibles.

Es aceptado que los patrones de spoligotipado evolucionan por pérdida de secuencias espaciadoras que, en organismos clonales, no puede ser reparada mediante recombinación y permanecen ausentes en ese linaje para siempre. La frecuencia con la que estos espaciadores se pierden los hace buenos marcadores para estudios epidemiológicos, haciendo que esta técnica sea utilizada rutinariamente para trazar cadenas de transmisión de tuberculosis. Sin embargo, recientemente ha sido catalogada como problemática junto con MIRU-VNTR, dada su alta propensión hacia la evolución convergente y homoplasia [39]. Por lo tanto, en los últimos años se ha considerado de uso limitado para análisis filogenéticos y genética de poblaciones [39–42], proponiéndose además que, para estudios epidemiológicos, spoligotyping sea efectuado como test primario seguido de otro método de mayor poder discriminativo.

MIRU-VNTR

El genotipado de VNTR está basado en el análisis de múltiples *loci* de repetidos en tándem que muestran un gran polimorfismo en el número de copias de estos repetidos (*variable number tandem repeat*, VNTR) [43]. Originalmente identificado en eucariotas, *M. tuberculosis* fue una de las primeras especies bacterianas donde se localizaron estos *loci* de repetidos en tándem. En el

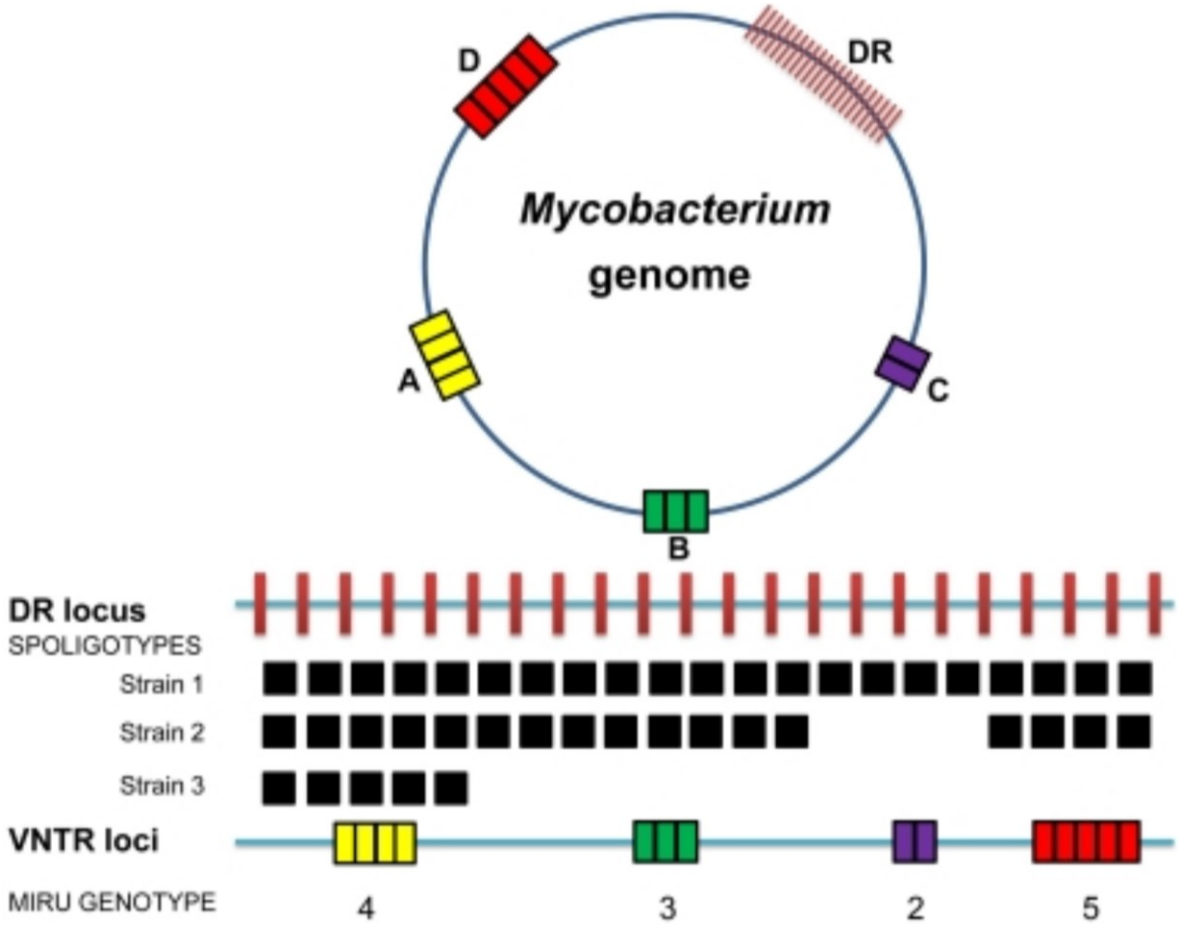

Figura 1-. Representación del genoma de un aislado del MTBC hipotético, mostrando las regiones utilizadas para *spoligotyping* y MIRU-VNTR. *Spoligotyping* está basado en la detección de espaciadores únicos en el *locus* de repetidos directos en los genomas de MTBC. Los patrones de *spoligotyping* se representan mediante cuadrados negros y blancos que indican la ausencia y presencia de espaciadores, respectivamente. La deleción de alguno de los 43 espaciadores estudiados permite diferenciar entre subespecies. MIRU-VNTR depende de la identificación de diferentes números de repetidos en varios *loci* a lo largo del genoma (A, B, C y D). El número de repetidos en cada *locus* se combina generando un código numérico utilizado para establecer relaciones filogenéticas y epidemiológicas entre cepas. Modificado de Comas et al., 2009 [39].

genoma de los miembros del MTBC existen numerosos loci VNTR identificados y, particularmente en *M. bovis*, el análisis de VNTR brinda una mayor resolución que sólo utilizando spoligotyping [44]. Se trata de una técnica basada en PCR donde se amplifican los *loci* VNTR elegidos con *primers* específicos para sus regiones flanqueantes (Figura 1). El número de unidades repetidas en cada *locus* se determina en base a la estimación del tamaño de los amplicones y al tamaño conocido de cada unidad de repetidos en el *locus* VTNR de estudio [24]. Se trata de un método barato, rápido y simple de realizar, con resultados no ambiguos y reproducibles. Los resultados se expresan en un formato simple, en donde cada digito representa el número de copias en un *locus* en particular. Para la interpretación de los resultados se utiliza una tabla de alelos que asocia

los diferentes tamaños de banda a su correspondiente número de repetidos para cada *locus*.

Inicialmente se reportaron 6 VNTR *loci* para *M. tuberculosis* (e*xact tandem repeats* ETR-A, -B, -C, -D, -E y -F) [45]. Pero la capacidad discriminativa con estos *loci* no sobrepasaba la alcanzada con IS6110-RFLP o *spoligotyping*. Con la culminación del proyecto genoma de *M. tuberculosis* H37Rv [46], se identificaron nuevos *loci* VNTR. Una clase específica de estos nuevos elementos son los MIRUs (*mycobacterial interspersed repetitive units*) [47]. Los mismos se describieron como repetidos en tándem de 46-101 bp diseminados en 41 *loci* en todo el cromosoma de *M. tuberculosis* H37Rv. Desde que se incorporaron los MIRUs la tipificación de micobacterias por VNTR se conoce como MIRU-VNTR. Originalmente se presentó un set de 12 *loci* como estándar [22], pero con limitaciones en cuanto a su poder discriminatorio [48]. En el año 2006, Supply y colaboradores propusieron un nuevo estándar de 15 *loci* para discriminación epidemiológica de rutina, así como un sistema de 24 *loci* como herramienta de mayor resolución para estudios filogenéticos [24]. Ambos han sido utilizados ampliamente desde entonces, demostrando su poder de resolución [49–52]. Los estudios realizados con aislados de *M. bovis* son más escasos, aunque se ha realizado tipificación por MIRU-VNTR en varios países para estudios epidemiológicos [44,53–55].

SNP-typing

Los *single nucleotide polymorphisms* (SNPs) se clasifican en dos grandes grupos: sinónimos y no sinónimos. Estos últimos provocan cambios aminoacídicos en las proteínas si se localizan en regiones codificantes, y por ende son frecuentemente evaluados para el estudio de los mecanismos subyacentes a procesos biológicos. Por el contrario, los cambios sinónimos producen alteraciones que no modifican la secuencia primaria de las proteínas y originalmente fueron considerados fenotípicamente neutrales. Actualmente se ha demostrado que algunos de estos cambios pueden alterar la estructura, función y nivel de expresión proteico (SNPs sinónimos funcionales). Si bien no son en su totalidad evolutivamente neutrales, las mutaciones sinónimas en su conjunto están sujetas a una menor presión selectiva que las mutaciones no sinónimas, por los que se las utiliza para estudios filogenéticos y de genética de poblaciones [56–58].

El SNP-*typing* (o *genotyping*) es un tipo de genotipado basado en la evaluación de SNPs previamente conocidos. Los SNPs son los marcadores filogenéticos más confiables para el estudio evolutivo del MTBC ya que poseen niveles despreciables de homoplasia [39]. A pesar de esto, debido a sesgos de muestreo (p*yhlogenetic discovery bias*), la información obtenida en un ensayo de SNP-typing es altamente dependiente de la estrategia utilizada para identificar el set de SNPs, y este es comúnmente obtenido al comparar un número limitado de cepas o cepas que no

representan la diversidad global existente [59].

Se han propuesto varias metodologías de SNP-*typing*, incluyendo métodos apropiados para laboratorios sin acceso a plataformas de secuenciación [60]. Muchos de estos han sido empleados para MTBC [61–67], y se han sugerido también paneles de SNPs para clasificar cepas dentro de los linajes principales del MTBC [39,41,68,69]. La clasificación de linajes en base a un set de SNPs dentro de *M. bovis* no ha sido puesta a punto hasta el momento.

Evolución del Complejo *Mycobacterium tuberculosis*

Durante mucho tiempo fue aceptada la teoría que *M. tuberculosis* se originó a partir de *M. bovis* por transmisión zoonótica del ganado al ser humano. La misma era consistente con la existencia de un "cuello de botella" evolutivo en el momento de la especiación del complejo, que habría ocurrido entre 15.000 a 20.000 años atrás y explicaría la baja diversidad genética del mismo [16]. Este número coincide con la datación de la domesticación del ganado (8.000 – 10.000 años atrás) [70]. Sin embargo, está teoría quedó invalidada luego de la publicación de los genomas de *M. tuberculosis* [46] y de *M. bovis* [71]. En efecto, diversos estudios demostraron la disimilitud de varias regiones genómicas en los miembros del MTBC mediante técnicas de genómica comparativa, introduciendo así el concepto de RDs [21,72,73]. Al no exhibir en el complejo evidencias de transferencia horizontal de genes, se postuló a estas regiones como polimorfismos de secuencia larga (LSPs), ya sean inserciones o deleciones, que representan eventos evolutivos únicos difícilmente reversibles, y por ende toda la progenie de una cepa que haya perdido una región heredará esta deleción. A partir de esto, se hipotetizó que la especiación de MTBC ocurrió en base a la pérdida de RDs y a la paralela aparición de polimorfismos. La filogenia resultante modificó el paradigma original de evolución de MTBC cuestionando la hipótesis del origen reciente de la tuberculosis humana, como se ilustra en la Figura 2, donde *M. tuberculosis* se sitúa en una posición ancestral y con un genoma más extenso, mientras que *M. bovis* se sitúa en una posición derivada. Actualmente se acepta que las cepas animales (*M. mungi*, bacilo *Dassie*, *M. suricattae*, *M. orygis*, *M. microti*, *M. pinnipedii*, *M. caprae*, *M. bovis*) divergieron de cepas humanas. Estudios posteriores resolvieron además que la emergencia del MTBC ocurrió en África hace aproximadamente 70.000 años [74]. Esto es reforzado por la restricción geográfica al cuerno de África del grupo más cercano a su ancestro común, *M. canetti*, así como por la restricción de *M. africanum* a África Occidental y la presencia también en este continente de los siete linajes descritos para *M. tuberculosis* [75].

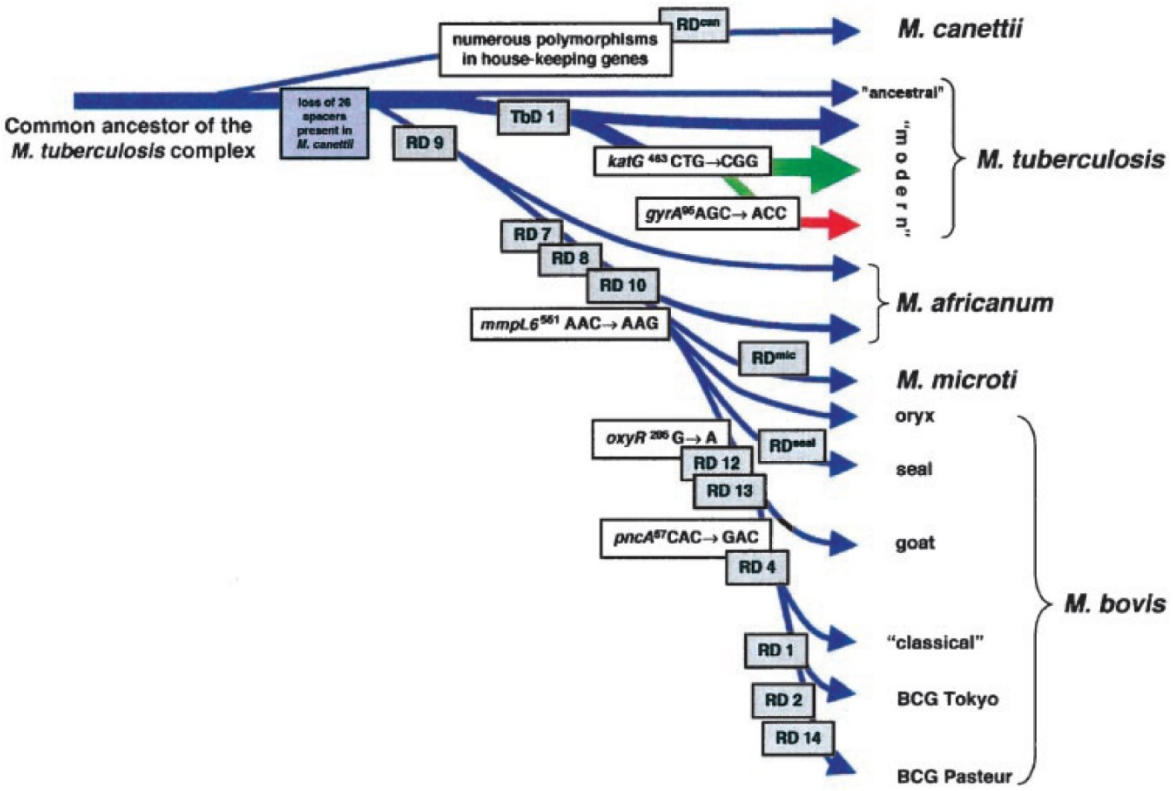

Figura 2- Esquema evolutivo propuesto para los organismos del MTBC en base a sucesivas deleciones de regiones genómicas en ciertos linajes (cajas grises) y polimorfismos. Tomado de Brosch et al., 2002 [76].

El uso de RDs como marcadores filogenéticos robustos ha sido demostrado en múltiples ocasiones desde entonces [17,76,77]. Numerosos estudios sobre las regiones de diferencia permitieron una mejor identificación de los linajes de MTBC. Por ejemplo, las cepas originadas de la cepa atenuada *M. bovis* BCG carecen de la región RD1. De la misma manera, las cepas adaptadas a hospederos animales no humanos forman un linaje marcado por la ausencia de una región denominada RD9 [76]. Estos estudios iniciales de RDs aportaron un panorama general sobre el grado de variabilidad y conservación de algunas regiones genómicas en un gran conjunto de organismos del MTBC [76].

La importancia del uso de *spoligotyping* como forma de tipado molecular en cepas adaptadas a animales fue destacada en el 2006 por Smith y colaboradores [35,78]. La presencia o ausencia de algunos espaciadores puede ser utilizada como marcador filogenético de manera similar a la pérdida de RDs en aquellos linajes adaptados a animales que surgen luego de la pérdida de RD9 [53], complementando así la filogenia propuesta anteriormente. Efectivamente, el conjunto de patrones tanto de spoligotipos, de mutaciones puntuales (SNPs) como de RDs permitieron definir una serie de clados anidados que presentan asociación a hospederos particulares (Figura 3). Así,

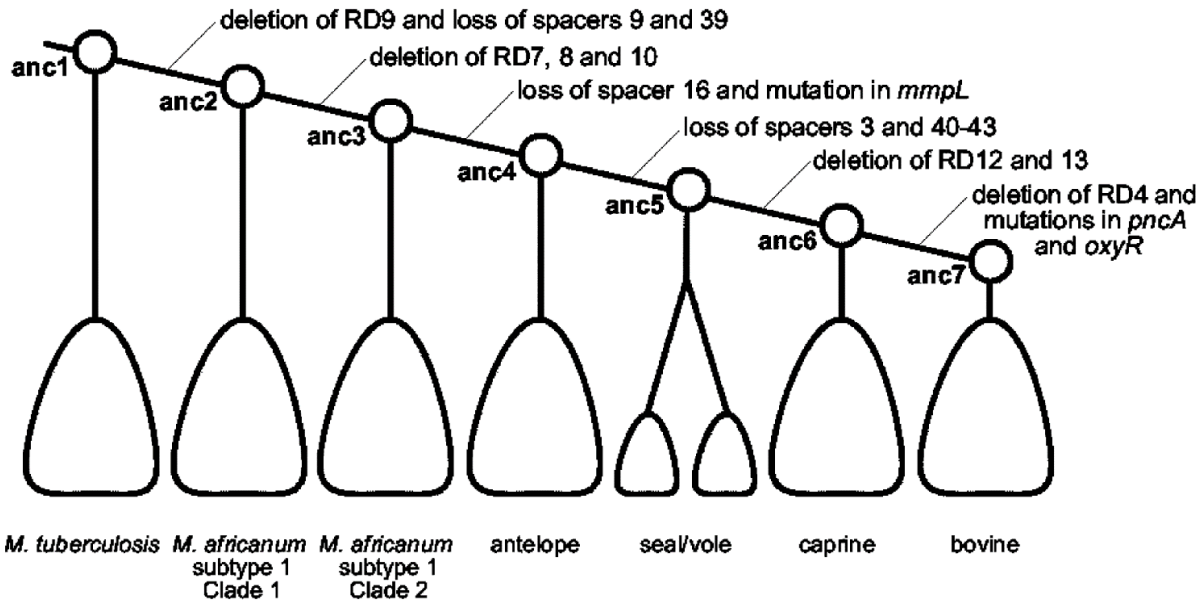

Figura 3- Filogenia del linaje resultante de la deleción de RD9 junto con *M. tuberculosis*. Se observan todas las deleciones informativas (RD), SNPs y deleción de espaciadores que Smith et al., 2006 [53] utilizaron para definir los clados.

Smith y colaboradores introducen la noción de ecotipos para los miembros del MTBC. Un ecotipo evoluciona a partir de que una cepa se vuelve inmune a eventos de selección que afectan a la población ancestral [53,79]. Cada ecotipo tiene una preferencia de hospedero distintiva que representa su nicho, y presenta diferencias moleculares comunes dentro de cada grupo. Hasta este momento se consideraba a *M. bovis* como generalista (capaz de infectar diversos tipos de hospederos), y las aproximaciones experimentales buscaban entender los rasgos genéticos o mecanismos que hacen a *M. bovis* capaz de infectar diversos tipos de hospederos o aquellos que restringen a *M. tuberculosis* a hospederos humanos. Por el contrario, el concepto de ecotipos promovió priorizar la identificación de los rasgos genéticos que controlan los diferentes eventos de adaptación a hospederos [53].

Estructura poblacional de *M. bovis*

La población actual de *M. bovis* es el resultado de una serie de expansiones clonales donde los clones adquieren altas frecuencias entre la población [80], haciendo que la misma se vuelva muy estructurada en vez de uniforme, y se compone de un conjunto de complejos clonales de mayor o menor tamaño. Los complejos clonales se definen como grupos de cepas descendientes de una única célula, la cual representa el ancestro común más reciente del complejo (MRCA). Todos los miembros de un complejo clonal comparten marcadores moleculares que definen su pertenencia al grupo.

Los marcadores moleculares de preferencia son los RDs ya que son estables e improbables de

aparecer independientemente en diferentes linajes, aunque es posible recurrir a SNPs si no se pueden identificar RDs característicos. Como sondeo preliminar, actualmente se acude al spoligotipado para identificar miembros de un complejo clonal [81], ya que todas las cepas que pertenecen a un complejo van a presentar patrones de spoligotipo derivados del encontrado en el MRCA.

Hasta el momento se han definido cuatro complejos clonales en *M. bovis*: European 1 (Eu1) [82], European 2 (Eu2) [83], African 1 (Af1) [84] y African 2 (Af2) [85]. Eu1 se caracteriza por tener una distribución global, presentándose en altas frecuencias en las islas británicas, en algunas antiguas colonias británicas y en América con excepción de Brasil, donde se encuentran en menor frecuencia (<10%). También se localizan en menor frecuencia en España y Portugal. Eu1 presenta una deleción de 806bp en el gen *tre*Y (RDEu1) y pérdida del espaciador 11. Por su parte, Eu2 está predominantemente en la Península Ibérica; presenta frecuencias altas en España y Portugal, y bajas en Francia e Italia. Se caracteriza por la ausencia del espaciador 21 y un SNP en el gen *gua*A. Los representantes de Af1 se han encontrado en alta frecuencia únicamente en África Centro-Oeste. Se caracterizan por la pérdida del espaciador 30 y la deleción RDAf1 (5.3kb). Por último, Af2 es dominante en África Este, presenta una deleción de 14.1kb (RDAf2) y ausencia de los espaciadores 3 al 7.

Tuberculosis bovina

Mycobacterium bovis es el agente etiológico de la tuberculosis bovina (bTB), una enfermedad crónica infecciosa respiratoria que incide sobre el ganado vacuno y una amplia variedad de mamíferos susceptibles. La misma está entre las enfermedades de ganado más importantes debido a su impacto socioeconómico negativo que afecta las actividades agropecuarias y el rendimiento animal, repercutiendo sobre la producción de leche y carne y la comercialización de sus subproductos [86,87].

En Uruguay existe una alta incidencia de bTB en ganado que produce pérdidas importantes en la producción de leche y carne. Entre los años 2012 y 2017 se han reportado 133 nuevos focos de bTB en establecimientos ganaderos, con un promedio de 22 por año [88]. Durante años, el control de la bTB en ganado se ha logrado mediante la aplicación de programas sistemáticos de diagnóstico y eliminación del reactivo, monitoreo de frigoríficos, restricciones en el movimiento de ganado y en ocasiones sacrificio de rodeos enteros frente a focos en donde no se puede establecer la fuente de infección. Si bien la sospecha o presencia de esta enfermedad es de denuncia obligatoria, su prevalencia aún no tiene una explicación epidemiológica clara. Los casos reportados de bTB en

nuestro país presentan características fenotípicas diversas, con prevalencia, infectividad y virulencia variadas. El estudio de la variabilidad genómica, sumado a la información de los genotipos presentes en nuestro país, permitirá conocer la epidemiología de las cepas responsables de los casos de tuberculosis bovina reportados en el Uruguay.

Patogénesis y ciclo de vida del agente infeccioso

La bTB se caracteriza por el desarrollo lento y progresivo de granulomas o tubérculos en los tejidos afectados, siendo las principales lesiones formadas en los ganglios linfáticos y pulmones. El bacilo es transmitido por otro hospedero con la enfermedad en estado activo e infecta a su hospedero mamífero en los tejidos mencionados [89]. La transmisión se produce principalmente por vía aerógena, y la exposición a aerosoles es considerada la ruta más frecuente para la infección [90]. Sin embargo, la infección por la ingesta de material contaminado también puede ocurrir.

En algunas ocasiones la bacteria permanece en estado latente en el organismo hospedero sin desencadenar la enfermedad. Los diferentes estadios de la infección son reflejo del balance entre el bacilo y los mecanismos de defensa del hospedero [91].

Naturalmente, se conoce con mayor detalle el ciclo de vida y factores de virulencia del patógeno humano, *M. tuberculosis*, pero la mayoría de los descubrimientos son directamente aplicables también a *M. bovis*. El ciclo de vida de este bacilo se resume en la entrada al hospedero, obtención de un nicho para desarrollarse, multiplicación y salida del hospedero (Figura 4). Este proceso lo logra evadiendo, modulando y explotando el sistema inmune del hospedero [92].

Los macrófagos son la célula hospedera primaria para el crecimiento intracelular de *M. bovis* y la mayoría de las micobacterias patogénicas; luego de la deposición del bacilo en la superficie respiratoria, se reclutan macrófagos que serán infectados y funcionarán como transporte de la bacteria a tejidos más profundos. Las especies del MTBC han desarrollado numerosos mecanismos para evitar el ambiente hostil del macrófago, tales como inhibir la fusión lisosoma-fagosoma y escapar de ambientes ácidos dentro del fagolisosoma [93]. Sucesivamente se desarrolla el granuloma: un agregado de macrófagos y otras células inmunitarias. En etapas tempranas, el granuloma es en cierta forma responsable de expandir la infección permitiendo a las bacterias diseminarse hacia nuevos macrófagos. La fagocitosis asegura la interacción con células del sistema inmune innato y adquirido del hospedero y, a medida que se desarrolla la inmunidad adaptativa, el granuloma podrá finalmente restringir el crecimiento bacteriano. Pero en muchas circunstancias, los macrófagos que conforman este granuloma pueden comenzar un proceso de necrosis, formando un núcleo necrótico que volverá a permitir el crecimiento

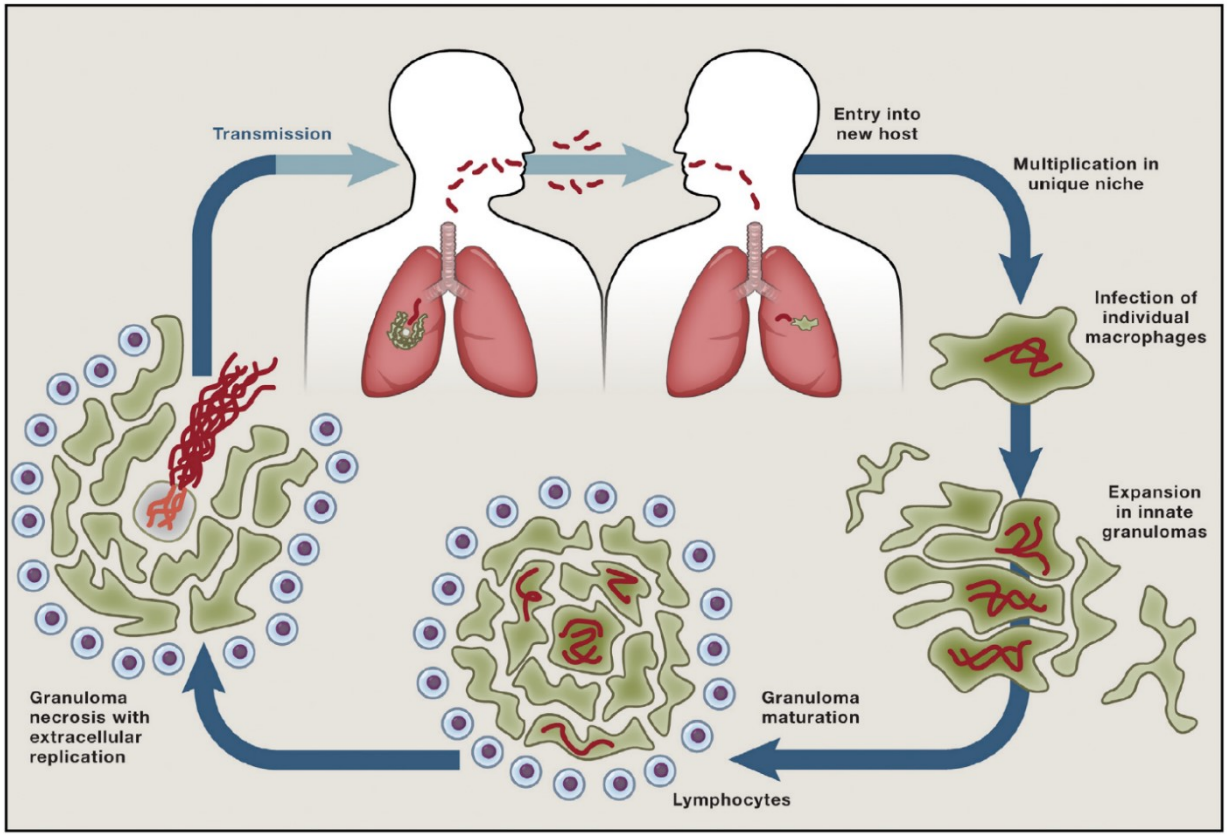

Figura 4- Ciclo de vida e infección de M. tuberculosis. Tomado de Cambier et al., 2014 [92].

bacteriano y luego su transmisión a nuevos hosederos (Figura 4) [94].

A pesar de que la mayoría de los animales con infección latente no mueren de tuberculosis, el mayor peligro reside en la reactivación y subsecuente transmisión por contacto cercano.

La bTB es más frecuente en vacas lecheras que en ganado cárnico debido a que las vacas comparten tiempo en el tambo, así como los mismos bebederos y en muchos casos se produce hacinamiento. Por otro lado, y tienen más tiempo de exposición a la infección e incubación de la enfermedad dado a que alcanzan una mayor edad [95]. Sin embargo, los rodeos de producción cárnica pueden verse afectados con alta morbilidad si entran animales infectados y comparten el mismo bebedero. La principal vía de ingreso de la enfermedad a un rebaño es la introducción de bovinos enfermos o portadores de *M. bovis*. Hay un alto riesgo de nuevas infecciones cuando se da un estrecho contacto entre los animales, principalmente en ganadería intensiva [87]. En general, cuanto mayor sea la población del establecimiento, mayor es la probabilidad de introducción y persistencia del bacilo. En humanos, la infección de *M. bovis* puede ocurrir por vía aerógena o alimentaria, ya sea por contacto cercano con ganado o humanos tuberculosos como por el consumo de leche contaminada con el bacilo, respectivamente [89].

Diagnóstico

El diagnóstico actual de bTB se basa en la detección directa del agente etiológico en el caso de una infección activa y la detección indirecta por relevamiento de la respuesta inmune resultante de la infección para infecciones latentes. Para la enfermedad activa, la detección directa se realiza por cultivo bacteriológico, siendo el *gold standard* para la confirmación de la infección en un establecimiento. Cuando se presume infección latente, las micobacterias no son directamente detectables y, por lo tanto, las pruebas diagnósticas miden la presencia de una respuesta inmune adaptativa contra las mismas. La dificultad de esta estrategia es que la detección de una respuesta podría representar exposición sin infección o infección previa. La prueba de tuberculina es hasta el momento el principal método recomendado para el diagnóstico de infección latente [96]. La misma implica la inyección intradermal de un extracto purificado de proteínas de M. *tuberculosis*, M. *bovis* y/o M. *avium*. Esto resulta en una respuesta de hipersensibilidad retardada que produce un endurecimiento cutáneo en el lugar de inyección en las 48 a 72 horas siguientes.

En los reaccionantes positivos se debe realizar una prueba comparada utilizando dos inóculos: un extracto proteico de M. *bovis* y otro de M. *avium*. La prueba se considera positiva a M. *bovis* si el extracto de M. *bovis* produce un engrosamiento cutáneo mayor a la reacción producida por el extracto de M. *avium*. En este caso el bovino es sacrificado y todo el rebaño queda interdicto, prohibiendo en el movimiento de ganado en todo el predio.

Estas pruebas tienen de un 10 a 25% de falsos negativos que pueden perpetuar la enfermedad en el rebaño, y de 1 a 5% de falsos positivos [97], pudiéndose presentar errores en la interpretación por parte del operador, reacciones cruzadas con micobacterias ambientales y/o anergia por infecciones recientes [87].

PARTE 1

La primera parte de este trabajo está destinada a la tipificación y caracterización genómica de cepas uruguayas de *M. bovis*. Se realizó un análisis de la estructura poblacional conformada por 23 cepas uruguayas de *M. bovis* y una posterior búsqueda de aquellas características genómicas que explicaran la estructuración observada. A partir de los resultados obtenidos, elaboramos un trabajo que fue publicado en BMC Genomics.

HIPOTESIS

La caracterización de cepas de *M. bovis* responsables de los casos de tuberculosis bovina en Uruguay y posterior secuenciación permitirá profundizar en el conocimiento del genoma de esta especie.

OBJETIVOS GENERALES

Caracterización de cepas de *M. bovis* uruguayas y contextualización a nivel global para el estudio de la estructura poblacional de este patógeno en Uruguay.

OBJETIVOS ESPECÍFICOS

1. Genotipado de cepas uruguayas de *M. bovis* mediante spoligotyping.
2. Evaluación tanto de la variabilidad como de características genómicas comunes entre las cepas seleccionadas.
3. Reconstrucción filogenética de las cepas uruguayas seleccionadas a partir de sus variantes genéticas e investigación de su estructura poblacional.

Lasserre et al. BMC Genomics (2018) 19:2
DOI 10.1186/s12864-017-4249-6

BMC Genomics

RESEARCH ARTICLE Open Access

Whole genome sequencing of the monomorphic pathogen *Mycobacterium bovis* reveals local differentiation of cattle clinical isolates

Moira Lasserre[1†], Pablo Fresia[2†], Gonzalo Greif[1], Gregorio Iraola[2], Miguel Castro-Ramos[3], Arturo Juambeltz[3], Álvaro Nuñez[3], Hugo Naya[2], Carlos Robello[1,4*] and Luisa Berná[1*]

Abstract

Background: Bovine tuberculosis (bTB) poses serious risks to animal welfare and economy, as well as to public health as a zoonosis. Its etiological agent, *Mycobacterium bovis*, belongs to the *Mycobacterium tuberculosis* complex (MTBC), a group of genetically monomorphic organisms featured by a remarkably high overall nucleotide identity (99.9%). Indeed, this characteristic is of major concern for correct typing and determination of strain-specific traits based on sequence diversity. Due to its historical economic dependence on cattle production, Uruguay is deeply affected by the prevailing incidence of *Mycobacterium bovis*. With the world's highest number of cattle per human, and its intensive cattle production, Uruguay represents a particularly suited setting to evaluate genomic variability among isolates, and the diversity traits associated to this pathogen.

Results: We compared 186 genomes from MTBC strains isolated worldwide, and found a highly structured population in *M. bovis*. The analysis of 23 new *M. bovis* genomes, belonging to strains isolated in Uruguay evidenced three groups present in the country. Despite presenting an expected highly conserved genomic structure and sequence, these strains segregate into a clustered manner within the worldwide phylogeny. Analysis of the non-pe/ppe differential areas against a reference genome defined four main sources of variability, namely: regions of difference (RD), variable genes, duplications and novel genes. RDs and variant analysis segregated the strains into clusters that are concordant with their spoligotype identities. Due to its high homoplasy rate, spoligotyping failed to reflect the true genomic diversity among worldwide representative strains, however, it remains a good indicator for closely related populations.

Conclusions: This study introduces a comprehensive population structure analysis of worldwide *M. bovis isolates*. The incorporation and analysis of 23 novel Uruguayan *M. bovis* genomes, sheds light onto the genomic diversity of this pathogen, evidencing the existence of greater genetic variability among strains than previously contemplated.

Keywords: Bovine tuberculosis, Comparative genomics, Phylogenetics, Genetically monomorphic bacteria, European 1

* Correspondence: robello@pasteur.edu.uy; lberna@pasteur.edu.uy
†Equal contributors
[1]Unidad de Biología Molecular, Institut Pasteur de Montevideo, Montevideo, Uruguay
Full list of author information is available at the end of the article

Background

Bovine tuberculosis (bTB) is a chronic respiratory disease of livestock characterized by the development of granulomas in affected tissues, caused by *Mycobacterium bovis*. bTB has serious animal welfare and economic consequences, affecting animal performance and the trade value of sub-products [1]. Despite milk pasteurization and cattle sanitary programs that initially succeeded to control bTB in many developed countries, wildlife reservoirs contribute to disease spillover back to domesticated animals, hampering control efforts [1]. *M. bovis* is also a zoonotic pathogen, being of major concern in the developing world, where human populations are at greater risk for transmission due to close animal-human interactions, and high HIV prevalence puts a greater number of immunocompromised individuals at risk [1, 2]. The absence of active surveillance programs and limited epidemiological studies in many countries have underestimated the local prevalence and its impact on humans locally, as well as worldwide [1, 3].

M. bovis belongs to the *Mycobacterium tuberculosis* complex (MTBC), a group of closely related species that share 99.9% of their nucleotide sequences, and have identical 16S rRNA genes [4, 5]. While originally considered to be a genetically monomorphic group, current evidence points to the existence of considerable genomic diversity among strains [6–8]. It has been shown that the population structure of *M. bovis* is the result of a series of clonal expansions where clones acquired high frequencies within the population [9]. To date, four clonal complexes of *M. bovis* have been defined based on distinct spoligotype signatures and deletions; European 1, European 2, African 1, African 2. These distinct spoligotypes derived from the ancestral BCG-like spoligotype SB0120 [10–15]. Nonetheless, standardized epidemiological methods for strain typing within the MTBC, such as spoligotyping, exhibits a high propensity for homoplasy [16]. On the other hand, single nucleotide polymorphisms (SNPs) are well distributed throughout the genome in both intragenic and intergenic regions, and have low reverse mutation rates and homoplasy indexes [16]. Therefore, SNPs arise as a reliable and robust tool for establishing phylogenetic relationships and for population structure studies of the MTBC [17].

Uruguay has the world's highest number of cattle per capita (3.6), with over 12 million bovines. Although bTB prevalence has been low for the past 50 years due to the implementation of a national surveillance program [18], several outbreaks were reported between 2011 and 2013. Importantly, the country's economy is largely dependent on the cattle industry [19]. Costly control programs and significant production decreases caused by *M. bovis*, mainly affecting the dairy cattle industry, appreciably threaten the country's economy.

In the present study, we set out to characterize the genomic variability among *M. bovis* isolates present in the country. We sequenced and analyzed the genomic relationship of 23 *M. bovis* Uruguayan isolates, obtained from representative dairy farms in Uruguay, to worldwide strains selected to represent the highest clonal complex diversity available. Subsequent comparative genomics allowed us to explore their genomic variability and to determine local diversity. To assess the relevance of spoligotyping as a complimentary source of information on the variability of strains, we analyzed the isolates´ spoligotype patterns in silico. We found that, despite presenting the expected highly conserved genomic structure, these strains displayed key variability traits that contributed to the formation of a distinctly structured population.

Results

Whole genome sequencing, assembly and genotyping of 23 Uruguayan strains

The calculated coverage, N50 and genome sizes for the 23 sequenced local strains range from 21X to 161X, 44,571 to 107,175 bp and 3.49 to 4.50 Mbp, respectively. Detailed information on the sequencing statistics can be found in the Additional File 1: Figure S1. Epidemiological data for the selected Uruguayan strains of *M. bovis* is shown in Additional file 2: Table S1. All 23 strains were found to have 1 rRNA operon. Strain MbURU-002 exhibits the highest number of predicted CDS, which correlates to it having the largest genome size among the strains. The total numbers of tRNA, ranges from 50 to 58 genes. Strain-specific detailed information can be found in in Additional file 3: Figure S2. In silico spoligotyping of the sequenced Uruguayan strains showed five different patterns: SB0274 (35%), SB0145 (30%), SB0130 (26%), SB0140 (4%), and SB1072 (4%).

Phylogenomics of *M. bovis* strains portray structured populations

To comprehend the largest possible genetic diversity in the analysis, we included genomes from isolates with geographically distinct origins, representing three clonal complexes. We uncovered that an average of ~58% of genes are shared for any individual genome; 2370 genes are shared between the Uruguayan strains sequenced and 163 MTBC strains isolated worldwide (see Additional File 4: Table S2 for details). The phylogenetic relationships reconstructed by maximum likelihood based on the core genome (i.e. the 2370 shared genes) show three main clusters of *M. bovis*, each corresponding to the clonal complexes European 1 (Eu1), European 2 (Eu2) and BCG-like (Fig. 1a). All the Uruguayan strains cluster within the widely distributed and highly structured clonal

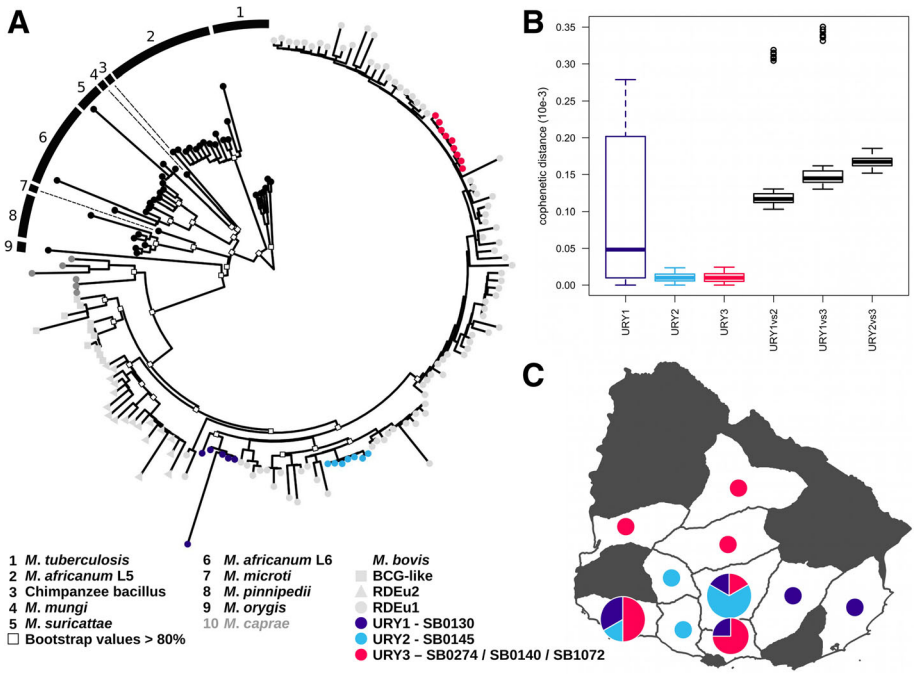

Fig. 1 (**a**) Maximum likelihood phylogeny of *Mycobacterium tuberculosis* complex obtained based on 2370 core genes, (**b**) Cophenetic distance, showing the diversity, within the three Uruguayan groups (URY1, URY2, URY3) and between them, (**c**) Distribution by group of the 23 *M. bovis* isolates in Uruguay. While *M. bovis* was isolated from all but four of the departments of this country (data not shown), white areas specify those where the 23 sequenced strains where isolated from. The size of the circles is representative of the number of strains isolated from each department

complex Eu1. Within Eu1, the Uruguayan strains form three divergent groups (URY1, URY2, and URY3). Fig. 1c shows the geographical distribution of *M. bovis* groups URY1, URY2 and URY3 within Uruguay. URY1 shows the highest within group genetic distance (Fig. 1b), and is a well-supported cluster (bootstrap >80). URY2 shows a low within group genetic distance (Fig. 1b), and is clustered with strains from Mexico and the USA. URY3 also shows a low within group genetic distance (Fig. 1b), but groups with the most diverse and widely distributed cluster which includes strains from Argentina, Brazil, Canada, Mexico, South Korea, UK and USA (Additional File 4: Table S2). Using the core genome, we estimated π and θW for each of the groups, which measure genetic diversity both for synonymous and non-synonymous sites (Table 1). As expected, the overall variability is low. On average, URY3 is the most variable group for synonymous sites, but variability estimates based on non-synonymous sites are similar for the three groups. Variant based phylogenetic tree and Principal Component Analysis (PCA) cluster the Uruguayan strains according to their group identities (Fig. 2). Interestingly, we found there is a correlation between variants and spoligotypes. However, this is only clearly evident at the country scale, whereby MbURU-022 (SB0140) shares more variants with the rest of URY3 than with AF2122/97 (SB0140).

Variability throughout *M. bovis* genomes

We delved into the genomic differences among strains which could further explain the observed population structure among the Uruguayan strains. Figure 3 summarizes the genomic variability of the sequenced strains with respect to the reference genome of *M. bovis* [20]. Strains MbURU-012, 014 and 020 exhibit sizable areas of missing information, which can be attributed to the low sequencing depth of their genomes, resulting in smaller genome sizes (3.48 to 4.09 Mbp). To prevent skewing in the data analysis, we defined a minimum genomic coverage of 30X from 100 bp paired end data, for variant analysis, annotation, and other downstream analyses. The scarcity of low identity regions among the

Table 1 The mean number of segregating sites and estimates of synonymous and non-synonymous genetic diversity for each group. SS: Segregating sites across the core genes. π: The average pairwise differences per site. θW: Watterson's estimator of genetic diversity based on the number of segregating sites

	Synonymous			Non-Synonymous		
	SS	π	θW	SS	π	θW
URY1	312	0.000006	0.00005	264	0.00004	0.00004
URY2	162	0.0000004	0.00002	139	0.00002	0.00001
URY3	523	0.00001	0.00005	370	0.00003	0.00004

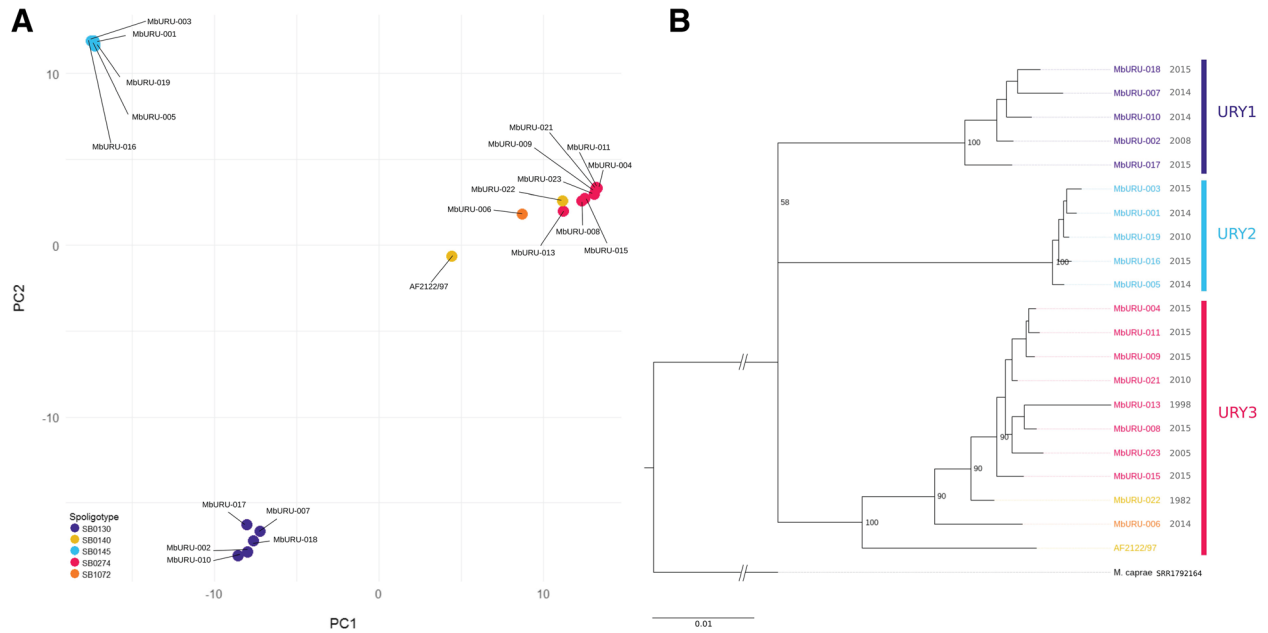

Fig. 2 Clustering of the Uruguayan strains, evidenced by (**a**) PCA of their variants against the reference strain AF2122/97, and (**b**) phylogenetic tree of these variants obtained with RAxML. Year of isolation is showed next to each of the strain names. *Mycobacterium caprae* (SRR1792164) was chosen as outgroup

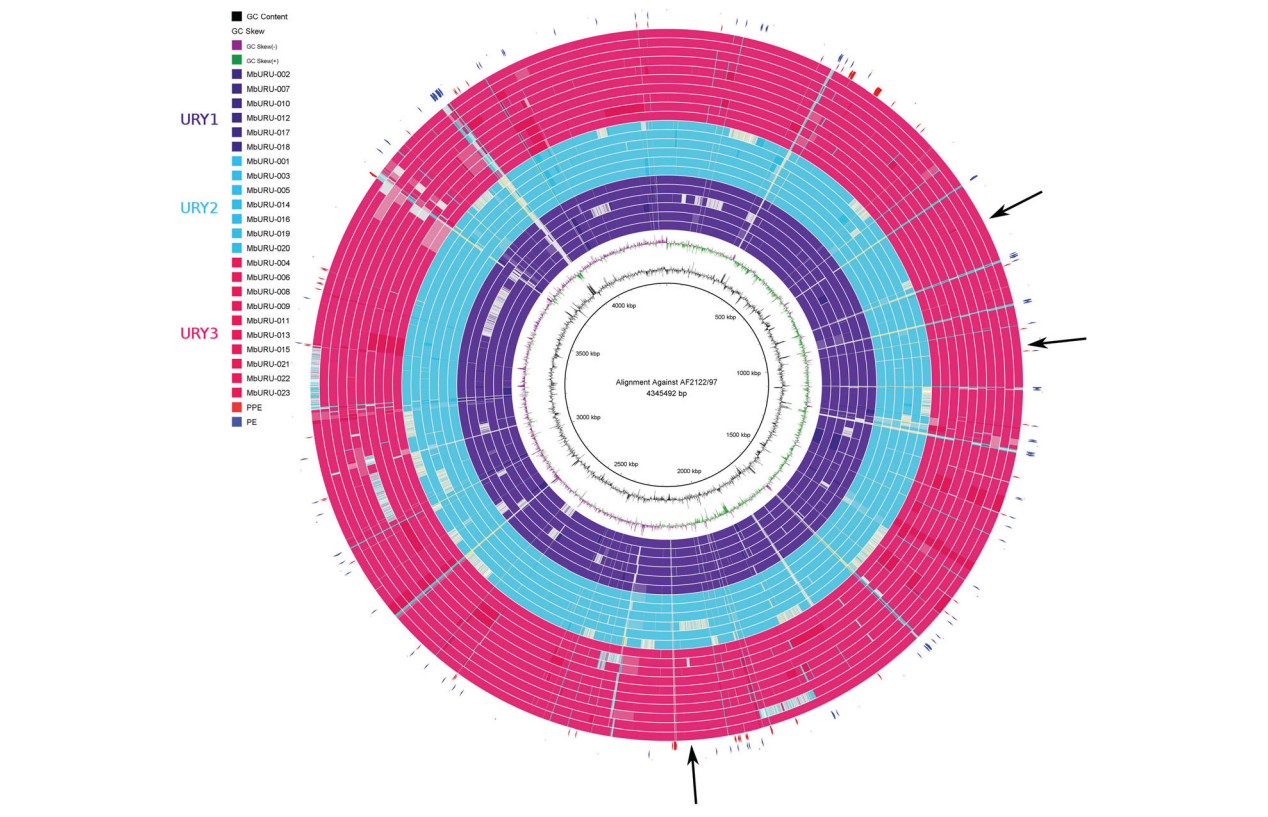

Fig. 3 Pairwise alignments of each of the 23 sequenced Uruguayan strains against the reference strain. White regions denote those areas that have identity scores <95% or that are absent in the strain. Blue and red arrowheads indicate the location of gene families PE and PPE, respectively, while black arrows exemplify low identity or absent regions without PE and PPE genes

local strains is consistent with the proposed 99.9% overall genetic identity among members of the MTBC [20, 21]. Low identity regions identified mainly correspond to genes coding for PE and PPE protein families (blue and red arrowheads in Fig. 3, respectively). Nonetheless, we identified novel areas exhibiting low identity, or even missing, with respect to the reference genome. First, we localized low identity regions, which harbored 25 genes (Additional File 5: Table S3). Among them we found *pks12* and two glycosyltransferases (*Mb1551* and *Mb1553c*) with identity scores between 86% and 93% in nine of the strains. Manual inspection of the alignment shows that several unequivocal mapped reads support these polymorphisms, evidencing that the observed diversity is real. On the contrary, identity scores found in the glycosyltransferases are the result of gene duplications in these strains or the corresponding deletions in the reference strain, such that the observed variability is due to mismapping. (Additional File 6: Figs. S3A-S3B). This is evidenced by the high read coverage of the glycosyltransferases, which reaches two to ten times the genome-wide average in all strains except for MbURU-001, 008 and 021.

Regions of difference (RDs) were identified by inspecting the genome-wide coverage density. We found ten novel RDs are present in the Uruguayan strains when compared against the reference (Table 2). To confirm the validity of these results, we tested the four found to be specific to a given spoligotype (bolded in Table 2). Two of these regions were confirmed by PCR (RDbov130a and RDbov145b, Additional File 7: Figure S4A), and all of them were manually curated by visual inspection of the paired-end read alignments both against the reference genome and against their respective assemblies (Additional File 7: Figure S4B). All strains of spoligotypes SB0130 and SB0145 lack RD3, previously reported as absent in other MTBC strains including *M. bovis*, BCG and *M. tuberculosis* [22]. The remaining nine RDs have not been described previously. Interestingly, many of these regions of difference appear to be spoligotype-specific. For instance, strains with spoligotype SB0130 lack a 994 bp region that comprises the gene *Mb3923c*. Likewise, strains with spoligotype SB0145, share the loss of three different regions, of 604, 546 and 1448 bp, which correspond to the group of genes *Mb0026-Mb0027, xylB* and *Mb0930-Mb0931*, respectively. Other lost regions code for PE and PPE proteins, exhibiting also a spoligotype-specific pattern of loss (Table 2).

Duplicated genes in the Uruguayan strains were identified by performing a coverage analysis. A total of 217

Table 2 Details of the regions of difference (RD) of lengths greater than 500 bp in the sequenced strains from Uruguay, as well as deletions found in PE/PPE-coding genes. Each RD is defined by an ID, and all deletions show the number of ORFs that cover, starting and ending coordinates, length, strains presenting these deletions and spoligotypes associated with them. Marked in bold are the RDs with more robust association to a given spoligotype

ID		Genes covered	Start	End	Length	Strains (MbURU-)	Spoligotypes associated
Previously described	RD3	14, Mb 1598-Mb1611c	1,764,652	1,773,872	9221	001, 002, 003, 005, 007, 0010, 012, 014, 016	SB0145, SB0130
	RDbov145a	2, Mb0026-Mb0027	29,475	30,078	604	001, 003, 005, 014, 016, 019, 020	SBO145
	RDbovl 31	5, Mb0677c-Mb0681c	756,105	757,654	1550	010	X
	RDbov145b	1, serA2	824,054	824,599	546	001, 003, 005, 014, 016, 019, 020	SB0145
	RDbov145c	2, Mb0930-Mb0931	1,009,480	1,010,927	1448	001, 003, 005, 014, 016, 019, 020	SB0145
	RDbovl 072	7, Mb1908-Mb1914c	2,116,213	2,122,816	6604	006	SB1072
	RDbovl 32	1, rmlB2	3,832,661	3,833,913	1253	007	X
	RDbov133	2, bisC-Mb1478c	1,616,260	1,619,455	3196	012	X
	RDbov134	1, Mb3756	4,115,825	4,116,744	920	012	X
	RDbov130a	1, Mb3923c	4,310,701	4,311,694	994	002, 007, 010, 012, 017, 018	SB0130
PE/PPE		1, PE_PGRS19	1,189,852	1,190,034	183	001, 003, 005, 014, 016, 019, 020	SB0145
		1, PE_PGRS20	1,192,321	1,192,742	422	001, 003, 005, 014, 016, 019, 020	SB0145
		1, PE_PGRS24	1,486,342	1,486,528	187	001, 003, 005, 014, 016, 019, 020	SB0145
		1, PPE30	2,033,505	2,033,624	120	002, 007, 010, 012, 017, 018	SB0130
		1, PE_PGRS42d	2,762,816	2,762,980	165	004, 008, 009, 011, 013, 015, 021, 022, 023	SB0140, SB0274
		1, PE_PGRS50b	3,691,479	3,691,697	219	004, 006, 008, 009, 011, 013, 015, 021, 022, 023	SB0140, SB0274, SB1072
		1, PE_PGRS50b	3,694,231	3,694,311	81	006	SB1072
		1, PE_PGRS51	3,733,072	3,733,247	176	001, 003, 005, 014, 016, 019, 020	SB1045

duplicated unique genes were identified in all strains (Additional File 8: Table S4). Three of the strains show no duplications (002, 006 and 021), and there are no duplicated genes specifically associated to a given spoligotype. Strain MbURU-018 accounts for 35% (144) of the total set of duplicated genes. In fact, Gene Ontology (GO) analysis of the duplicated genes revealed that only three strains, MbURU-007, 009 and 018, exhibit significant enrichment of duplicated genes. MbURU-007 displays enrichment in the Diterpenoid biosynthesis route. On the other hand, the duplications found in MbURU-009 and 018 enrich for immune system process, defense responses to viruses and defense responses.

We also identified five putative novel genes in the local strains that are not present in the reference genome. One PE_PGRS and one uncharacterized protein are strain specific, present only in MbURU-011 and 015, respectively. The other correspond to novel genes found in more than one strain; an ABC transporter that has been very recently re-cataloged as present in the reference genome after re-sequencing and annotation of AF2122/97 [23], a Ser/Thr protein kinase in 11 of our strains, and a hypothetical protein in strains MbURU-007 and 009 (Additional File 9: Table S5).

To further assess the variability of our strains, we evaluated the presence of polymorphisms. The Uruguayan strains harbor 231 to 551 SNPs, and 18 to 55 insertions/deletions (indels) with respect to the reference genome AF2122/97. Comparative analyses revealed a total of 1366 non-repetitive variants, 499 unique variants and a set of 43 variants common to the 20 strains. As the phylogenetic tree shows in Fig. 2, isolates in group URY3 show less variants than those in groups URY1 and URY2, suggesting these are genetically more akin to the reference strain. URY1, URY2 and URY3 show a set of 161, 296 and 47 unique variants, and 391, 529 and 182 common variants per group, respectively (Table 3).

To evaluate the potential existence of SNP clustering in the genome, we calculated the SNP density throughout the genomes in all the Uruguayan strains by establishing a sliding window of 5 kb. The resulting SNP density graph shows a non-random SNP distribution through the genome (binomial test, $p < 0.001$ after Bonferroni correction) with 78 statistically significant regions (Fig. 4, red peaks). Detailed information on these regions (density > 0.006 SNP/kpb) can be found in Additional File 10: Table S6. The genes with low identity values mentioned before, *pks12*, *Mb1551* and *Mb1553c*, had the highest detected SNP density.

We then determined the variant frequency (SNP/kpb) among genes in the 20 high coverage strains, taking into consideration their identities as URY1, URY2 or URY3. Figure 5 shows variants separated in three categories according to their impact on genes (high, moderate and low, as described in [24]). We found 376, 569 and 95 unique genes with low, moderate and high impact variants, respectively. We noted a significant difference in the incidence of variants depending on the group observed.

In order to determine whether genes affected to the same extent by SNPs are involved in common processes we analyzed their GO enrichment for each category (low, moderate and high). While genes with low impact mutations are not enriched in any GO term, we observed some interesting trends for genes moderately and highly impacted.

Genes bearing high impact mutations are not enriched in any particular biological process pathway. While some of these genes are found to have few to no paralogs, others perform functions that are interesting to note. For instance, *Mb0119*, one of the only two annotated sugar kinases in the reference genome of *M. bovis*, exhibits frameshifts in 11 of the studied strains that belong to spoligotypes SB0130 and SB0145. Heat shock protein *hsp* also shows a frameshift in seven of the eight strains of spoligotype SB0274, and the strain belonging to SB0140. Both *iniB* and *iniC*, two of the three genes of the *iniBAC* operon, bear frameshifts in strains. *iniB* appears mutated in strains belonging to spoligotypes

Table 3 Details of the variants found in each of the local strain, separated by spoligotype identity

Group	Spoligotype	Strain (MbURU-)	Total per strain	Unique per strain	Common per group	Unique per group
URY1	SB0130	002	539	8	391	161
		007	456	6		
		010	518	16		
		017	563	50		
		018	511	17		
URY2	SB0145	001	586	10	529	296
		003	592	6		
		005	571	8		
		016	592	16		
		019	578	4		
URY3	SB0274	004	387	9	182	47
		008	316	3		
		009	415	14		
		011	396	11		
		013	249	8		
		015	333	10		
		021	361	5		
		023	329	9		
	SB1072	006	361	85		
	SB0140	022	347	25		

SB0274 and SB0130, while frameshifts in *iniC* are only observed for strains of spoligotype SB1072. While we found no frameshifts in the remaining member of the *iniBAC* operon, *iniA*, interestingly, we also found one missense and one synonymous variant for all SB0145 strains. *sppA*, a putative protease IV, possibly involved in the digestion of signal peptides for the secretion of mature proteins across the membrane, was found to have acquired a frameshift in strain 009. Synonymous variants at this gene were also found in strains of the SB0145 and SB1072 spoligotypes. All truncated genes found are shared within a spoligotype, or by population groups. Of particular interest is *plcD*, the only phospholipase C found in *M. bovis*, which was found to be truncated in all strains belonging to spoligotype SB0274. The complete list of truncated genes for these strains can be found in Additional File 11: Table S7.

Genes with moderate impact mutations are enriched in peptidoglycan-based cell wall biogenesis and carbohydrate metabolic processes. All studied strains contribute equally to the enrichment of these terms, whereby strains from URY2 bear the highest number of genes associated to these terms (Additional File 12: Figure S5).

PE and PPE families of proteins represent a major source of antigenic variability in MTBC members [25–27], Notably, even though PE and PPE families are known for their high variability among strains due to their strong association to pathogenicity [28], their coding genes do not display a high frequency of variants relative to the total set of affected genes. Of the 136 known PE/PPE genes, 58 were found to show at least one variant with respect to the reference, and only 22 show a mutation rate higher than the mutation average for all affected genes (9.43 variant/kbp).

Discussion

Assessing the existing diversity of *M. bovis* strains and its correlation to pathogenicity and severity of disease is of paramount importance to the economic growth and public health in Uruguay, where cattle outnumber

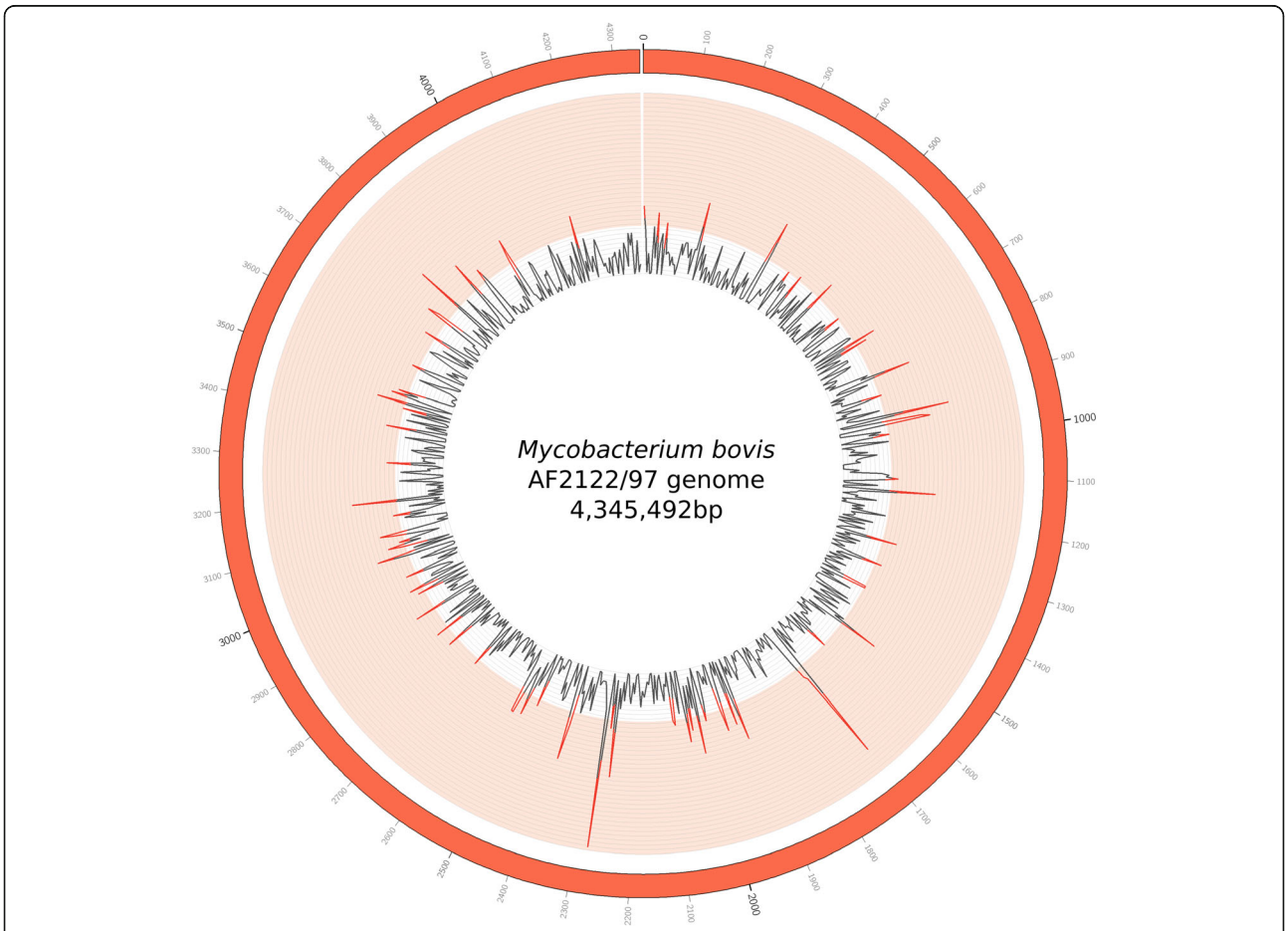

Fig. 4 Visualization of the density of variants in the entire length of the reference genome (AF2122/97). Inner rays represent the variant density of every 5 kb region: a region that surpasses a density of 0.006 (innermost circle of the light red band) is significantly high density (binomial test, $p < 0.001$ after Bonferroni correction)

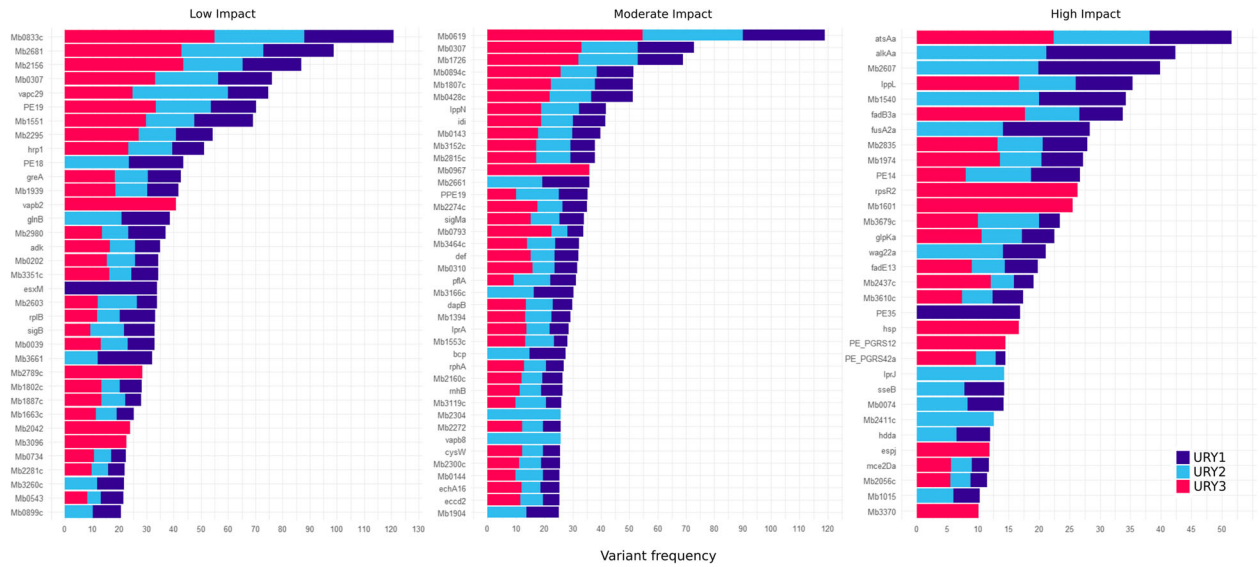

Fig. 5 Genes with the highest variant frequency (SNP/kb), categorized according to their impact in the genes as high, moderate and low impact. Each strain is colored according to the group they belong to, as defined by their phylogenetic structure: URY01, URY02 and URY03

people by almost four to one, and the cattle industry is a major contributor to the country's GDP. In this study, we report the genome structure of 23 novel Uruguayan *M. bovis* isolates, and compare them to the reference strains.

Currently, four main clonal complexes of *M. bovis* have been described, European 1 (Eu1) [13], European 2 (Eu2) [14], African 1 (Af1) [12], and African 2 (Af2) [11]. The 23 Uruguayan strains sequenced here show spoligotype patterns lacking spacer 11, concordant with strains from Eu1, which were originally identified throughout Britain and Ireland as well as former British colonies (South Africa, Australia, New Zealand, Canada and the United States), and Latin America with the exception of Brazil [10]. This clonal complex is rare in mainland Europe [13], a region dominated by Eu2 which is also the most frequent in Brazil [14]. Af1 and Af2 clonal complexes are restricted to countries from West-Central and East Africa, respectively [10, 29].

Although all Uruguayan strains belong to Eu1, we identified three highly structured groups (URY1, URY2, and URY3), showing high genetic diversity among them, consistent with the genomic variability reported in this work. The estimated diversity statistics (i.e. π and θW) for synonymous sites revealed an excess of rare alleles (θW > π) in URY1 and URY2, but not in URY3, indicating a deviation from neutrality due to demographic processes and/or population structure. For non-synonymous sites, π and θW values are not significantly different.

Eu1 likely reached its current global distribution with the trade of modern British cattle breeds, perhaps within the last 200 years [10]. The high genetic diversity found among Uruguayan *M. bovis* can probably be attributed to the extensive history of British cattle breeding in the country, and the lack of clear geographic barriers separating Uruguay from Brazil.

These three highly structured groups were congruently devised both by core genome and variant calling analyses. URY1 represents a genetically more heterogeneous group than URY2 and URY3. At the same time, these last two show closer genetic similarity between them. When compared to the reference, groups URY1 and URY2 present a higher number of variants than URY3, and a differential pattern of variant incidence between these two sets is evidenced. This clustering was also evidenced by the absence of region of difference RD3 in URY1 and URY2, which contains ORFs from phiRv1 prophage, present in URY3. Strains from URY3 are more closely related to the reference strain of *M. bovis*, AF2122/97 (SB0140) [20], than to the remaining strains. Spoligotype SB0140 is known to be a predecessor of spoligotypes SB0274 and SB1072. This is in accordance with our phylogeny of SNPs and indels, where the reference shares more variants with strains from group URY3 than with the other two. The complete set of distinctive polymorphisms for the three groups can be found in Additional File 13: Table S8.

Our comparative analysis revealed a set of genomic traits that begin to explain the observed strain clustering and/or are informative as sources of variability previously unknown. PE and PPE families of proteins represent a major source of antigenic variability in MTBC members [25–27], but a small amount of variable areas

that do not belong to these families represent useful sources of variability for the characterization of these strains. Among these are RDs which cover up to seven ORFs, most of which are annotated as conserved hypothetical proteins. The fact that some RDs are found to be absent in all strains of a given spoligotype leads us to propose that spoligotypes, which focus on a very small region of the genome of MTBC [30], can be complemented by analyzing the absence of specific RDs to localize lineages and sub lineages. While spoligotyping by itself is ineffective for phylogenetic applications given its high homoplasy rate [16], it is a good epidemiological method when used along with other techniques such as MIRU-VNTR [31]. The idea of complementing these markers with large sequence polymorphisms adjusts itself harmoniously with the suggestion that particular "signature" spoligotyping patterns can be indeed informative for population genetic analyses, where many strains can be grouped using such "signature" patterns [16]. RD typing has had a widespread use in resolving phylogenetic relationships, and it was later succeeded by SNP typing. It has also been used as a companion approach to SNP calling for predictive purposes, in search for the development of new SNP-typing barcodes [32]. Identifying genomic features like RDs that are specific to sub lineages represents a useful source of information to delineate the fine differences among strains of identical spoligotype patterns, which our data suggests do not share a common ancestor. This, taken together with our results, supports that RDs are good markers for sub lineages and should therefore not be left aside. IDs shown in bold in Table 2 indicate RDs suggested for use as markers to characterize the spoligotypes found in our samples: RD145a, RD145b and RD145c and RD130a.

SNPs are also a contributing factor to the variability of some regions, including several of the low identity genes found in these strains: the glycosyltransferases *Mb1551* and *Mb1553c,* and pks12, the longest gene in bTB and tuberculosis genomes coding for a polyketide synthase implicated in antigenic variation [33]. Of these three, lower identities in the glycosyltransferases can be explained by their exceptionally high read coverage in all but three of our strains. This hints at duplicated genes, that the assembler collapsed, which were consequently perceived as bearing low identity to the reference in strains MbURU-002, 003, 007, 009, 011, 016, 022 and 023. A high degree of diversity was found among the identified duplicated genes. Particularly, strains 009 and 018 have an exceptionally high number of duplications relative to the other strains. Large numbers of duplicated genes could be associated with a genomic adaptation to a changing environment by means of a gene dosage effect. It has been proposed before that while gene duplication might not necessarily double gene dosage, due to potential negative feedback loops, it generally leads to its increase [34–36]. Biological assays will be required to determine whether a functional relationship exists between this genomic feature and the enriched biological processes of defense response to virus and immune system.

An average of 131 (36%) synonymous and 229 (64%) non-synonymous SNPs were found among our strains. The genomes present a high dN/dS ratio, which is consistent with previous assertions of low levels of purifying selection in MTBC members [37]. We found a bias in moderate impact variants towards genes involved in cell wall organization pathways, as it is the case for *iniBAC* promoter genes, which are induced by inhibitors of cell wall biosynthesis. At the same time, this enrichment is not apparent when the strains are studied individually. There seems to be a lack of a consensus strategy for cell wall organization, which implies different means to the same end. Heterogeneity of cell surfaces is an adaptive trait among populations of pathogenic bacteria, which allows different sub populations to thrive in a variety of environments within the host. This would allow the organisms to adhere to a greater variety of surfaces, putting cell wall component genes under diversifying selection [38, 39]. On the other hand, high impact variants in our strains affected many genes with unique or noteworthy functions, however, no bias in biological pathways were identified. Further analysis of the functional consequences of these mutations and examination of these in larger sample numbers is required. The number of variants in PE and PPE genes drew a different picture for this family. Unexpectedly, most members of this family do not present a higher frequency of variants than the average. On one hand, many of these genes are known to exhibit extensive genetic variation, with many mechanisms having been reported that contribute to this [28]. On the other hand, the main mechanism of variation within these families could be events of homologous recombination that result in deletions or duplications of whole regions in the genes, instead of a high frequency of variants. This is highly likely, given that most members in these families contain domains comprised of series of tandem repeats, making them more prone to undergo recombination [40].

The analysis presented here sheds light onto *M. bovis* genomic variability worldwide, as well as, the nature of Uruguayan strains. Moreover, it supports the existence of a larger pattern of genomic variability than that provided by a spoligotype classification. In small populations, spoligotypes are good indicators of the genomic similarity among strains, as it is evidenced by the spoligotype/SNP correlation observed in the Uruguayan strains. However, for the purposes of comparison among multiple populations, in a worldwide context, spoligotypes

fall short. High homoplasy rates blur the initial correlation between these two markers evidencing the clear need to resort to additional typing means. RDs have shown to be genomic markers with good resolution in defining sublineages and we therefore proposed them as appropriate markers. The incorporation of meta data will also be crucial for a more thorough characterization of the Uruguayan population of *M. bovis*, allowing a definitive association between all genomic traits described here and their potential phenotypic effects.

Conclusions

We have sequenced and analyzed 23 genomes of *M. bovis* in Uruguay. Comparative studies of these strains showed that, while they all belong to the clonal complex European 1, they exhibit a surprisingly structured phylogenetic tree, and a high level of genomic variability. Localization of the genomic sources of variability for Uruguayan strains was attained. We observed duplicated genes with high diversity among strains, differential distribution of regions of difference, distinct SNP patterns of incidence which specifically grouped with spoligotype patterns, novel genes absent in the reference, and variable genes. In future studies, it will be interesting to assess the association of this high variability of such genomic features with phenotypical traits in order to gain insights into the functional consequences of this diversification.

Methods

M. bovis Strains

All samples were granted by Dirección General de Servicios Ganaderos (DGSG), obtained during routine surveillance and eradication campaigns conducted in slaughter plants.

Phylogenetic reconstruction of 186 MTBC strains

The 164 MTBC strains used to compare the *M. bovis* strains from Uruguay were downloaded from public databases, assembled with SPAdes [41], and annotated with Prokka [42]. The pan genome of all the 186 strains was estimated with Roary (v3.6.8) [43]; core genes were defined as those present in all 186 isolates with a 90% ID cut-off. Recombination was inspected using fastGEAR [44]. Finally, core genes aligned with Mafft [45] were used to reconstruct the phylogenetic relationships by maximum likelihood using RAxML [46], with GTRCAT model and autoMRE for bootstraping. For each Uruguayan group (i.e. URY1, URY2, URY3), two measures of genetic diversity were estimated, the average pairwise nucleotide diversity per site (π, [47]) and Watterson's θ (θW), which is based on the number of segregating sites [48]. For both measurements, synonymous and non-synonymous diversity was calculated separately using the R package PopGenome [49].

Strain selection, culture and DNA extraction

23 *M. bovis* strains were used for whole genome sequencing, obtained from ten departments of Uruguay in the years 1982, 1998, 2005, 2008, 2010, 2014 and 2015. Samples were decontaminated using the 5% oxalic acid decontamination method [50], using equal parts of both decontaminant and sample for 15 min at 37 °C. Tissues were homogenized (Stomacher 400) and later centrifuged at 2800 rpm for 30 min. The resulting sediments were inoculated on Löwenstein-Jensen and Stonebrink media, incubated at 37 °C and screened weekly for macroscopic growth until eight weeks. Identification of mycobacteria was based on observation of smears submitted to the Ziehl-Neelsen method, growth characteristics such as time, temperature and colony morphology. Suspected colonies were then evaluated with biochemical tests [51]. Genomic DNA was purified from supernatants of the strains diluted in TE (Tris-EDTA) and warmed at 100 °C for 10 min. Experimental corroboration of the extracted DNA as belonging to *Mycobacterium tuberculosis* complex (MTBC) was performed by PCR of the ETR-D fragment as previously described [52].

Whole genome sequencing and typification

Sequencing of the 23 strains was performed at the Institut Pasteur de Montevideo on an Illumina MiSeq platform from a paired-end library (2 × 75 cycles). Briefly, Nextera XT (Illumina, USA) library preparation kit was used from 1 ng of total DNA according to manufacturer instructions. Index primers were added to each library to allow sequence multiplexing. After 12 PCR cycles, the final library was purified with AMPure XP (Benchman, USA) and quantified with the Qubit dsDNA HS assay kit (Invitrogen, USA). Quality and length of the librarywere assessed with the Agilent high-sensitivity DNA kit (Agilent, USA) using the 2100 Bioanalyzer (Agilent, USA). Quality assessment of the resulting reads was performed using NGSQCToolkit (v2.3.3) [53]. Those reads with overall quality score below 20 were filtered out. From the remaining reads, we calculated the resulting coverage of each genome. If the coverage was lower than 30X, we rejected the whole sequencing project. This value was chosen to ensure a good quality of the variants called in downstream analysis. Typification of all strains was performed in silico with the tool SpolPred [54]. SpolPred predicts the spoligotype pattern of a strain based on the reads of a whole genome sequencing project.

Genome assembly

Velvet (v1.2.10) [55] was used to perform a de novo assembly of the local strains. Multiple independent assemblies were performed for each strain, differing in the chosen k-mer value, which ranged from 17 to 61. We chose the best k-mer and respective assembly based on the amount of resulting contigs, the N50 and the relation between the overall length of the assembly and the sum of both the lengths of the contigs under and over than 1000 bp. We expect a lower amount of contigs with a high N50 and an overall assembly length close to the known length of the reference strain AF2122/97. We also performed two iterations of confirmatory assembly using SPAdes (v3.6.1) [41], using the previous Velvet assembly as known data and changing the set of trial k-mers in each. All three resulting assemblies were integrated with the software CISA (v1.3) [56].

Genome improvement

PAGIT toolkit [57] was used to further improve the quality of the assembly by closing gaps on scaffolds, correcting base errors and generate an annotation of CDS based on the reference genome AF2122/97. The final assembly was also automatically annotated with RAST server [58] and Prokka [42]. tRNA and rRNA genes were identified with ARAGORN [59] and barnap (v0.6), respectively.

Genome alignment against the reference

Genome comparison of the Uruguayan strains relative to the reference was done using BLAST Ring Image Generator (BRIG) software [60], which performed pairwise alignments for the 23 genomes. Chosen identity values to be displayed were 98% and 95% as upper and lower identity threshold, respectively.

Mapping to the reference and coverage analyses

Paired-end reads were first aligned to the GRCh38 human genome assembly (GenBank accession GCF_000001405.28) with BWA (v0.7.12-r1039) [61], in order to filter out possible human sequenced reads resulting from manipulation errors in the sequencing stage. From the remaining reads, we then performed a second mapping to the reference strain *M. bovis* AF2122/97 (GenBank accession NC_002945), allowing up to 3 mismatches in each seed of length 15 bp. Samtools mpileup [62] was ran on the mapped reads to get the coverage at each base. This was used to identify RDs and duplicated genes. For the former we extracted those regions with coverage 0 and a length higher than 500 bp. For the latter, we qualified as duplicated genes those whose median coverage was higher than 2 in more than 70% of the length of each gene.

Validation of regions of difference

To verify the predicted RDs we used two strategies. Firstly, by PCR. The PCR reactions targeted at a region inside each RD consisted of Mango-Taq (Bioline, London, UK), primers listed in Additional File 7: Figure S4A (Integrated DNA Technologies), and the following cycling parameters: 4 min at 95 °C, 37 cycles at 94 °C for 30 s, at 55 °C for 20 s and 72 °C for 30 s, with 4 min at 72 °C for the final extension. We analyzed the reaction results on a 1% agarose gel and determined whether the primers amplified or not (showing absence or presence of the RD, respectively), or if the results were not conclusive (True, False and Inconclusive in Additional File 7: Figure S4A). Secondly, we further investigated the regions by mapping and performing manual inspection of the sequenced reads mapped to the reference and the reads mapped to the assembled genome though the Integrative Genomics Viewer (IGV, [63]). If the RD was an artifact, the number of mapped reads would slowly decrease before reaching zero. Furthermore, we analyzed the insert sizes of those reads flanking the absent regions, where we expected red-colored reads (with insert sizes larger than expected) flanking the RDs in an alignment against the reference genome. Conversely, mapping against the assembled genomes should not is not expected to show blue-colored reads (insert sized smaller than expected). Finally, we assessed the support of reads in the position where the deletion would occur.

Novel gene prediction

From the mapping to the reference performed with BWA, we kept the unaligned reads of the 23 Uruguayan strains and assembled each subset de novo with Velvet. RAST was used to find ORFs from the resulting contigs and annotate them. We filtered out those sequences that were smaller than 225 nucleotides and performed a blastp of all remaining sequences against each other to find common novel genes between strains and against the NCBI database to locate the highest identity matches against them.

GO annotation

The Gene Ontology (GO) terms for any set of genes were analyzed as follows. The orthologous genes for each gene in *M. tuberculosis* H37Rv were obtained by reciprocal blastp. Given the high identity between these two members of MTBC, there was no need for obtaining orthologous genes by alternative methods. For this set of genes, we acquired the prioritized biological process GO terms ($p < 0.05$) and their fold enrichment from the Gene Ontology project [64].

Variant calling and clustering

The reads and pairs that mapped to the reference were filtered in with Samtools to later perform variant calling

(v0.1.18) [62], not including the three low coverage strains (< 30X). Samtools mpileup and VarScan (v2.3.7) [65] were used for variant calling, filtering both indels and SNPs with a minimum of 20 supporting reads at a position to call variants (–-min-reads2) and a minimum variant allele frequency threshold of 0.2 (–-min-var.-freq). In order to cluster the local strains and visualize their relatedness we performed a principal component analysis (PCA) from the variants of the 23 Uruguayan strains using the package *adegenet* [66]. A maximum likelihood phylogeny was estimated from these variants using RAxML (v8.2.7) with GTRGAMMA model and 100 bootstrapping iterations [46], choosing *M. caprae* as outgroup (SRA Accession: SRR1792164). From the SNPs identified, we also calculated the SNP density along the length of the reference strain AF2122/97 using a sliding window of 5 kb (SNP absolute frequency divided by the length of the window). To visualize if there were heterogeneous SNP-dense regions in the genome, we visualized these densities with Circos (v0.69) [67]. We used SnpSift (v4.2) [68], a vcf-manipulation tool, to extract all the genes contained between the SNP-densest regions.

SNP annotation

SnpEff was used to annotate variants of the local strains (v4.2, build 2015–12-05) [24], classifying them as synonymous or non-synonymous and obtaining their respective impact according to their incidence in the resulting gene (high, medium, low. See SnpEff documentation). Based on this information, we also calculated the number of variants each gene showed for each strains divided by the gene length in kilobases as a relative measure of SNPs (SNP/kb). Finally, we performed a GO search to obtain the most affected terms.

Additional files

Additional file 1: Figure S1. Sequencing and genome annotation statistics for the 23 Uruguayan strains of *M. bovis*. (PDF 115 kb)

Additional file 2: Table S1. *M. bovis* strains isolated from a bovine host in Uruguay. (ODS 26 kb)

Additional file 3: Figure S2. List of tRNAs and corresponding codons in the Uruguayan genomes under study. (PDF 66 kb)

Additional file 4: Table S2. Details of the 163 MTBC strains downloaded from NCBI and the 23 Uruguayan strains from this study, used to resolve a phylogenetic tree out of their core genomes. (ODS 34 kb)

Additional file 5: Table S3. Resulting blastn of all non-pe/ppe coding sequences against *M. bovis* AF2122/97. (ODS 24 kb)

Additional file 6: Figure S3a. Interactive Genome Visualizer (IGV) visualization of reads aligned to reference genome AF2122/97 (exemplified by strains MbURU-002, MbURU-003 and MbURU-010), showing a region containing glycosyltransferase-coding genes Mb1551 and Mb1553c (wbbL2). Coverage is represented in the upper track of each strain as a gray bar chart, and SNPs are showed as colored bars. The average coverage for these genes can reach up to 10 times the average genome coverage. **Figure S3b.** IGV visualization of reads from strain MbURU-002 aligned to reference genome AF2122/97, showing a region containing polyketide synthase *pks12* gene. The lower alignment represents only reads with mapping qualities higher than 0 and show therefore no multi-mapping reads, while the upper alignment represents the original data from MbURU-002. Coverage is represented in the upper track of each strain as a gray bar chart, and SNPs are showed as colored bars. Note that there are very few called SNPs in this region and some are still present once multimapping reads have been removed. (PDF 2630 kb)

Additional file 7: Figure S4a. Validation of in silico RD typing with PCR on 21 of the Uruguayan *M. bovis* strains. **Figure S4b.** Alignment of the sequenced reads of strain MbURU-003 against the assembled genome of the same strain. Selected pair of reads in red exemplify one of the reads that flanks both sides of a region of difference (RDbov145a) that is absent in this strain. (PDF 1643 kb)

Additional file 8: Table S4. List of the over represented genes found in the 23 Uruguayan *M. bovis* strains. (ODS 18 kb)

Additional file 9: Table S5. Putative novel genes identified in the Uruguayan strains by assembly of the unmapped reads. (ODS 14 kb)

Additional file 10: Table S6. Regions with high SNP density. (ODS 24 kb)

Additional file 11: Table S7. List of truncated genes for the Uruguayan strains studied. (ODS 20 kb)

Additional file 12: Figure S5. Bar plots displaying the number of genes affected per strain that are associated to the GO terms carbohydrate metabolic process and peptidoglycan-based cell wall biogenesis. (PDF 1342 kb)

Additional file 13: Table S8. Details on the distinctive characteristics defining each of the three groups described in this study: URY1, URY2 and URY3. (ODS 250 kb)

Abbreviations

Af1: African 1 clonal complex; Af2: African 2 clonal complex; bTB: Bovine tuberculosis; Eu1: European 1 clonal complex; Eu2: European 2 clonal complex; GDP: Gross domestic product; GO: Gene Ontology; MIRU-VNTR: Mycobacterial interspersed repetitive unit-variable number of tandem repeat; MTBC: *Mycobacterium tuberculosis* complex; ORF: Open Reading Frame; PCA: Principal Component Analysis; RD: Region of difference; SNP: Single nucleotide polymorphism

Acknowledgements

We thank Eugenia Francia for her crucial English and style corrections of the manuscript.

Funding

This research received support from the Fondo de Promoción de Tecnología Agropecuaria (FPTA): grant N° 328 for DNA sequencing and consumables, from the Agencia Nacional de Investigación e Innovación (UY): grants POS_NAC_2015_1_109466 for Moira Lasserre's fellowship, DCI-ALA/2011/023–502 "Contrato de apoyo a las políticas de innovación y cohesión territorial" for a postdoctoral fellowship for Luisa Berná, and Fondo Caldeyro Barcia for Pablo Fressia's postdoctoral fellowship, and finally from FOCEM (MERCOSUR Structural Convergence Fund): grant COF 03/11 for reagents and consumables.

Availability of data and materials

This Whole Genome Shotgun project has been deposited at DDBJ/ENA/GenBank under the accessions LFGY00000000 and NAZJ00000000 to NBAE00000000. The versions described in this paper are versions LFGY01000000 and NAZJ01000000 to NBAE01000000 (MbURU-001 and MbURU-002 to MbURU-023, respectively).

Authors' contributions

ML, PF, AN, GI, CR and LB conceived the study. ML, PF and LB and CR developed the methodology. MCR and AJ cultured the strains and extracted the genomic DNA. GG performed the sequencing of strains and collected the data. ML and PF performed the bioinformatic analyses and HN was in charge of the statistical analyses and assisted in the data analysis. ML, PF, LB,

GI and CR wrote the manuscript. All the authors reviewed critically the manuscript and approved the final version.

Ethics approval and consent to participate
Not applicable.

Consent for publication
Not applicable.

Competing interests
The authors declare that they have no competing interests.

Publisher's Note

Author details
[1]Unidad de Biología Molecular, Institut Pasteur de Montevideo, Montevideo, Uruguay. [2]Unidad de Bioinformática, Institut Pasteur de Montevideo, Montevideo, Uruguay. [3]Departamento de Bacteriología, División de Laboratorios Veterinarios (DI.LA.VE.) "Miguel C. Rubino", Montevideo, Uruguay. [4]Departamento de Bioquímica, Facultad de Medicina, Universidad de la República, Montevideo, Uruguay.

Received: 8 June 2017 Accepted: 31 October 2017
Published online: 02 January 2018

References

1. Michel AL, Müller B, van Helden PD. Mycobacterium Bovis at the animal-human interface: a problem, or not? Vet Microbiol. 2010;140:371–81.
2. Ayele WY, Neill SD, Zinsstag J, Weiss MG, Pavlik I. Bovine tuberculosis: an old disease but a new threat to Africa. Int J Tuberc Lung Dis Off J Int Union Tuberc Lung Dis. 2004;8:924–37.
3. Bermingham ML, Bishop SC, Woolliams JA, Pong-Wong R, Allen AR, McBride SH, et al. Genome-wide association study identifies novel loci associated with resistance to bovine tuberculosis. Heredity. 2014;112:543–51.
4. Sreevatsan S, Pan X, Stockbauer KE, Connell ND, Kreiswirth BN, Whittam TS, et al. Restricted structural gene polymorphism in the mycobacterium tuberculosis complex indicates evolutionarily recent global dissemination. Proc Natl Acad Sci U S A. 1997;94:9869–74.
5. Böddinghaus B, Rogall T, Flohr T, Blöcker H, Böttger EC. Detection and identification of mycobacteria by amplification of rRNA. J Clin Microbiol. 1990;28:1751–9.
6. Otal I, Martín C, Vincent-Lévy-Frebault V, Thierry D, Gicquel B. Restriction fragment length polymorphism analysis using IS6110 as an epidemiological marker in tuberculosis. J Clin Microbiol. 1991;29:1252–4.
7. Kamerbeek J, Schouls L, Kolk A, van Agterveld M, van Soolingen D, Kuijper S, et al. Simultaneous detection and strain differentiation of mycobacterium tuberculosis for diagnosis and epidemiology. J Clin Microbiol. 1997;35:907–14.
8. Supply P, Allix C, Lesjean S, Cardoso-Oelemann M, Rüsch-Gerdes S, Willery E, et al. Proposal for standardization of optimized Mycobacterial interspersed repetitive unit-variable-number tandem repeat typing of mycobacterium tuberculosis. J Clin Microbiol. 2006;44:4498–510.
9. Smith NH, Dale J, Inwald J, Palmer S, Gordon SV, Hewinson RG, et al. The population structure of Mycobacterium Bovis in great Britain: Clonal expansion. Proc Natl Acad Sci U S A. 2003;100:15271–5.
10. Smith NH. The global distribution and phylogeography of Mycobacterium Bovis clonal complexes. Infect Genet Evol J Mol Epidemiol Evol Genet Infect Dis. 2012;12:857–65.
11. Berg S, Garcia-Pelayo MC, Müller B, Hailu E, Asiimwe B, Kremer K, et al. African 2, a Clonal complex of Mycobacterium Bovis epidemiologically important in East Africa. J Bacteriol. 2011;193:670–8.
12. Müller B, Hilty M, Berg S, Garcia-Pelayo MC, Dale J, Boschiroli ML, et al. African 1, an epidemiologically important clonal complex of Mycobacterium Bovis dominant in Mali, Nigeria, Cameroon, and Chad. J Bacteriol. 2009;191:1951–60.
13. Smith NH, Berg S, Dale J, Allen A, Rodriguez S, Romero B, et al. European 1: a globally important clonal complex of Mycobacterium Bovis. Infect. Genet. Evol. J. Mol. Epidemiol. Evol. Genet. Infect. Dis. 2011;11:1340–51.
14. Rodriguez-Campos S, Schürch AC, Dale J, Lohan AJ, Cunha MV, Botelho A, et al. European 2–a clonal complex of Mycobacterium Bovis dominant in the Iberian peninsula. Infect. Genet Evol J Mol Epidemiol Evol Genet Infect Dis. 2012;12:866–72.
15. Haddad N, Ostyn A, Karoui C, Masselot M, Thorel MF, Hughes SL, et al. Spoligotype diversity of Mycobacterium Bovis strains isolated in France from 1979 to 2000. J Clin Microbiol. 2001;39:3623–32.
16. Comas I, Homolka S, Niemann S, Gagneux S. Genotyping of genetically monomorphic bacteria: DNA sequencing in mycobacterium tuberculosis highlights the limitations of current methodologies. PLoS One. 2009;4:e7815.
17. Joshi D, Harris NB, Waters R, Thacker T, Mathema B, Krieswirth B, et al. Single nucleotide polymorphisms in the Mycobacterium Bovis genome resolve Phylogenetic relationships. J Clin Microbiol. 2012;50:3853–61.
18. Picasso C, Alvarez J, VanderWaal KL, Fernandez F, Gil A, Wells SJ, et al. Epidemiological investigation of bovine tuberculosis outbreaks in Uruguay (2011–2013). Prev Vet Med. 2017;138:156–61.
19. Buckman RT. Latin America 2013. Rowman & Littlefield; 2013.
20. Garnier T, Eiglmeier K, Camus J-C, Medina N, Mansoor H, Pryor M, et al. The complete genome sequence of Mycobacterium Bovis. Proc Natl Acad Sci U S A. 2003;100:7877–82.
21. Brosch R, Gordon SV, Marmiesse M, Brodin P, Buchrieser C, Eiglmeier K, et al. A new evolutionary scenario for the mycobacterium tuberculosis complex. Proc Natl Acad Sci U S A. 2002;99:3684–9.
22. Cole ST. Comparative and functional genomics of the mycobacterium tuberculosis complex. Microbiol Read Engl. 2002;148:2919–28.
23. Malone K, Farrell D, Stuber T, Schubert O, Aebersold R, Robbe-Austerman S, Gordon S. Updated Reference Genome Sequence and Annotation of Mycobacterium bovis AF2122/97. Genome Announcements. 2017;5:e00157-17.
24. Cingolani P, Platts A, Wang LL, Coon M, Nguyen T, Wang L, et al. A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff. Fly (Austin). 2012;6:80–92.
25. Sampson SL, Mycobacterial PE. PPE proteins at the host-pathogen Interface. J. Immunol Res. 2011;2011:e497203.
26. Karboul A, Mazza A, Gey van Pittius NC, Ho JL, Brousseau R, Mardassi H. Frequent homologous recombination events in mycobacterium tuberculosis PE/PPE multigene families: potential role in antigenic variability. J Bacteriol. 2008;190:7838–46.
27. Talarico S, Cave MD, Marrs CF, Foxman B, Zhang L, Yang Z. Variation of the mycobacterium tuberculosis PE_PGRS33 Gene among clinical isolates. J Clin Microbiol. 2005;43:4954–60.
28. Fishbein S, van Wyk N, Warren RM, Sampson SL. Phylogeny to function: PE/PPE protein evolution and impact on mycobacterium tuberculosis pathogenicity. Mol Microbiol. 2015;96:901–16.
29. Carvalho RCT, Vasconcellos SEG, Issa Mde A, PMS F, PMPC M, Araújo FR, et al. Molecular typing of Mycobacterium Bovis from cattle reared in Midwest Brazil. PLoS One. 2016;11:e0162459.
30. Jagielski T, van Ingen J, Rastogi N, Dziadek J, Mazur P, et al. Current methods in the molecular typing of mycobacterium tuberculosis and other Mycobacteria. Biomed Res Int. 2014;2014:e645802.
31. Jafarian M, Aghali-Merza M, Farnia P, Ahmadi M, Masjedi MR, Velayati AA. Synchronous comparison of mycobacterium tuberculosis epidemiology strains by "MIRU-VNTR" and "MIRU-VNTR and Spoligotyping" technique. Avicenna J. Med. Biotechnol. 2010;2:145–52.
32. Coll F, McNerney R, Guerra-Assunção JA, Glynn JR, Perdigão J, Viveiros M, et al. A robust SNP barcode for typing mycobacterium tuberculosis complex strains. Nat Commun. 2014;5:4812.
33. Matsunaga I, Bhatt A, Young DC, Cheng T-Y, Eyles SJ, Besra GS, et al. Mycobacterium tuberculosis pks12 produces a novel polyketide presented by CD1c to T cells. J Exp Med. 2004;200:1559–69.
34. Zhou J, Lemos B, Dopman EB, Hartl DL. Copy-number variation: the balance between gene dosage and expression in Drosophila Melanogaster. Genome Biol Evol. 2011;3:1014–24.
35. Schuster-Böckler B, Conrad D, Bateman A. Dosage sensitivity shapes the evolution of copy-number varied regions. PLoS One. 2010;5:e9474.
36. Stranger BE, Forrest MS, Dunning M, Ingle CE, Beazley C, Thorne N, et al. Relative impact of nucleotide and copy number variation on gene expression phenotypes. Science. 2007;315:848–53.
37. Hershberg R, Lipatov M, Small PM, Sheffer H, Niemann S, Homolka S, et al. High functional diversity in mycobacterium tuberculosis driven by genetic drift and human demography. PLoS Biol. 2008;e311:6.

38. Pepperell CS, Casto AM, Kitchen A, Granka JM, Cornejo OE, Holmes EC, et al. The role of selection in shaping diversity of natural M. Tuberculosis populations. PLoS Pathog. 2013;9:e1003543.
39. Mei HC. van der, Busscher HJ. Bacterial Cell Surface Heterogeneity: A Pathogen's Disguise. PLoS Pathog. 2012;8:e1002821.
40. Phelan JE, Coll F, Bergval I, Anthony RM, Warren R, Sampson SL, et al. Recombination in pe/ppe genes contributes to genetic variation in mycobacterium tuberculosis lineages. BMC Genomics. 2016;17:151.
41. Bankevich A, Nurk S, Antipov D, Gurevich AA, Dvorkin M, Kulikov AS, et al. SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. J Comput Biol J Comput Mol Cell Biol. 2012;19:455–77.
42. Seemann T. Prokka: rapid prokaryotic genome annotation. Bioinforma. Oxf. Engl. 2014;30:2068–9.
43. Page AJ, Cummins CA, Hunt M, Wong VK, Reuter S, Holden MTG, et al. Roary: rapid large-scale prokaryote pan genome analysis. Bioinforma Oxf Engl. 2015;31:3691–3.
44. Mostowy R, Croucher NJ, Andam CP, Corander J, Hanage WP, Marttinen P. Efficient inference of recent and ancestral recombination within bacterial populations. Mol Biol Evol. 2017;
45. Katoh K, Standley DM. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. Mol Biol Evol. 2013; 30:772–80.
46. Stamatakis A. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. Bioinformatics. 2014;30:1312–3.
47. Nei M, Li WH. Mathematical model for studying genetic variation in terms of restriction endonucleases. Proc Natl Acad Sci. 1979;76:5269–73.
48. Watterson GA. On the number of segregating sites in genetical models without recombination. Theor Popul Biol. 1975;7:256–76.
49. Pfeifer B, Wittelsbürger U, Ramos-Onsins SE, Lercher MJ. PopGenome: an efficient Swiss army knife for population genomic analyses in R. Mol Biol Evol. 2014;31:1929–36.
50. Tacquet A, Tison F, Devulder B, Roos P. Techniques for decontamination of pathological specimens for culturing mycobacteria. Bull Int Union Tuberc. 1967;39:21–4.
51. Bossé J. Manual of standards for diagnostic tests and vaccines. Can Vet J. 1998;39:183.
52. Coitinho C, Greif G, Robello C, van Ingen J, Rivas C. Identification of mycobacterium tuberculosis complex by polymerase chain reaction of exact tandem repeat-D fragment from mycobacterial cultures. Int J Mycobacteriology 2012;1:146–148.
53. Patel RK, Jain M. NGS QC toolkit: a toolkit for quality control of next generation sequencing data. PLoS One. 2012;7:e30619.
54. Coll F, Mallard K, Preston MD, Bentley S, Parkhill J, McNerney R, et al. SpolPred: rapid and accurate prediction of mycobacterium tuberculosis spoligotypes from short genomic sequences. Bioinforma. Oxf. Engl. 2012;28: 2991–3.
55. Zerbino DR, Birney E. Velvet: algorithms for de novo short read assembly using de Bruijn graphs. Genome Res. 2008;18:821–9.
56. Lin S-H, Liao Y-C. CISA: Contig integrator for sequence assembly of bacterial genomes. PLoS One. 2013;8:e60843.
57. Swain MT, Tsai IJ, Assefa SA, Newbold C, Berriman M, Otto TD. A post-assembly genome-improvement toolkit (PAGIT) to obtain annotated genomes from contigs. Nat Protoc. 2012;7:1260–84.
58. Aziz RK, Bartels D, Best AA, De Jongh M, Disz T, Edwards RA, et al. The RAST server: rapid annotations using subsystems technology. BMC Genomics. 2008;9:75.
59. Laslett D, Canback B. ARAGORN, a program to detect tRNA genes and tmRNA genes in nucleotide sequences. Nucleic Acids Res. 2004;32:11–6.
60. Alikhan N-F, Petty NK, Ben Zakour NL, Beatson SABLAST. Ring image generator (BRIG): simple prokaryote genome comparisons. BMC Genomics. 2011;12:402.
61. Li H, Durbin R. Fast and accurate short read alignment with burrows-wheeler transform. Bioinforma Oxf Engl. 2009;25:1754–60.
62. Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, et al. The sequence alignment/map format and SAMtools. Bioinforma. Oxf. Engl. 2009; 25:2078–9.
63. Robinson JT, Thorvaldsdóttir H, Winckler W, Guttman M, Lander ES, Getz G, et al. Integrative genomics viewer. Nat Biotechnol. 2011;29:24–6.
64. Gene Ontology Consortium: going forward. Nucleic Acids Res. 2015;43: D1049–56.
65. Koboldt DC, Zhang Q, Larson DE, Shen D, McLellan MD, Lin L, et al. VarScan 2: somatic mutation and copy number alteration discovery in cancer by exome sequencing. Genome Res. 2012;22:568–76.
66. Jombart T. Adegenet: a R package for the multivariate analysis of genetic markers. Bioinformatics. 2008;24:1403–5.
67. Krzywinski M, Schein J, Birol I, Connors J, Gascoyne R, Horsman D, et al. Circos: an information aesthetic for comparative genomics. Genome Res. 2009;19:1639–45.
68. Cingolani P, Patel VM, Coon M, Nguyen T, Land SJ, Ruden DM, et al. Using Drosophila Melanogaster as a model for Genotoxic chemical mutational studies with a new program, SnpSift. Front Genet. 2012;3:35.

MATERIAL SUPLEMENTARIO

Additional file 1: Figure S1: Sequencing and genome annotation statistics for the 23 Uruguayan strains of *M. bovis*.

Additional file 2: Table S1: *M. bovis* strains isolated from a bovine host in Uruguay.

Additional file 3: Figure S2: List of tRNAs and corresponding codons in the Uruguayan genomes under study.

Additional file 4: Table S2: Details of the 163 MTBC strains downloaded from NCBI and the 23 Uruguayan strains from this study, used to resolve a phylogenetic tree out of their core genomes.

Additional file 5: Table S3: Resulting blastn of all non-pe/ppe coding sequences against *M. bovis* AF2122/97.

Additional file 6: Figure S3: **Figure S3a** Interactive Genome Visualizer (IGV) visualization of reads aligned to reference genome AF2122/97 (exemplified by strains MbURU-002, MbURU-003 and MbURU-010), showing a region containing glycosyltransferase-coding genes Mb1551 and Mb1553c (wbbL2). Coverage is represented in the upper track of each strain as a gray bar chart, and SNPs are showed as colored bars. The average coverage for these genes can reach up to 10 times the average genome coverage. **Figure S3b**. IGV visualization of reads from strain MbURU-002 aligned to reference genome AF2122/97, showing a region containing polyketide synthase pks12 gene. The lower alignment represents only reads with mapping qualities higher than 0 and show therefore no multimapping reads, while the upper alignment represents the original data from MbURU-002. Coverage is represented in the upper track of each strain as a gray bar chart, and SNPs are showed as colored bars. Note that there are very few called SNPs in this region and some are still present once multimapping reads have been removed.

Additional file 7: Figure S4: **Figure S4a.** Validation of in silico RD typing with PCR on 21 of the Uruguayan M. bovis strains. **Figure S4b**. Alignment of the sequenced reads of strain MbURU-003 against the assembled genome of the same strain. Selected pair of reads in red exemplify one of the reads that flanks both sides of a region of difference (RDbov145a) that is absent in this strain.

Additional file 8: Table S4: List of the over represented genes found in the 23 Uruguayan M. bovis strains.

Additional file 9: Table S5: Putative novel genes identified in the Uruguayan strains by assembly of the unmapped reads.

Additional file 10: Table S6: Regions with high SNP density.

Additional file 11: Table S7: List of truncated genes for the Uruguayan strains studied.

Additional file 12: Figure S5: Bar plots displaying the number of genes affected per strain that are associated to the GO terms carbohydrate metabolic process and peptidoglycan-based cell wall biogenesis.

Additional file 13: Table S8: Details on the distinctive characteristics defining each of the three groups described in this study: URY1, URY2 and URY3.

PARTE 2

La segunda parte del presente trabajo se centra en el análisis en profundidad de la estructura poblacional de *M. bovis* y la búsqueda de un conjunto mínimo de polimorfismos suficientemente informativos para reconstruir la diversidad observada. Se busca que este set de SNPs pueda ser utilizado como marcador y referencia en el futuro frente a la aparición de nuevas cepas de *M. bovis*, permitiendo su rápida caracterización por SNP-*typing*.

HIPOTESIS

Existe un set mínimo de SNPs suficientes para la distinción en sublinajes en *M. bovis* y la caracterización de nuevas cepas, los cuales podrán obtenerse a partir del análisis de los genomas actualmente disponibles.

OBJETIVOS GENERALES

Obtención de un conjunto mínimo de SNPs que describan la diversidad mundial observada de *M. bovis*.

OBJETIVOS ESPECÍFICOS

1. Reconstrucción filogenética de las cepas mundiales disponibles de *M. bovis* a partir de variantes genéticas e investigación de su estructura poblacional.
2. Obtención *in silico* de los patrones de spoligotipado y presencia o ausencia de regiones de diferencia para todas las cepas analizadas.
4. Obtención de los linajes y sublinajes principales en la estructura poblacional de *M. bovis* resultante y validación de estos agrupamientos con la información obtenida en el objetivo anterior.
5. Filtrado de SNPs en base a un criterio de selección definido para obtener un set mínimo final que mantenga la información filogenética original y reconstruya la diversidad, linajes y sublinajes obtenidos con el grupo de SNPs incial.
6. Análisis de un nuevo set de cepas control para predecir sus identidades a partir del conjunto de SNPs elegido y constatar la robustez informativa del mismo.

METODOLOGÍA

Cepas de *M. bovis*

Se obtuvieron un total de 1316 cepas disponibles de *M. bovis* (último acceso: 11 de Marzo del 2017) en la base de datos SRA de NCBI [98] (Illumina MiSeq y HiSeq, promedio de largo de reads = 100bp). Además, se incorporaron las 23 cepas uruguayas utilizadas en la primera parte de este trabajo. De esta manera, se contó inicialmente con un total de 1.339 cepas representativas de distintos orígenes geográficos, incluyendo la mayor diversidad genética disponible.

Filtrado de reads por calidad y alineamiento

Los reads pareados de estas cepas fueron procesados en base a sus calidades de secuenciación con la herramienta NGSQCToolkit (v2.3.3) [99], donde se recortaron las bases con baja calidad (*Phred quality scores*) desde el extremo 3', hasta la aparición de al menos una base con calidad superior a 20 para reducir la incidencia de errores de secuenciación. Luego de esto se filtraron los pares de reads donde al menos uno de ellos presentara una calidad promedio menor a 20. En base a experiencia previa, se definió una cobertura genómica [100] mínima de 30X para evitar sesgos frente a la falta de información. Se eliminaron aquellas cepas cuyos genomas presentaran una cobertura inferior a este valor luego de recortar y filtrar sus reads (ver script). Como resultado, se obtuvieron un total de 1.217 cepas de *M. bovis* de calidad y profundidad suficiente para realizar los análisis propuestos.

Posteriormente, los reads fueron alineados mediante BWA (v0.7.12-r1039) [101] al genoma de referencia humano (versión GRCh38, número de acceso GCF_000001405.28), permitiendo hasta 3 *mismatches*, de modo de eliminar aquellos reads de origen humano que pudieran estar contaminando las muestras. De los reads restantes, se realizó alineamiento contra la referencia de *M. bovis*, AF2122/97 (número de acceso NC_002945), permitiendo nuevamente hasta 3 errores por secuencia de 15bp. Se filtraron los reads de acuerdo con sus propiedades de mapeo mediante la herramienta samtools (v0.1.18) [102], manteniendo solo los pares de reads mapeados. El pipeline de alineamiento y su filtrado se puede encontrar en este repositorio público.

SNP-*calling* y anotación de variantes

A partir de los alineamientos se realizó la obtención de variantes utilizando samtools mpileup y VarScan (v2.3.7) [103]. Se consideraron como variantes reales a aquellas que presentaran más de 20 reads de soporte y una frecuencia mínima del 20% (ver script), resultando en un total de 20.097 SNPs para las 1.217 cepas analizadas. Las variantes encontradas fueron luego anotadas funcionalmente con el software SnpEff (v4.2) [104]. Este programa utiliza el genoma de referencia

para la anotación, e informa sobre el efecto que las variantes producen y el gen o genes que afecta, por ejemplo (ver script).

Genotipado *in silico*

Todas las cepas fueron genotipificadas *in silico* para obtener una predicción de los *spoligotipos* utilizando el programa SpoTyping [105] a partir de sus reads. SpoTyping presenta una mayor proporción de predicciones acertadas y menor cantidad de cepas sin predicción asignada que el más conocido SpolPred [106].

Dada la ausencia de programas para la obtención de nuevas regiones de diferencia (no reportados previamente), se desarrolló un conjunto de scripts para la obtención de esta información. A partir de los alineamientos se extrae la cobertura de reads posición a posición en cada cepa con respecto a la referencia AF2122/97, utilizando la opción *genomecov* de bedtools (ver script). Esta información es luego utilizada como *input* para obtener todas las regiones del genoma de un largo superior a x que tienen cobertura menor o igual a 1 (un read), siendo x un entero definido por el usuario (por defecto = 75bp, ver parseRD.py). Una vez obtenidos estos *gaps* genómicos por cepa, se integra la información de todos los individuos con el fin de obtener RDs comunes entre grupos (ver script). Para esto se siguen los siguientes pasos:

1.- Para cada cepa, se unen aquellos *gaps* encontrados a una distancia menor a 100bp, que corresponde al largo aproximado de un read para la mayoría de las cepas analizadas, de forma de aumentar la sensibilidad.

2.- Se filtran los *gaps* por tamaño, conservando como RD sólo aquellos mayores a un largo definido por usuario (por defecto = 1.000bp, en base a reportes previos de otros RDs).

3.- Se almacena el tamaño, posición de inicio, posición de fin y cepa de todos los RD individuales encontrados en las cepas analizadas y se ordena el set total de forma ascendente de acuerdo con la posición de inicio.

5.- Dado un RD x_0, se define como solapante potencial S al conjunto de RDs individuales posteriores a x_0 cuyas posiciones de inicio en el genoma son menores a la posición de fin de x_0, de la siguiente forma:

$$S = \{ x, x \in C \mid x_{(inicio)} < x_{0\,(fin)} \},$$

siendo C el conjunto total de RDs identificados. Estos solapantes potenciales serán candidatos para representar un único RD común.

6.- Se define el tamaño solapado dentro de un conjunto S. Para cada RD individual x perteneciente

al conjunto S, si el tamaño de *x* es mayor al 95% del tamaño solapado, se actualiza tanto el tamaño del inicio como del fin solapados, manteniendo a *x* en el conjunto S. Si es menor al 95% del tamaño solapado, se elimina este x del conjunto S.

7.- Una vez recorrido un bloque entero de solapantes potenciales, se almacena la información resultante del RD común, asignando un ID único al mismo ("RDinicio_fin").

8.- Para cada RD individual *x*, si este fue ingresado y posteriormente removido de un conjunto S, se calcula un nuevo bloque a partir de este RD.

9.- Se repite el paso 5 para la totalidad de los RDs individuales restantes.

Los RDs obtenidos serán luego filtrados en base a su asociación con linajes y sublinajes dentro de la estructura poblacional observada de *M. bovis*.

Reconstrucción filogenética

Se reconstruyó una filogenia a partir de las 20.097 variantes utilizando el programa basado en máxima verosimilitud RAxML (v8.2.7) [107], con el modelo gamma de sustitución (GTRGAMMA). *M. caprae* fue elegido como grupo externo, al ser la especie más cercana dentro del linaje de MTBC (número de acceso: SRR1792164). Se constató el soporte de la filogenia resultante efectuando 100 iteraciones de *bootstrapping*. La filogenia obtenida fue visualizada y anotada utilizando el paquete ETE Toolkit 2 [108], integrándola con metadatos relativos a ubicación geográfica, spoligotipos y RDs. Para esto se utilizó el script implementado por Wong et al. 2016 [109] con modificaciones para mejorar la visualización de los datos (colores de fondo, ocultar ramas y ordenar ramas; ver plotTree.py).

Para discernir claramente las relaciones entre los clados de la filogenia global, se reconstruyó la estructura básica de su topología utilizando un representante por cada grupo de tercer nivel de jerarquía sugerido por BAPS (ver abajo), eliminando las restantes hojas del árbol mediante la función *tree.prune*(["*rama1*", "…", "*ramaN*"]) del paquete ETE Toolkit.

Identificación de linajes y sublinajes

Se realizó un estudio de la estructura poblacional de *M. bovis* utilizando además una aproximación exenta del contexto filogenético descrito anteriormente. Para localizar los principales clusters dentro de la estructura poblacional observada se utilizó una variante del análisis Bayesiano implementado en el programa BAPS (*Bayesian Analysis of Population Structure*) [110]. Este programa establece la mejor partición de un set dado de individuos en sub poblaciones bajo el supuesto de admixia (donde pueden existir individuos con una composición genética proveniente

de más de una de las poblaciones establecidas), y posteriormente asigna a estos individuos en los grupos o poblaciones estimados, sin suponer poblaciones predefinidas. Si bien el MTBC es altamente clonal, asumimos un modelo que supone admixia ya que existen reportes de recombinación dentro de este complejo [111–113]. La variante de BAPS utilizada, denominada hierBAPS [114], realiza un agrupamiento jerárquico de las cepas de acuerdo a la similitud de sus secuencias, donde un grupo estimado en cualquier nivel de la jerarquía será a su vez reagrupado en el próximo nivel.

En este análisis se utilizaron las variantes obtenidas para cada cepa con el fin de determinar los linajes y sublinajes responsables de la estructura poblacional observada. El análisis de clusterización jerárquico se realizó en 5 corridas independientes, eligiéndose un límite superior de clusters que varió entre 5 y 10 en las diferentes corridas, utilizando 3 niveles de jerarquía. Los resultados obtenidos fueron comparados con la filogenia previamente estimada y con los datos disponibles de RDs.

Se denominaron como sublinajes a los clusters de segundo nivel jerárquico, los cuales a su vez pertenecen a linajes representados por los clusters de primer nivel jerárquico. Con el fin de mantener una consistencia con la filogenia, se realizaron modificaciones menores en algunos de los sublinajes obtenidos mediante BAPS, específicamente uniendo a aquellos localizados en zonas con bajos valores de *bootstrap* dentro de la filogenia.

Paralelamente, la identificación de los complejos clonales conocidos se realizó en base a la ausencia de los espaciadores característicos antes mencionados dentro del patrón de spoligotipos (ver Estructura Poblacional de M. bovis).

Identificación de SNPs específicos de (sub)linajes para genotipado

Se filtraron las variantes de todas las cepas en estudio, de forma de alcanzar un set final de SNPs capaz de reconstruir la diversidad filogenética observada con el set total de SNPs. Ya que M. *bovis* es un grupo altamente homogéneo, la cantidad de SNPs promedio por cepa con respecto a una referencia no es tan numerosa como lo observado entre todo el MTBC [68]. Por ello, se determinó que para la obtención de SNPs específicos de sublinajes se requiere el cumplimiento de una de las siguientes condiciones (ver filterSNPs.py):

1.- La variante debe estar presente exclusivamente en todos los individuos de un (sub)linaje, sin presentarse en individuos que no pertenecen a este (SNPs exclusivos).

2.- La variante debe estar presente en todos los individuos de un (sub)linaje, y no debe encontrarse en todos los individuos de otro (sub)linaje (SNPs distintivos).

3.- La variante debe encontrarse en todos los individuos de 2 o 3 (sub)linajes completos.

Luego, se discriminan los tres conjuntos de SNPs en base al efecto de la variante, priorizando a los SNPs sinónimos. Se buscó representar a cada sublinaje (total = 16) con al menos 2 SNPs cada uno. Para esto se seleccionaron todas las variantes sinónimas exclusivas (condición 1), que aseguran la especificidad de los sublinajes. Luego, a los linajes sub representados se le asignaron al azar variantes sinónimas características (condición 2) y, en el caso de seguir existiendo sublinajes sin variantes representativas, a estos se le asignaron al azar variantes sinónimas de la condición 3. Como caso extremo, cuando un sublinaje no es capaz de ser diferenciado de otro por falta de variantes específicas, recurrimos a seleccionar variantes con modificaciones no sinónimas en el mismo orden de prioridad dentro de las condiciones estipuladas anteriormente (esto se realizó solamente en dos clados). Se identificó un set mínimo de 56 SNPs capaz de genotipar cepas de *M. bovis* dentro de 5 linajes y 16 sublinajes.

Incorporación de nuevas cepas control para predicción de su linaje

Se incorporaron 21 cepas nuevas para predecir sus genotipos y testear la capacidad discriminante del set de SNPs elegido (Anexo 1). Estas cepas fueron procesadas como se detalló anteriormente (filtrado de reads, genotipado *in silico* y SNP-*calling*), y se realizó una reconstrucción filogenética a partir del nuevo set de variantes totales (i.e. las X variantes descritas más las Y obtenidas a partir de las 21 cepas), utilizando las configuraciones mencionadas previamente.

Puesta a prueba de la robustez informativa del set mínimo de SNPs

Se implementó el *script* *BovisCode* dedicado a la búsqueda de aquellas variantes pertenecientes al set mínimo de SNPs obtenido en este trabajo. Este programa busca y compara la presencia y ausencia de cada una de estas variantes, lo cual busca dar por resultado una identidad inequívoca en linaje y sublinaje. Esto se aplicó para cada una de las 21 nuevas cepas. A partir de esta información, los genotipos predichos fueron comparados con los genotipos identificados en base al set completo de variantes.

Acceso público de *scripts* implementados

Los pipelines y scripts desarrollados para este objetivo se encuentran disponibles en `https://github.com/emmaielle/thesis_mbovis_genomics-part2`.

RESULTADOS

Estructura poblacional de *M. bovis*

Se utilizaron datos de secuenciación masiva de todas las cepas disponibles de *M. bovis* en bases de datos internacionales (n = 1.217) para visualizar la diversidad existente entre estos ejemplares e identificar la estructura poblacional y los linajes y sublinajes que la conforman. Las características detalladas de estas cepas, incluyendo sus spoligotipos *in silico*, se encuentran en el Anexo 1.

Se identificaron un total de 20.097 variantes; de estas, 41.8% se encontraron en cepas únicas. Por otro lado, el 88% del total pertenecen a regiones codificantes y de estos, el 57% ocasionan cambios no-sinónimos (las proporciones de las variantes identificadas se ilustran en el Anexo 2). El árbol filogenético resultante del alineamiento de estas variantes evidenció una estructura global congruente con la clasificación de complejos clonales definida hasta el momento [115]. Si bien en este análisis no se vieron representados todos los complejos clonales, los complejos BCG-like (n = 3; 0.24%), Eu1 (n = 1.090; 89.4%), Eu2 (n = 117; 9.3%) y Af2 (n = 2; 0.16%) se segregaron claramente dentro del árbol final, observándose una mayoría de cepas pertenecientes a Eu1.

La integración de la topología filogenética con la inferencia poblacional realizada por BAPS permitió identificar 5 linajes que definen la diversidad existente y se dividen en 2 o 3 sublinajes cada uno, alcanzando un total de 13. Cada genotipo se denominó utilizando una nomenclatura descriptiva de linaje y sublinaje (por ejemplo, el linaje 3 presenta dos sublinajes: 3.1 y 3.2, donde los números son asignados en forma ascendente en base a la historia evolutiva evidenciada filogenéticamente) (Figura 1).

Los spoligotipos se mantuvieron coherentes con la información filogenética obtenida, con excepciones específicas que incluyen los patrones SB0484, SB1757, SB1812, SB0292 y SB1040, los cuales se observaron entre 2 y 5 cepas. Estos indicios de homoplasia compusieron inesperadamente una proporción mínima dentro del muestreo total (aproximadamente 1%). Estadísticas sobre la incidencia de spoligotipos se resumen en el Anexo 3. En estas se puede observar la frecuencia de los spoligotipos, en particular SB0130, SB0673, SB0145 y SB0140 en su conjunto componen más de la mitad del muestreo total.

La incidencia de RDs permitió a su vez respaldar los genotipos observados, si bien muchos de los clados no presentaron RDs características. Utilizando el script desarrollado para la identificación de RDs detallado en Métodos, se encontraron 6 nuevas regiones principales. Estas RDs, y otras

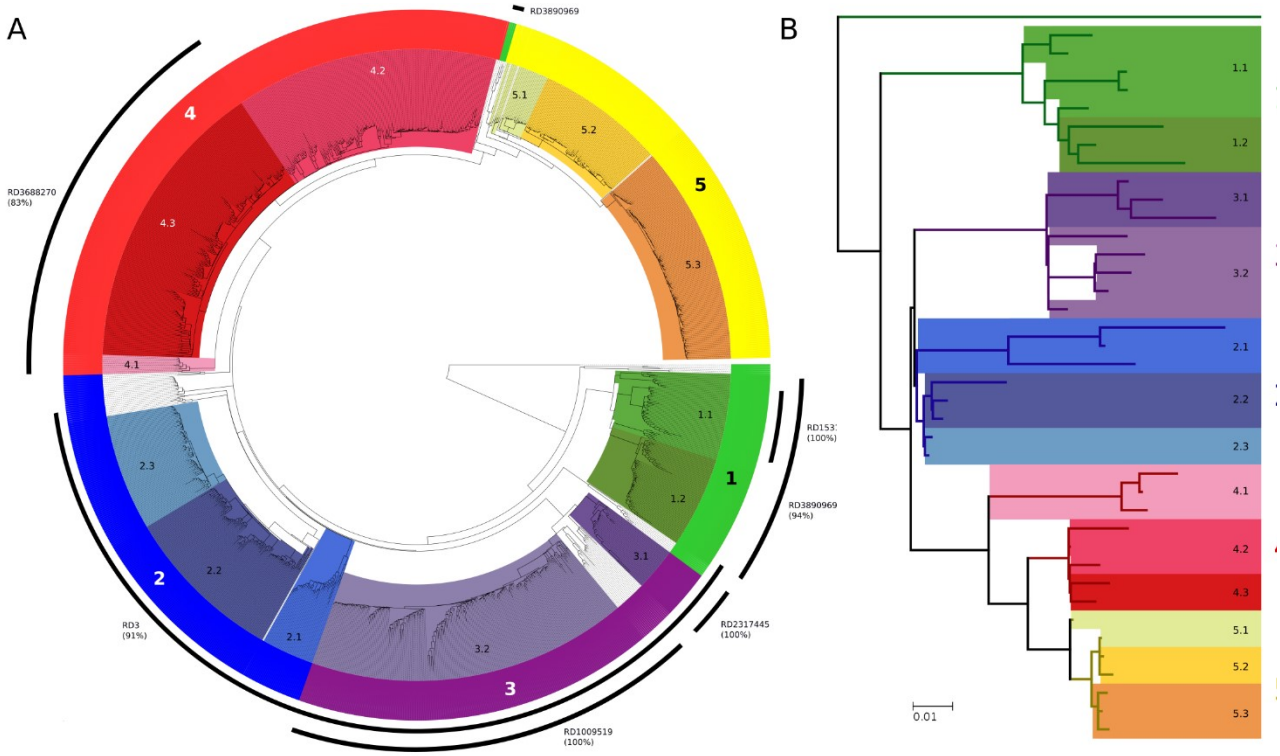

Figura 1.- Estructura poblacional de *M. bovis* en base a variantes genómicas. **(A)**- Filogenia reconstruida a partir de 20.097 variantes entre los 1.217 aislados de *M. bovis*, utilizando a *M. caprae* como grupo externo. Se identifican en el anillo externo los cinco linajes principales y en color de fondo los sublinajes que contiene cada uno. Los arcos negros señalan RDs específicas de (sub) linajes, junto con un porcentaje de especificidad del clado. **(B)**- Esqueleto de la topología encontrada en (A) para mostrar las relaciones entre los sublinajes. Cada hoja representa un clado diferente identificado dentro de los sublinajes.

que no son determinantes de (sub)linajes, se detallan en el Anexo 4, donde a su vez se puede observar la distribución de todas los RDs identificadas en la filogenia mundial.

El linaje 1 (verde en Figura 1) representa el conjunto de cepas más ancestral de *M. bovis*, contiene representantes del complejo Af2, BCG-like y Eu2, y está compuesto por 2 sublinajes y un subgrupo de cepas sin suficiente diferenciación como para formar un sublinaje propio. Este subgrupo presenta la mayor variabilidad de cepas, con spoligotipos BCG-like, Af2 y Eu2, aunque en muy baja representación. Las dos cepas Af2 encontradas presentan la deleción RDAf2 característica de este complejo clonal (denominada en este trabajo RD681589) y ausencia de espaciadores 3 al 7. Se observó también que estas dos cepas de origen asiático comparten de forma exclusiva la ausencia de un total de cuatro RDs. Un mayor número de cepas analizadas podría permitir la identificación de otro clado de BAPS que separara a estas cepas del subgrupo antes mencionado. La filogenia observada muestra algunas incongruencias con la designación de este linaje realizada por BAPS, ya que algunos de sus representantes están localizados como descendientes del linaje 5. La

presencia de una deleción común a la mayor parte del linaje 1 sugiere que estos representantes, constituidos por cuatro cepas, presentan variantes únicas al linaje 5 que los situaron incorrectamente en la topología filogenética. El sublinaje 1.1, caracterizado por el spoligotipo SB0121 y sus descendientes SB1308 y SB1345, pertenece al complejo Eu2. Un clado de este sublinaje carece de la región RD1531687, de ~12.000 bp. El sublinaje 1.2 también corresponde a Eu2, y se caracteriza por presentar los patrones SB0121 y SB0265. Los sublinajes 1.1 y 1.2 comparten la ausencia de la región RD3890969 (~3.000 bp).

La deleción de la región RD3 (en este trabajo identificada como RD1766212) es una característica común del linaje 3 y la mayoría de las cepas del linaje 2 (violeta y azul en Figura 1, respectivamente). Esta región previamente descrita [116], contiene ORFs del fago phiRv1. Las cepas del linaje 2 que no comparten un ancestro común con el linaje 3 no contienen esta deleción. Por otro lado, RD3 también está ausente en algunas cepas pertenecientes a otros grupos (algunas cepas de 1, 4.2, 4.3 y 5.2). La información filogenética aportada por esta RD ha sido cuestionada ya que esta región se ha perdido múltiples veces en la historia evolutiva de *M. bovis* [77], si bien se la observa claramente representando a dos de los grandes linajes que encontramos en esta especie. El linaje 2 se caracteriza por presentar spoligotipos derivados de SB0130 y se divide en tres sublinajes, sin presentar deleciones específicas. Como descendiente del linaje 2, el linaje 3 abarca dos sublinajes y dos deleciones mayores: RD2317445, incidiendo sobre dos clados de 3.1 (~1.200bp), y RD1009519, y abarcando todo 3.2. Este último sublinaje se encontraba originalmente conformado por dos grupos BAPS, pero fue integrado en uno sólo debido a la baja congruencia filogenética evidenciada por muy bajos valores de *bootstrap* (ver reconstrucción filogenética original en el Anexo 5). El linaje 3 se caracteriza por presentar spoligotipos derivados del SB0145.

El linaje 4 se divide en tres grandes grupos (rojo en Figura 1), de los cuales 4.1 y 4.3 comparten la deleción de RD3688270 (~1.000 pb). Este linaje se caracteriza por presentar patrones de spoligotipo descendientes de SB0140. El sublinaje 4.1, externo al resto del linaje, está representado exclusivamente por el spoligotipo SB1499, mientras que 4.3 contiene cepas tipo SB0673, patrón primariamente estadounidense. El sublinaje 4.2 contiene una mayor variedad de spoligotipos y se presenta como ancestral tanto del sublinaje anterior como de todo el linaje 5.

Finalmente, el grupo más derivado es el linaje 5, el cual contiene los sublinajes 5.1 (al que pertenece la referencia AF2122/97) a 5.3. Es un linaje homogéneo y aparenta un origen más reciente ya que está constituido casi exclusivamente por el patrón SB0140. Esto puede deberse a una falta de muestreo dentro de este linaje, ya que está constituido predominantemente por representantes de Nueva Zelanda e Irlanda del Norte. No se encontraron deleciones particulares para este grupo.

Identificación de un set mínimo de SNPs para genotipado de *M. bovis*

Del total de 20.097 variantes identificadas se obtuvo un subconjunto de 294 SNPs exclusivos de sublinajes y 675 SNPs distintivos de sublinajes. También se identificaron 488 SNPs que se encuentran entre dos y tres sublinajes, para aumentar la resolución del set mínimo de SNPs final en aquellos clados ambiguos. La cantidad de variantes entre las diferentes categorías funcionales para cada uno de estos subgrupos se mantuvo proporcional (Anexo 2) y las variantes elegidas para conformar el set mínimo fueron seleccionadas al azar. Se priorizó la selección de variantes sinónimas, aunque debido a la falta de variabilidad en el sublinaje 4.3 se debió incluir un SNP no sinónimo en el mismo. Para la reconstrucción de la estructura poblacional observada y el genotipado de cepas futuras, se identificó un set mínimo de 56 SNPs, cuyas características se detallan en el Anexo 6. Como se observa en la Figura 2, una reconstrucción filogenética a partir de este set de SNPs evidenció los mismos clados que aquellos identificados utilizando toda la

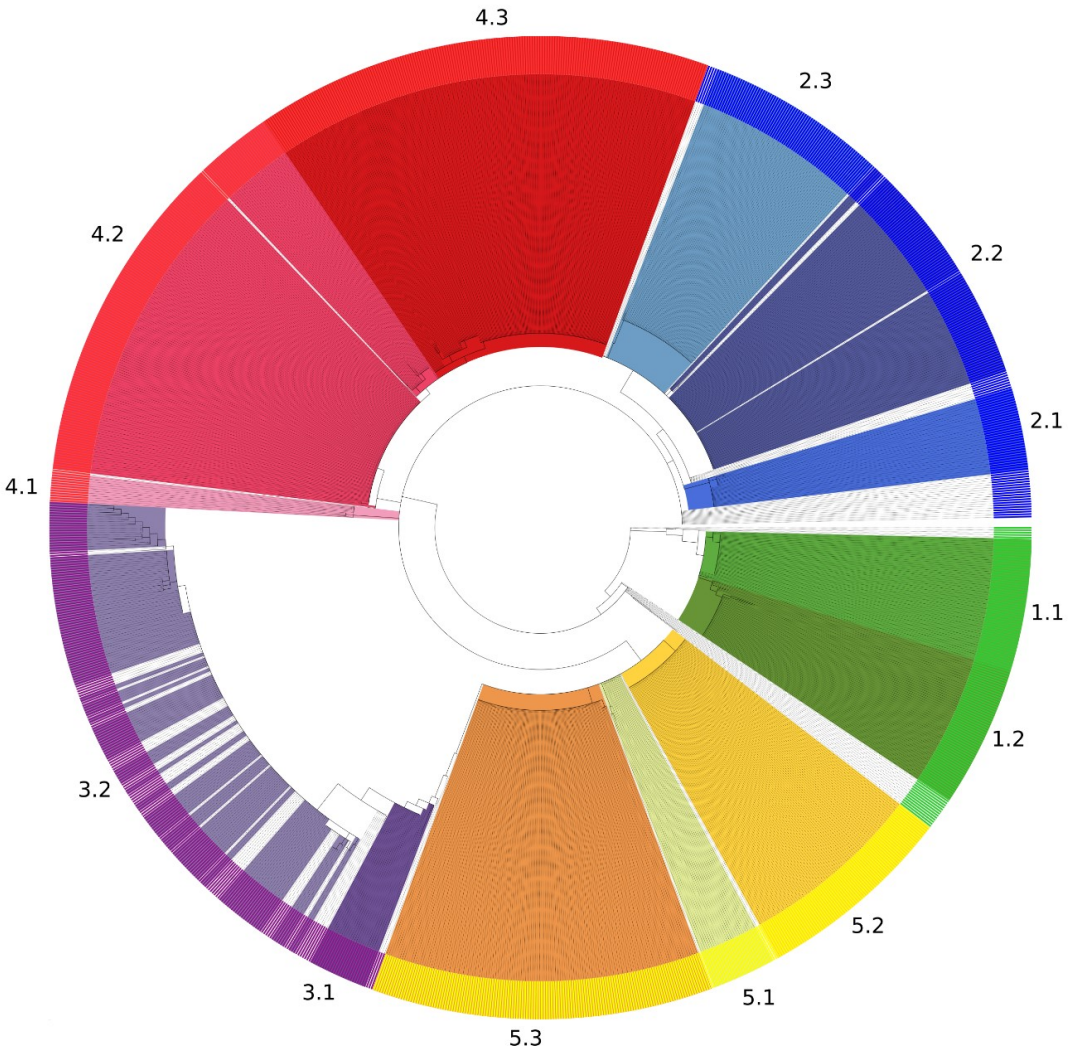

Figura 2.- Filogenia global de *M. bovis* reconstruida a partir de un set mínimo de 56 SNPs informativos.

variabilidad observada a partir del set completo, pudiendo genotipar las cepas de *M. bovis* dentro de los cinco linajes primarios, discriminando hasta el nivel de sublinaje. Todos los sublinajes definidos previamente presentan una segregación clara.

Predicción de linajes en cepas control

Se introdujeron 21 cepas nuevas de *M. bovis* para corroborar la robustez del set de SNPs elegido para el genotipado rápido de cepas. En la Figura 3 se observa la reconstrucción filogenética a partir de las variantes del total de 1.238 cepas, indicando con negro a las nuevas incorporaciones. La estructura poblacional de esta nueva filogenia permaneció fiel a los resultados anteriores, a excepción de tres cepas dentro del Linaje 4, que ocuparon una posición más derivada que el resto de los sublinajes a los que pertenecen de acuerdo con el resultado de BAPS.

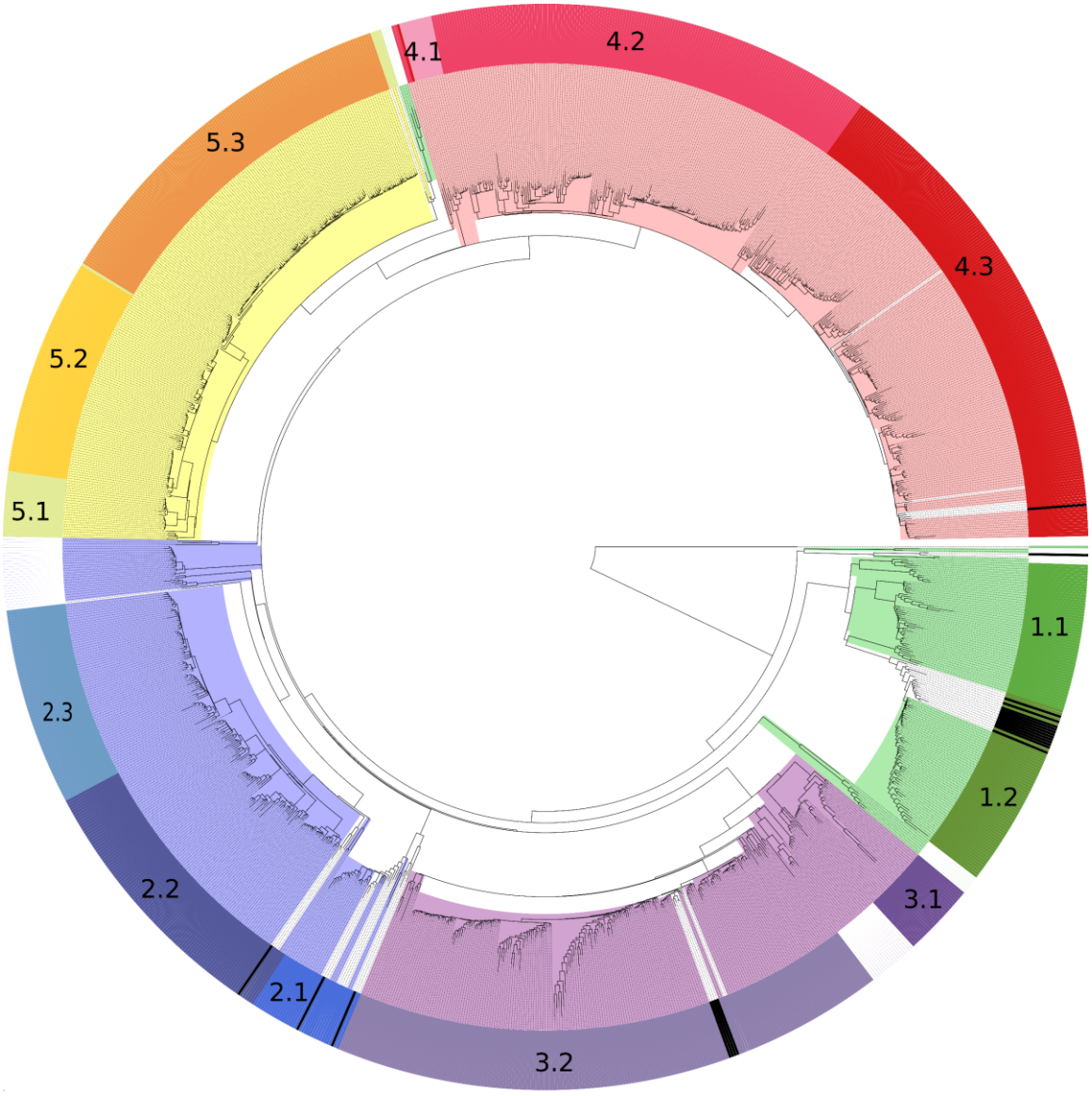

Figura 3.- Filogenia actualizada con la incorporación de 21 nuevas cepas para la posterior predicción de sus linajes (en negro). Se encontraron trece cepas del Linaje 1, tres cepas del Linaje 2, cuatro del Linaje 3 y una del Linaje 4.

Los linajes y sublinajes de estas nuevas cepas, obtenidas en base a su ubicación en la filogenia completa, fueron contrastadas con la información obtenida por el set de SNPs mínimo y *BovisBarcode*. Ambos resultados se comparan en la Tabla 1, donde se puede observar un acierto total de las predicciones a nivel de linaje. Para el sublinaje, se logró un 48% de aciertos exactos y 52% de aciertos parciales (donde el resultado abarca dos posibles sublinajes).

	Variantes totales		**SNP barcode**	
	Linaje	**Sublinaje**	**Linaje**	**Sublinaje**
SRR5485729	1	1.1	**1**	**1.1**
SRR1791776	1	1.3	**1**	1.2, 1.3
SRR1792163	1	1.3	**1**	1.2, 1.3
SRR1792173	1	1.3	**1**	1.2, 1.3
SRR1792188	1	1.3	**1**	1.2, 1.3
SRR1792205	1	1.3	**1**	1.2, 1.3
SRR1792207	1	1.3	**1**	1.2, 1.3
SRR1792249	1	1.3	**1**	1.2, 1.3
SRR1792251	1	1.3	**1**	**1.3**
SRR1792258	1	1.3	**1**	1.2, 1.3
SRR1792263	1	1.3	**1**	1.2, 1.3
SRR1792486	1	1.3	**1**	1.2, 1.3
SRR1792495	1	1.3	**1**	**1.3**
SRR1791795	2	2.1	**2**	**2.1**
SRR1792077	2	2.1	**2**	**2.1**
SRR1792176	2	2.2	**2**	**2.2**
SRR1791828	3	3.2	**3**	**3.2**
SRR1792069	3	3.2	**3**	**3.2**
SRR1791940	3	3.2	**3**	**3.2**
SRR5817715	3	3.2	**3**	**3.2**
SRR1791771	4	4.3	**4**	4.2, 4.3

Tabla 1.- Comparación de la identidad de nuevas cepas (representadas por sus números de acceso en el SRA), obtenida a partir tanto de su ubicación filogenética (columnas 2 y 3) como de la predicción realizada con un set de 56 SNPs seleccionados (columnas 4 y 5).

DISCUSION Y CONCLUSIONES

La incidencia de *M. bovis* representa un riesgo tanto para el bienestar animal como para la salud pública mundial. Sin embargo, hay escasa información relativa a los genotipos que presenta este patógeno y a su estructura poblacional. Estudios previos han identificado cuatro complejos clonales: de distribución mundial (Eu1) y de distribución regional (Eu2, Af1 y Af2), y un complejo clonal ancestral (BCG-like) [82–85,115], pero no existen estudios sobre la estructura poblacional interna de los mismos. Esto representa una limitante para la búsqueda de estrategias de genotipado rápidas y accesibles.

En este trabajo se reporta la estructura poblacional de *M. bovis* a nivel mundial con un alcance no descrito previamente. A partir de su variabilidad genómica (SNPs y RDs) se establecieron relaciones filogenéticas entre 1.217 cepas de *M. bovis*. Esto permitió la identificación de 5 linajes principales y 13 sublinajes. Los linajes obtenidos filogenéticamente son congruentes con los obtenidos poblacionalmente hasta el nivel de sublinaje con escasas excepciones, e introducen una clasificación distinta de *M. bovis*.

Si bien las cepas utilizadas disponibles en la base datos NCBI son de orígenes variados, se observa claramente la existencia de un sesgo geográfico inevitable asociado al propósito de los proyectos de secuenciación disponibles. Una gran proporción de las cepas disponibles tienen origen norteamericano (Estados Unidos 38% y México 17%), mientras que el resto de las cepas se distribuyen más homogéneamente entre los países restantes. Aun así, la inclusión de todos los representantes de *M. bovis* disponibles al momento de este trabajo describe una estructura poblacional que será de utilidad para análisis posteriores ya que sugiere una contraparte interesante a nivel genómico y posiblemente fenotípico.

Los complejos clonales conocidos presentan una direccionalidad evolutiva que se origina desde los complejos Af2 y BCG-like como grupo más basal y próximo al grupo externo (*M. caprae*), el cual diverge en los complejos Eu2 y Eu1. Este último complejo representa el grupo cosmopolita de mayor diseminación geográfica. En el Anexo 3 se delinean estos complejos dentro de la representación filogenética. A su vez se identificaron un conjunto de cepas que no presentan las características esperadas dentro de los complejos clonales conocidos. Por un lado, algunas cepas no pertenecen a ninguno de estos complejos, en base a las características genómicas que los definen. Entre estas se encuentran cepas con el spoligotipo SB1069, las cuales pertenecen a un subgrupo no identificado dentro del linaje 1 y presentan la ausencia de espaciadores 3 al 7 característica de Af2, pero no presentan la deleción RDAf2. Esta observación se ha reportado recientemente, ponderando la existencia de otros complejos no descritos y promoviendo una

revisión de la clasificación actual [117]. Frente a esta problemática, la diversidad presentada en este trabajo es evidencia de la resolución que se puede alcanzar en estudios poblacionales cuando se utiliza toda la información genética disponible.

Por otro lado, reportamos el primer caso de dos ejemplares pertenecientes al complejo clonal Af2 con un origen distinto al comúnmente encontrado [118,119]. En este caso, las dos cepas (SRA SRR1791710 y SRR1791712) son provenientes de China y fueron extraídas de un hospedero primate. Resultaría interesante analizar con mayor detalle la historia geográfica de estas cepas, ya que presentan las señales que caracterizan a este complejo (RDAf2 y ausencia de espaciadores 3 a 7), pero Af2 es reconocido por presentar una extensión geográfica reducida exclusivamente a África del Este. Además, estas dos cepas presentan otras cuatro RDs características que podríamos corroborar como exclusivas del complejo clonal si tuviéramos un muestreo mayor.

Las RDs identificadas fueron útiles para determinar el conjunto de linajes y sublinajes dentro de la muestra analizada si bien existen algunos inconvenientes resultantes de la metodología de detección de RDs concebida en este trabajo. Un ejemplo es el de la región RD3688270, que está ausente en los sublinajes 4.1 y 4.3; ya que 4.1 es ancestral al resto del linaje 4, la pérdida de un RD se debería evidenciar en todas las cepas derivadas. Esto no ocurre en 4.2, lo cual sugiere que esta región tiene la tendencia a perderse de manera paralela en distintos grupos. Otras RDs con estas características se han observado en la muestra analizada, como lo es RD3. Los polimorfismos de secuencia larga (LSP) en *M. tuberculosis* se clasifican en tres grupos de acuerdo con su capacidad de originarse más de una vez de forma independiente: El grupo A es el único que se considera como marcador filogenético preferible, ya que puede ser explicado por un evento único en un ancestro común, mientras que los grupos B y C representan polimorfismos resultantes de presiones selectivas y se han originado al menos en dos (B) o en innumerables (C) ocasiones. Muchos LSP de los grupos B y C se encuentran flanqueados por elementos IS6110 o contienen secuencias ricas en GC. La recombinación originada por ambas características genómicas podría ser responsable de la recurrencia de los LSP. Ningún LSP perteneciente al grupo A, por el contrario, presenta estas características [120]. Se conoce que la información filogenética que aporta RD3 es cuestionable por su incidencia en diferentes puntos de la historia evolutiva de *M. bovis* [77]. Teniendo en cuenta esta clasificación, y extendiendo su alcance a *M. bovis*, la región RD3688270 arriba mencionada podría no ser adecuada para su uso en estudios filogenéticos y de poblaciones, aunque es útil como un sondeo preliminar de los grupos de cepas relacionados entre sí. Podría ser interesante analizar la composición nucleotídica de todos los RD identificados para confirmar su utilidad filogenética, en busca de secuencias ricas en GC y/o elementos IS6110

flanqueantes. Se esperaría que si una región pertenece al grupo A no presente ninguna de estas características [120].

Otro inconveniente que se origina de esta estrategia es la identificación errónea de RDs, en donde se predicen falsas regiones ausentes por distintas razones. Algunas de estas incluyen baja cobertura de reads ya sea por baja cobertura genómica como por errores de secuenciación. Por otro lado, el alto porcentaje de identidad utilizado como umbral para definir un RD único produce regiones que van disminuyendo en tamaño como resultado de la comparación y solapamiento de gaps. Esto resulta en regiones cuyo tamaño y coordenadas no coinciden exactamente con lo que se conoce experimentalmente. Este es el caso de RD3, el cuál se define en la literatura como una región de aproximadamente 9.000 pb, mientras que el tamaño obtenido *in silico* con esta estrategia es de 7.000 pb. A efectos filogenéticos, la diferencia observada no afecta el resultado, ya que la prioridad en este caso es la de encontrar cepas que comparten la región. Sin embargo, sería necesario corroborar experimentalmente las coordenadas de cada uno de los RDs identificados y, a futuro, mejorar la sensibilidad del algoritmo. Cabe destacar que no existe disponible hasta la fecha ningún método de predicción *in silico* de RDs *de novo*, y lo más semejante permite encontrar *in silico* regiones de diferencia previamente descritas a partir de datos genómicos [121].

De la misma manera que en *M. tuberculosis* se obtuvieron 62 SNPs, se designó un panel de 56 SNPs capaces de reproducir la filogenia de la especie. En comparación, es una cantidad total equivalente, pero esto no es indicador de una similitud en la diversidad de ambas especies. Al contrario, *M. tuberculosis* presenta una mayor variabilidad y heterogeneidad genética, lo cual facilita la obtención de variantes con alto poder discriminativo entre linajes. *M. bovis* presenta una historia evolutiva más reciente y un menor número de variantes entre sus integrantes. Esto reduce las posibilidades de identificar variantes útiles para la caracterización inequívoca de (sub)linajes. Los 56 SNPs obtenidos para *M. bovis* logran en conjunto segregar las cepas dentro de la clasificación encontrada, pero es importante destacar que el estudio en *M. tuberculosis* localizó exitosamente un panel de 62 SNPs de 413 alternativas.

Nuestro panel de *M. bovis* fue probado en un set de cepas control, las cuales fueron situadas dentro de la estructura descrita con un nivel de acierto satisfactorio. Este set control, sin embargo, es muy reducido y fue seleccionado en base a nuevos proyectos de secuenciación que disponibilizaron luego de la obtención del set inicial. Por esta razón, si bien observamos que las mismas pertenecen a diversos de los linajes descritos, debemos reconocer el sesgo existente dentro de una muestra que deberá ser expandida en el futuro.

La designación de un set mínimo de SNPs pretende que pueda ser utilizado como referencia en el futuro para su caracterización por SNP-*typing*. Las aplicaciones que surgen de esto son numerosas, esperándose que su utilización permita conocer rápidamente las características epidemiológicas de un animal infectado y así mantener bajo control su incidencia y diseminación. Esto será así una vez que se cuente con una asociación fenotípica a los diferentes linajes y sublinajes descritos.

En conclusión, exploramos la diversidad de *M. bovis* y su estructuración poblacional, contrastando esta información con el conocimiento disponible de complejos clonales descritos. Si bien algunos complejos no tuvieron una representación muestral grande, lo fue suficiente para establecer una potencial historia evolutiva de los mismos. Se desarrollaron además dos metodologías *in silico* de tipificado con alto potencial: una para la identificación de regiones de diferencia *de novo* y otra para la clasificación de cepas basada en el set de SNPs descrito anteriormente. Ambos algoritmos pueden ser optimizados, pero establecen una base inicial para la tipificación de cepas *in silico* de *M. bovis*.

PERSPECTIVAS

Es de interés probar el set mínimo de SNPs en un número mayor de cepas control para consolidar la robustez filogenética que este set presenta. La incorporación de metadatos dentro de la estructuración poblacional obtenida es altamente necesaria, por lo cual simplemente se podría probar este set dentro de un conjunto de cepas con información fenotípica y epidemiológica asociada.

Si bien los resultados obtenidos aquí representan un aporte novedoso a la clasificación de *M. bovis*, los datos de los que partimos deberían tener una distribución de los diferentes complejos clonales más equitativa. Esto permitiría por un lado la obtención de una mayor cantidad de SNPs únicos para los sub linajes que estaban menor caracterizados. Por otro lado, posibilitaría la dilucidación de posibles sub linajes no descritos aquí por falta de representantes que evidenciaran la existencia de clados únicos.

Finalmente, es necesario analizar las RDs identificadas en este trabajo para determinar si son filogenéticamente informativas. Es decir, tanto sus regiones flanqueantes como su contenido en GC en búsqueda de *hotspots* de recombinación. Independientemente de si son informativas para este análisis, también resultaría interesante confirmar experimentalmente sus coordenadas de inicio y fin para aportar al conocimiento de las RDs. Esto va de la mano con la posibilidad de modificar el algoritmo desarrollado para mejorar la predicción de las mismas.

REFERENCIAS

1. Whitman WB, Goodfellow M, Kämpfer P, Busse H-J, Trujillo ME, Ludwig W, et al. Bergey's Manual of Systematic Bacteriology: Volume 5: The Actinobacteria. Springer Science & Business Media; 2012.

2. van Soolingen D, Hoogenboezem T, de Haas PE, Hermans PW, Koedam MA, Teppema KS, et al. A novel pathogenic taxon of the *Mycobacterium tuberculosis* complex, Canetti: characterization of an exceptional isolate from Africa. Int J Syst Bacteriol. 1997;47:1236–45.

3. Koch R. Die Atiologie der Tuberkulose. Berl Klin Wochenschr. 1882;19:221–30.

4. Castets M, Rist N, Boisvert H. La variété africaine du bacille tuberculeux humain. Médecine Afr Noire. 1969;321–2.

5. Alexander KA, Laver PN, Michel AL, Williams M, van Helden PD, Warren RM, et al. Novel *Mycobacterium tuberculosis* Complex Pathogen, M. mungi. Emerg Infect Dis. 2010;16:1296–9.

6. Wagner JC, Buchanan G, Bokkenheuser V, Leviseur S. An acid-fast bacillus isolated from the lungs of the Cape hyrax, Procavia capensis (Pallas). Nature. 1958;181:284–5.

7. Parsons SDC, Drewe JA, Gey van Pittius NC, Warren RM, van Helden PD. Novel Cause of Tuberculosis in Meerkats, South Africa. Emerg Infect Dis. 2013;19:2004–7.

8. Lomme JR, Thoen CO, Himes EM, Vinson JW, King RE. *Mycobacterium tuberculosis* infection in two East African oryxes. J Am Vet Med Assoc. 1976;169:912–4.

9. Reed GB. Genus *Mycobacterium* (species affectin warm-blooded animal except those causing leprosy). Bergeys Man Determinative Bacteriol. Baltimore: The Williams and Wilkins Co; 1957.

10. Cousins DV, Bastida R, Cataldi A, Quse V, Redrobe S, Dow S, et al. Tuberculosis in seals caused by a novel member of the *Mycobacterium tuberculosis* complex: *Mycobacterium pinnipedii* sp. nov. Int J Syst Evol Microbiol. 2003;53:1305–14.

11. Aranaz A, Cousins D, Mateos A, Domínguez L. Elevation of *Mycobacterium tuberculosis* subsp. *caprae* Aranaz et al. 1999 to species rank as *Mycobacterium caprae* comb. nov., sp. nov. Int J Syst Evol Microbiol. 2003;53:1785–9.

12. Karlson AG, Lessel EF. *Mycobacterium bovis* nom. nov. Int J Syst Evol Microbiol. 1970;20:273–82.

13. Calmette A. La vaccination préventive contre la tuberculose par le "BCG." Paris: Masson et cie; 1927.

14. Brosch R, Gordon SV, Pym A, Eiglmeier K, Garnier T, Cole ST. Comparative genomics of the mycobacteria. Int J Med Microbiol IJMM. 2000;290:143–52.

15. Böddinghaus B, Rogall T, Flohr T, Blöcker H, Böttger EC. Detection and identification of mycobacteria by amplification of rRNA. J Clin Microbiol. 1990;28:1751–9.

16. Sreevatsan S, Pan X, Stockbauer KE, Connell ND, Kreiswirth BN, Whittam TS, et al. Restricted structural

gene polymorphism in the *Mycobacterium tuberculosis* complex indicates evolutionarily recent global dissemination. Proc Natl Acad Sci. 1997;94:9869–74.

17. Hirsh AE, Tsolaki AG, DeRiemer K, Feldman MW, Small PM. Stable association between strains of *Mycobacterium tuberculosis* and their human host populations. Proc Natl Acad Sci U S A. 2004;101:4871–6.

18. Rastogi N, Legrand E, Sola C. The mycobacteria: an introduction to nomenclature and pathogenesis. Rev Sci Tech Int Off Epizoot. 2001;20:21–54.

19. Ford C, Yusim K, Ioerger T, Feng S, Chase M, Greene M, et al. *Mycobacterium tuberculosis*--heterogeneity revealed through whole genome sequencing. Tuberc Edinb Scotl. 2012;92:194–201.

20. Otal I, Martín C, Vincent-Lévy-Frebault V, Thierry D, Gicquel B. Restriction fragment length polymorphism analysis using IS6110 as an epidemiological marker in tuberculosis. J Clin Microbiol. 1991;29:1252–4.

21. Gordon SV, Brosch R, Billault A, Garnier T, Eiglmeier K, Cole ST. Identification of variable regions in the genomes of tubercle bacilli using bacterial artificial chromosome arrays. Mol Microbiol. 1999;32:643–55.

22. Supply P, Mazars E, Lesjean S, Vincent V, Gicquel B, Locht C. Variable human minisatellite-like regions in the *Mycobacterium tuberculosis* genome. Mol Microbiol. 2000;36:762–71.

23. Kamerbeek J, Schouls L, Kolk A, van Agterveld M, van Soolingen D, Kuijper S, et al. Simultaneous detection and strain differentiation of *Mycobacterium tuberculosis* for diagnosis and epidemiology. J Clin Microbiol. 1997;35:907–14.

24. Supply P, Allix C, Lesjean S, Cardoso-Oelemann M, Rüsch-Gerdes S, Willery E, et al. Proposal for Standardization of Optimized Mycobacterial Interspersed Repetitive Unit-Variable-Number Tandem Repeat Typing of *Mycobacterium tuberculosis*. J Clin Microbiol. 2006;44:4498–510.

25. Joshi D, Harris NB, Waters R, Thacker T, Mathema B, Krieswirth B, et al. Single Nucleotide Polymorphisms in the *Mycobacterium bovis* Genome Resolve Phylogenetic Relationships. J Clin Microbiol. 2012;50:3853–61.

26. Telenti A, Imboden P, Marchesi F, Lowrie D, Cole S, Colston MJ, et al. Detection of rifampicin-resistance mutations in *Mycobacterium tuberculosis*. Lancet Lond Engl. 1993;341:647–50.

27. Roring S, Brittain D, Bunschoten AE, Hughes MS, Skuce RA, van Embden JD, et al. Spacer oligotyping of *Mycobacterium bovis* isolates compared to typing by restriction fragment length polymorphism using PGRS, DR and IS6110 probes. Vet Microbiol. 1998;61:111–20.

28. Ross BC, Raios K, Jackson K, Dwyer B. Molecular cloning of a highly repeated DNA element from *Mycobacterium tuberculosis* and its use as an epidemiological tool. J Clin Microbiol. 1992;30:942–6.

29. Hermans PW, van Soolingen D, Bik EM, de Haas PE, Dale JW, van Embden JD. Insertion element IS987 from *Mycobacterium bovis* BCG is located in a hot-spot integration region for insertion elements in

Mycobacterium tuberculosis complex strains. Infect Immun. 1991;59:2695–705.

30. Groenen PM, Bunschoten AE, van Soolingen D, van Embden JD. Nature of DNA polymorphism in the direct repeat cluster of *Mycobacterium tuberculosis*; application for strain differentiation by a novel typing method. Mol Microbiol. 1993;10:1057–65.

31. McEvoy CRE, Falmer AA, Gey van Pittius NC, Victor TC, van Helden PD, Warren RM. The role of IS6110 in the evolution of *Mycobacterium tuberculosis*. Tuberc Edinb Scotl. 2007;87:393–404.

32. Jagielski T, van Ingen J, Rastogi N, Dziadek J, aw, Mazur P, et al. Current Methods in the Molecular Typing of *Mycobacterium tuberculosis* and Other Mycobacteria. BioMed Res Int. 2014;2014:e645802.

33. Schürch AC, van Soolingen D. DNA fingerprinting of *Mycobacterium tuberculosis*: from phage typing to whole-genome sequencing. Infect Genet Evol J Mol Epidemiol Evol Genet Infect Dis. 2012;12:602–9.

34. Horvath P, Barrangou R. CRISPR/Cas, the immune system of bacteria and archaea. Science. 2010;327:167–70.

35. Kamerbeek J, Schouls L, Kolk A, van Agterveld M, van Soolingen D, Kuijper S, et al. Simultaneous detection and strain differentiation of *Mycobacterium tuberculosis* for diagnosis and epidemiology. J Clin Microbiol. 1997;35:907–14.

36. Cohen T, Wilson D, Wallengren K, Samuel EY, Murray M. Mixed-Strain *Mycobacterium tuberculosis* Infections among Patients Dying in a Hospital in KwaZulu-Natal, South Africa. J Clin Microbiol. 2011;49:385–8.

37. Kato-Maeda M, Metcalfe JZ, Flores L. Genotyping of *Mycobacterium tuberculosis*: application in epidemiologic studies. Future Microbiol. 2011;6:203–16.

38. Gormley E, Corner L a. L, Costello E, Rodriguez-Campos S. Bacteriological diagnosis and molecular strain typing of *Mycobacterium bovis* and *Mycobacterium caprae*. Res Vet Sci. 2014;97 Suppl:S30-43.

39. Comas I, Homolka S, Niemann S, Gagneux S. Genotyping of genetically monomorphic bacteria: DNA sequencing in *Mycobacterium tuberculosis* highlights the limitations of current methodologies. PloS One. 2009;4:e7815.

40. Warren RM, Streicher EM, Sampson SL, van der Spuy GD, Richardson M, Nguyen D, et al. Microevolution of the direct repeat region of *Mycobacterium tuberculosis*: implications for interpretation of spoligotyping data. J Clin Microbiol. 2002;40:4457–65.

41. Filliol I, Motiwala AS, Cavatore M, Qi W, Hazbón MH, Bobadilla del Valle M, et al. Global phylogeny of *Mycobacterium tuberculosis* based on single nucleotide polymorphism (SNP) analysis: insights into tuberculosis evolution, phylogenetic accuracy of other DNA fingerprinting systems, and recommendations for a minimal standard SNP set. J Bacteriol. 2006;188:759–72.

42. Costello E, O'Grady D, Flynn O, O'Brien R, Rogers M, Quigley F, et al. Study of Restriction Fragment

Length Polymorphism Analysis and Spoligotyping for Epidemiological Investigation of *Mycobacterium bovis* Infection. J Clin Microbiol. 1999;37:3217–22.

43. Nakamura Y, Leppert M, O'Connell P, Wolff R, Holm T, Culver M, et al. Variable number of tandem repeat (VNTR) markers for human gene mapping. Science. 1987;235:1616–22.

44. Roring S, Scott A, Brittain D, Walker I, Hewinson G, Neill S, et al. Development of variable-number tandem repeat typing of *Mycobacterium bovis*: comparison of results with those obtained by using existing exact tandem repeats and spoligotyping. J Clin Microbiol. 2002;40:2126–33.

45. Frothingham R, Meeker-O'Connell WA. Genetic diversity in the *Mycobacterium tuberculosis* complex based on variable numbers of tandem DNA repeats. Microbiol Read Engl. 1998;144 (Pt 5):1189–96.

46. Cole ST, Brosch R, Parkhill J, Garnier T, Churcher C, Harris D, et al. Deciphering the biology of *Mycobacterium tuberculosis* from the complete genome sequence. Nature. 1998;393:537–44.

47. Supply P, Magdalena J, Himpens S, Locht C. Identification of novel intergenic repetitive units in a mycobacterial two-component system operon. Mol Microbiol. 1997;26:991–1003.

48. García de Viedma D, de Viedma DG, Alonso Rodríguez N, Rodríguez NA, Andrés S, Martínez Lirola M, et al. Evaluation of alternatives to RFLP for the analysis of clustered cases of tuberculosis. Int J Tuberc Lung Dis Off J Int Union Tuberc Lung Dis. 2006;10:454–9.

49. Bouklata N, Supply P, Jaouhari S, Charof R, Seghrouchni F, Sadki K, et al. Molecular Typing of *Mycobacterium tuberculosis* Complex by 24-Locus Based MIRU-VNTR Typing in Conjunction with Spoligotyping to Assess Genetic Diversity of Strains Circulating in Morocco. PLOS ONE. 2015;10:e0135695.

50. Bidovec-Stojkovic U, Zolnir-Dovc M, Supply P. One year nationwide evaluation of 24-locus MIRU-VNTR genotyping on Slovenian *Mycobacterium tuberculosis* isolates. Respir Med. 2011;105 Suppl 1:S67-73.

51. Christianson S, Wolfe J, Orr P, Karlowsky J, Levett PN, Horsman GB, et al. Evaluation of 24 locus MIRU-VNTR genotyping of *Mycobacterium tuberculosis* isolates in Canada. Tuberc Edinb Scotl. 2010;90:31–8.

52. Vadwai V, Shetty A, Supply P, Rodrigues C. Evaluation of 24-locus MIRU-VNTR in extrapulmonary specimens: Study from a tertiary centre in Mumbai. Tuberculosis. 2012;92:264–72.

53. Smith NH, Kremer K, Inwald J, Dale J, Driscoll JR, Gordon SV, et al. Ecotypes of the *Mycobacterium tuberculosis* complex. J Theor Biol. 2006;239:220–5.

54. Aranaz A, Juan L de, Montero N, Sánchez C, Galka M, Delso C, et al. Bovine Tuberculosis (*Mycobacterium bovis*) in Wildlife in Spain. J Clin Microbiol. 2004;42:2602–8.

55. Romero B, Aranaz A, Sandoval Á, Álvarez J, de Juan L, Bezos J, et al. Persistence and molecular evolution of *Mycobacterium bovis* population from cattle and wildlife in Doñana National Park revealed by genotype variation. Vet Microbiol. 2008;132:87–95.

56. Al LB et. Silent Nucleotide Polymorphisms and a Phylogeny for *Mycobacterium tuberculosis* - Volume 10,

Number 9—September 2004 - Emerging Infectious Disease journal - CDC. [cited 2017 Oct 17]; Available from: https://wwwnc.cdc.gov/eid/article/10/9/04-0046_article

57. Gutacker MM, Mathema B, Soini H, Shashkina E, Kreiswirth BN, Graviss EA, et al. Single-nucleotide polymorphism-based population genetic analysis of *Mycobacterium tuberculosis* strains from 4 geographic sites. J Infect Dis. 2006;193:121–8.

58. Hunt R, Sauna ZE, Ambudkar SV, Gottesman MM, Kimchi-Sarfaty C. Silent (synonymous) SNPs: should we care about them? Methods Mol Biol Clifton NJ. 2009;578:23–39.

59. Achtman M. Evolution, Population Structure, and Phylogeography of Genetically Monomorphic Bacterial Pathogens. Annu Rev Microbiol. 2008;62:53–70.

60. Black WC, Vontas JG. Affordable assays for genotyping single nucleotide polymorphisms in insects. Insect Mol Biol. 2007;16:377–87.

61. Mestre O, Luo T, Dos Vultos T, Kremer K, Murray A, Namouchi A, et al. Phylogeny of *Mycobacterium tuberculosis* Beijing strains constructed from polymorphisms in genes involved in DNA replication, recombination and repair. PloS One. 2011;6:e16020.

62. Espinosa de los Monteros LE, Galán JC, Gutiérrez M, Samper S, García Marín JF, Martín C, et al. Allele-specific PCR method based on pncA and oxyR sequences for distinguishing *Mycobacterium bovis* from *Mycobacterium tuberculosis*: intraspecific M. bovis pncA sequence polymorphism. J Clin Microbiol. 1998;36:239–42.

63. Halse TA, Escuyer VE, Musser KA. Evaluation of a single-tube multiplex real-time PCR for differentiation of members of the *Mycobacterium tuberculosis* complex in clinical specimens. J Clin Microbiol. 2011;49:2562–7.

64. Bouakaze C, Keyser C, de Martino SJ, Sougakoff W, Veziris N, Dabernat H, et al. Identification and genotyping of *Mycobacterium tuberculosis* complex species by use of a SNaPshot Minisequencing-based assay. J Clin Microbiol. 2010;48:1758–66.

65. Bouakaze C, Keyser C, Gonzalez A, Sougakoff W, Veziris N, Dabernat H, et al. Matrix-assisted laser desorption ionization-time of flight mass spectrometry-based single nucleotide polymorphism genotyping assay using iPLEX gold technology for identification of *Mycobacterium tuberculosis* complex species and lineages. J Clin Microbiol. 2011;49:3292–9.

66. Bergval IL, Vijzelaar RNCP, Dalla Costa ER, Schuitema ARJ, Oskam L, Kritski AL, et al. Development of multiplex assay for rapid characterization of *Mycobacterium tuberculosis*. J Clin Microbiol. 2008;46:689–99.

67. Stucki D, Malla B, Hostettler S, Huna T, Feldmann J, Yeboah-Manu D, et al. Two new rapid SNP-typing methods for classifying *Mycobacterium tuberculosis* complex into the main phylogenetic lineages. PloS One. 2012;7:e41253.

68. Coll F, McNerney R, Guerra-Assunção JA, Glynn JR, Perdigão J, Viveiros M, et al. A robust SNP barcode for typing *Mycobacterium tuberculosis* complex strains. Nat Commun. 2014;5:4812.

69. Homolka S, Projahn M, Feuerriegel S, Ubben T, Diel R, Nübel U, et al. High resolution discrimination of clinical *Mycobacterium tuberculosis* complex strains based on single nucleotide polymorphisms. PloS One. 2012;7:e39855.

70. Kapur V, Whittam TS, Musser JM. Is *Mycobacterium tuberculosis* 15,000 years old? J Infect Dis. 1994;170:1348–9.

71. Garnier T, Eiglmeier K, Camus J-C, Medina N, Mansoor H, Pryor M, et al. The complete genome sequence of *Mycobacterium bovis*. Proc Natl Acad Sci U S A. 2003;100:7877–82.

72. Brosch R, Philipp WJ, Stavropoulos E, Colston MJ, Cole ST, Gordon SV. Genomic analysis reveals variation between *Mycobacterium tuberculosis* H37Rv and the attenuated M. tuberculosis H37Ra strain. Infect Immun. 1999;67:5768–74.

73. Behr MA, Wilson MA, Gill WP, Salamon H, Schoolnik GK, Rane S, et al. Comparative genomics of BCG vaccines by whole-genome DNA microarray. Science. 1999;284:1520–3.

74. Comas I, Coscolla M, Luo T, Borrell S, Holt KE, Kato-Maeda M, et al. Out-of-Africa migration and Neolithic coexpansion of *Mycobacterium tuberculosis* with modern humans. Nat Genet. 2013;45:1176–82.

75. Hershberg R, Lipatov M, Small PM, Sheffer H, Niemann S, Homolka S, et al. High functional diversity in *Mycobacterium tuberculosis* driven by genetic drift and human demography. PLoS Biol. 2008;6:e311.

76. Brosch R, Gordon SV, Marmiesse M, Brodin P, Buchrieser C, Eiglmeier K, et al. A new evolutionary scenario for the *Mycobacterium tuberculosis* complex. Proc Natl Acad Sci U S A. 2002;99:3684–9.

77. Mostowy S, Cousins D, Brinkman J, Aranaz A, Behr MA. Genomic deletions suggest a phylogeny for the *Mycobacterium tuberculosis* complex. J Infect Dis. 2002;186:74–80.

78. Durr PA, Clifton-Hadley RS, Hewinson RG. Molecular epidemiology of bovine tuberculosis. II. Applications of genotyping. Rev Sci Tech Int Off Epizoot. 2000;19:689–701.

79. Cohan FM. What are bacterial species? Annu Rev Microbiol. 2002;56:457–87.

80. Smith NH, Dale J, Inwald J, Palmer S, Gordon SV, Hewinson RG, et al. The population structure of *Mycobacterium bovis* in Great Britain: Clonal expansion. Proc Natl Acad Sci U S A. 2003;100:15271–5.

81. Smith NH, Upton P. Naming spoligotype patterns for the RD9-deleted lineage of the *Mycobacterium tuberculosis* complex; www.Mbovis.org. Infect Genet Evol J Mol Epidemiol Evol Genet Infect Dis. 2012;12:873–6.

82. Smith NH, Berg S, Dale J, Allen A, Rodriguez S, Romero B, et al. European 1: a globally important clonal complex of *Mycobacterium bovis*. Infect Genet Evol J Mol Epidemiol Evol Genet Infect Dis. 2011;11:1340–51.

83. Rodriguez-Campos S, Schürch AC, Dale J, Lohan AJ, Cunha MV, Botelho A, et al. European 2--a clonal complex of *Mycobacterium bovis* dominant in the Iberian Peninsula. Infect Genet Evol J Mol Epidemiol Evol Genet Infect Dis. 2012;12:866–72.

84. Müller B, Hilty M, Berg S, Garcia-Pelayo MC, Dale J, Boschiroli ML, et al. African 1, an epidemiologically important clonal complex of *Mycobacterium bovis* dominant in Mali, Nigeria, Cameroon, and Chad. J Bacteriol. 2009;191:1951–60.

85. Berg S, Garcia-Pelayo MC, Müller B, Hailu E, Asiimwe B, Kremer K, et al. African 2, a Clonal Complex of *Mycobacterium bovis* Epidemiologically Important in East Africa. J Bacteriol. 2011;193:670–8.

86. Michel AL, Müller B, van Helden PD. *Mycobacterium bovis* at the animal-human interface: a problem, or not? Vet Microbiol. 2010;140:371–81.

87. O'Reilly LM, Daborn CJ. The epidemiology of *Mycobacterium bovis* infections in animals and man: a review. Tuber Lung Dis Off J Int Union Tuberc Lung Dis. 1995;76 Suppl 1:1–46.

88. OIE World Animal Health Information System [Internet]. [cited 2017 Nov 7]. Available from: http://www.oie.int/wahis_2/public/wahid.php/Diseaseinformation/statusdetail

89. Wedlock DN, Skinner MA, de Lisle GW, Buddle BM. Control of *Mycobacterium bovis* infections and the risk to human populations. Microbes Infect. 2002;4:471–80.

90. Menzies FD, Neill SD. Cattle-to-cattle transmission of bovine tuberculosis. Vet J Lond Engl 1997. 2000;160:92–106.

91. van Crevel R, Ottenhoff THM, van der Meer JWM. Innate Immunity to *Mycobacterium tuberculosis*. Clin Microbiol Rev. 2002;15:294–309.

92. Cambier CJ, Falkow S, Ramakrishnan L. Host evasion and exploitation schemes of *Mycobacterium tuberculosis*. Cell. 2014;159:1497–509.

93. Alix E, Mukherjee S, Roy CR. Subversion of membrane transport pathways by vacuolar pathogens. J Cell Biol. 2011;195:943–52.

94. Pollock JM, Rodgers JD, Welsh MD, McNair J. Pathogenesis of bovine tuberculosis: the role of experimental models of infection. Vet Microbiol. 2006;112:141–50.

95. Ramírez-Villaescusa AM, Medley GF, Mason S, Green LE. Risk factors for herd breakdown with bovine tuberculosis in 148 cattle herds in the south west of England. Prev Vet Med. 2010;95:224–30.

96. OIE Terrestrial Manual. (2009). Bovine Tuberculosis. World assembly of delegates of the OIE in May 2009, Chapter 2.4.7.

97. Huebner RE, Schein MF, Bass JB. The tuberculin skin test. Clin Infect Dis Off Publ Infect Dis Soc Am. 1993;17:968–75.

98. Home - SRA - NCBI [Internet]. [cited 2017 Nov 9]. Available from: https://www.ncbi.nlm.nih.gov/sra

99. Patel RK, Jain M. NGS QC Toolkit: A Toolkit for Quality Control of Next Generation Sequencing Data. PLoS ONE [Internet]. 2012 [cited 2016 Dec 7];7. Available from: http://www.ncbi.nlm.nih.gov/pmc/articles/PMC3270013/

100. Lander ES, Waterman MS. Genomic mapping by fingerprinting random clones: a mathematical analysis. Genomics. 1988;2:231–9.

101. Li H, Durbin R. Fast and accurate short read alignment with Burrows-Wheeler transform. Bioinforma Oxf Engl. 2009;25:1754–60.

102. Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, et al. The Sequence Alignment/Map format and SAMtools. Bioinforma Oxf Engl. 2009;25:2078–9.

103. Koboldt DC, Zhang Q, Larson DE, Shen D, McLellan MD, Lin L, et al. VarScan 2: somatic mutation and copy number alteration discovery in cancer by exome sequencing. Genome Res. 2012;22:568–76.

104. Cingolani P, Platts A, Wang LL, Coon M, Nguyen T, Wang L, et al. A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff. Fly (Austin). 2012;6:80–92.

105. Xia E, Teo Y-Y, Ong RT-H. SpoTyping: fast and accurate in silico *Mycobacterium* spoligotyping from sequence reads. Genome Med [Internet]. 2016 [cited 2016 Oct 25];8. Available from: http://www.ncbi.nlm.nih.gov/pmc/articles/PMC4756441/

106. Coll F, Mallard K, Preston MD, Bentley S, Parkhill J, McNerney R, et al. SpolPred: rapid and accurate prediction of *Mycobacterium tuberculosis* spoligotypes from short genomic sequences. Bioinforma Oxf Engl. 2012;28:2991–3.

107. Stamatakis A. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. Bioinformatics. 2014;30:1312–3.

108. Huerta-Cepas J, Dopazo J, Gabaldón T. ETE: a python Environment for Tree Exploration. BMC Bioinformatics. 2010;11:24.

109. Wong VK, Baker S, Connor TR, Pickard D, Page AJ, Dave J, et al. An extended genotyping framework for *Salmonella enterica* serovar Typhi, the cause of human typhoid. Nat Commun. 2016;7:ncomms12827.

110. Corander J, Marttinen P, Sirén J, Tang J. Enhanced Bayesian modelling in BAPS software for learning genetic structures of populations. BMC Bioinformatics. 2008;9:539.

111. Phelan JE, Coll F, Bergval I, Anthony RM, Warren R, Sampson SL, et al. Recombination in pe/ppe genes contributes to genetic variation in *Mycobacterium tuberculosis* lineages. BMC Genomics. 2016;17:151.

112. Liu X, Gutacker MM, Musser JM, Fu Y-X. Evidence for Recombination in *Mycobacterium tuberculosis*. J Bacteriol. 2006;188:8169–77.

113. Namouchi A, Didelot X, Schöck U, Gicquel B, Rocha EPC. After the bottleneck: Genome-wide

diversification of the *Mycobacterium tuberculosis* complex by mutation, recombination, and natural selection. Genome Res. 2012;22:721–34.

114. Cheng L, Connor TR, Sirén J, Aanensen DM, Corander J. Hierarchical and Spatially Explicit Clustering of DNA Sequences with BAPS Software. Mol Biol Evol. 2013;30:1224–8.

115. Smith NH. The global distribution and phylogeography of *Mycobacterium bovis* clonal complexes. Infect Genet Evol J Mol Epidemiol Evol Genet Infect Dis. 2012;12:857–65.

116. Cole ST. Comparative and functional genomics of the *Mycobacterium tuberculosis* complex. Microbiol Read Engl. 2002;148:2919–28.

117. Zimpel CK, Brandão PE, Filho de S, F A, Souza D, F R, et al. Complete Genome Sequencing of *Mycobacterium bovis* SP38 and Comparative Genomics of *Mycobacterium bovis* and *M. tuberculosis* Strains. Front Microbiol [Internet]. 2017 [cited 2017 Dec 17];8. Available from: https://www.frontiersin.org/articles/10.3389/fmicb.2017.02389/full#B67

118. Yahyaoui-Azami H, Aboukhassib H, Bouslikhane M, Berrada J, Rami S, Reinhard M, et al. Molecular characterization of bovine tuberculosis strains in two slaughterhouses in Morocco. BMC Vet Res [Internet]. 2017 [cited 2017 Dec 17];13. Available from: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5574129/

119. El-Sayed A, El-Shannat S, Kamel M, Castañeda-Vazquez MA, Castañeda-Vazquez H. Molecular Epidemiology of *Mycobacterium bovis* in Humans and Cattle. Zoonoses Public Health. 2016;63:251–64.

120. Alland D, Lacher DW, Hazbón MH, Motiwala AS, Qi W, Fleischmann RD, et al. Role of Large Sequence Polymorphisms (LSPs) in Generating Genomic Diversity among Clinical Isolates of *Mycobacterium tuberculosis* and the Utility of LSPs in Phylogenetic Analysis. J Clin Microbiol. 2007;45:39–46.

121. Faksri K, Xia E, Tan JH, Teo Y-Y, Ong RT-H. In silico region of difference (RD) analysis of *Mycobacterium tuberculosis* complex from sequence reads using RD-Analyzer. BMC Genomics [Internet]. 2016 [cited 2018 Mar 3];17. Available from: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5093977/

ANEXO 1 – Características de las cepas utilizadas

			BAPS Clusters 3 levels, k=10					
Strain (SRA accession)	**Country**	**Spoligotype**	**Continent**	**Cluster3_10_1**	**Cluster3_10_2**	**Cluster3_10_3**	**Sub lineage**	**Obs.**
SRR1792164	China	SB1622	Asia	Cluster 4	Cluster2 14	Cluster3 40	XX	Outgroup
SRR1791710	China	SB2467	Asia	Cluster 4	Cluster2 16	Cluster3 42	1.1	
SRR1791712	China	SB2467	Asia	Cluster 4	Cluster2 16	Cluster3 42	1.1	
SRR1791768	Canada	SB0265	North America	Cluster 4	Cluster2 16	Cluster3 42	1.1	
SRR1792161	USA	SB1069	North America	Cluster 4	Cluster2 16	Cluster3 42	1.1	
SRR1792237	USA	Not_found	North America	Cluster 4	Cluster2 16	Cluster3 42	1.1	
SRR1792472	USA	SB0265	North America	Cluster 4	Cluster2 16	Cluster3 42	1.1	
SRR1792474	USA	SB1069	North America	Cluster 4	Cluster2 16	Cluster3 42	1.1	
SRR1792479	USA	SB1069	North America	Cluster 4	Cluster2 16	Cluster3 42	1.1	
SRR1792481	USA	SB1069	North America	Cluster 4	Cluster2 16	Cluster3 42	1.1	
SRR1792482	USA	Not_found	North America	Cluster 4	Cluster2 16	Cluster3 43	1.1	
SRR1792485	USA	SB0120	North America	Cluster 4	Cluster2 16	Cluster3 43	1.1	
SRR1792488	USA	SB1069	North America	Cluster 4	Cluster2 16	Cluster3 42	1.1	
SRR1792493	USA	SB1069	North America	Cluster 4	Cluster2 16	Cluster3 42	1.1	
SRR1792498	USA	SB0265	North America	Cluster 4	Cluster2 16	Cluster3 42	1.1	
SRR4117155	France	SB0120	Europe	Cluster 4	Cluster2 16	Cluster3 42	1.1	
SRR4199188	France	SB0120	Europe	Cluster 4	Cluster2 16	Cluster3 42	1.1	
SRR1791714	UNKNOWN	SB0121	UNKNOWN	Cluster 4	Cluster2 12	Cluster3 34	1.2	
SRR1791722	Mexico	SB0121	North America	Cluster 4	Cluster2 12	Cluster3 34	1.2	
SRR1791736	USA	SB0121	North America	Cluster 4	Cluster2 12	Cluster3 32	1.2	
SRR1791793	USA	SB0121	North America	Cluster 4	Cluster2 12	Cluster3 33	1.2	
SRR1791797	Mexico	SB0121	North America	Cluster 4	Cluster2 12	Cluster3 34	1.2	
SRR1791802	Mexico	SB0121	North America	Cluster 4	Cluster2 12	Cluster3 32	1.2	
SRR1791824	Mexico	SB1308	North America	Cluster 4	Cluster2 12	Cluster3 34	1.2	
SRR1791830	USA	SB0121	North America	Cluster 4	Cluster2 12	Cluster3 33	1.2	
SRR1791841	USA	SB0121	North America	Cluster 4	Cluster2 12	Cluster3 33	1.2	
SRR1791858	UNKNOWN	SB0121	UNKNOWN	Cluster 4	Cluster2 12	Cluster3 34	1.2	
SRR1791860	USA	SB1345	North America	Cluster 4	Cluster2 12	Cluster3 36	1.2	
SRR1791867	USA	SB1345	North America	Cluster 4	Cluster2 12	Cluster3 36	1.2	
SRR1791868	USA	SB1345	North America	Cluster 4	Cluster2 12	Cluster3 36	1.2	
SRR1791869	USA	SB1345	North America	Cluster 4	Cluster2 12	Cluster3 36	1.2	
SRR1791875	USA	SB1345	North America	Cluster 4	Cluster2 12	Cluster3 36	1.2	
SRR1791876	USA	SB1345	North America	Cluster 4	Cluster2 12	Cluster3 36	1.2	
SRR1791880	USA	SB1345	North America	Cluster 4	Cluster2 12	Cluster3 36	1.2	
SRR1791891	Mexico	SB0121	North America	Cluster 4	Cluster2 12	Cluster3 33	1.2	
SRR1791946	Mexico	SB0121	North America	Cluster 4	Cluster2 12	Cluster3 32	1.2	
SRR1791950	USA	SB0121	North America	Cluster 4	Cluster2 12	Cluster3 32	1.2	
SRR1791959	Mexico	SB0121	North America	Cluster 4	Cluster2 12	Cluster3 34	1.2	
SRR1791969	Mexico	SB0121	North America	Cluster 4	Cluster2 12	Cluster3 34	1.2	
SRR1792005	Mexico	SB0121	North America	Cluster 4	Cluster2 12	Cluster3 34	1.2	
SRR1792027	Mexico	SB0121	North America	Cluster 4	Cluster2 12	Cluster3 34	1.2	
SRR1792041	Mexico	SB0121	North America	Cluster 4	Cluster2 12	Cluster3 34	1.2	
SRR1792108	USA	SB0121	North America	Cluster 4	Cluster2 12	Cluster3 34	1.2	
SRR1792109	USA	SB0121	North America	Cluster 4	Cluster2 12	Cluster3 34	1.2	
SRR1792110	USA	SB0121	North America	Cluster 4	Cluster2 12	Cluster3 34	1.2	
SRR1792111	USA	SB0121	North America	Cluster 4	Cluster2 12	Cluster3 34	1.2	

SRR1792118	Mexico	SB0121	North America	Cluster 4	Cluster2 12	Cluster3 34	1.2	
SRR1792122	USA	SB0121	North America	Cluster 4	Cluster2 12	Cluster3 34	1.2	
SRR1792147	USA	SB0121	North America	Cluster 4	Cluster2 12	Cluster3 34	1.2	
SRR1792155	Mexico	SB0121	North America	Cluster 4	Cluster2 12	Cluster3 32	1.2	
SRR1792171	Mexico	SB0121	North America	Cluster 4	Cluster2 12	Cluster3 34	1.2	
SRR1792193	USA	SB0121	North America	Cluster 4	Cluster2 12	Cluster3 34	1.2	
SRR1792203	USA	SB0121	North America	Cluster 4	Cluster2 12	Cluster3 34	1.2	
SRR1792204	USA	SB0121	North America	Cluster 4	Cluster2 12	Cluster3 34	1.2	
SRR1792245	Mexico	SB0121	North America	Cluster 4	Cluster2 12	Cluster3 32	1.2	
SRR1792261	USA	SB0121	North America	Cluster 4	Cluster2 12	Cluster3 34	1.2	
SRR1792409	Mexico	SB0121	North America	Cluster 4	Cluster2 12	Cluster3 34	1.2	
SRR1792410	Mexico	SB0121	North America	Cluster 4	Cluster2 12	Cluster3 34	1.2	
SRR1792445	Mexico	SB0121	North America	Cluster 4	Cluster2 12	Cluster3 32	1.2	
SRR1792449	Mexico	SB0121	North America	Cluster 4	Cluster2 12	Cluster3 33	1.2	
SRR1792477	USA	SB0121	North America	Cluster 4	Cluster2 12	Cluster3 34	1.2	
SRR1792491	USA	SB0121	North America	Cluster 4	Cluster2 12	Cluster3 34	1.2	
SRR1792492	UNKNOWN	SB0121	UNKNOWN	Cluster 4	Cluster2 12	Cluster3 34	1.2	
SRR1792502	USA	SB0121	North America	Cluster 4	Cluster2 12	Cluster3 34	1.2	
ERR1815540		SB0121	Africa	Cluster 4	Cluster2 12	Cluster3 35	1.2	
ERR1815541		SB0121	Africa	Cluster 4	Cluster2 12	Cluster3 35	1.2	
ERR1815542		SB0121	Africa	Cluster 4	Cluster2 12	Cluster3 35	1.2	
ERR1815545		SB0121	Africa	Cluster 4	Cluster2 12	Cluster3 35	1.2	
ERR1815546		SB0121	Africa	Cluster 4	Cluster2 12	Cluster3 35	1.2	
ERR1815547		SB0121	Africa	Cluster 4	Cluster2 12	Cluster3 35	1.2	
ERR1815549		SB0121	Africa	Cluster 4	Cluster2 12	Cluster3 35	1.2	
SRR1791692	USA	SB0265	North America	Cluster 4	Cluster2 13	Cluster3 37	1.3	
SRR1791708	USA	SB0265	North America	Cluster 4	Cluster2 13	Cluster3 39	1.3	
SRR1791832	USA	SB0265	North America	Cluster 4	Cluster2 13	Cluster3 38	1.3	
SRR1792002	USA	SB0265	North America	Cluster 4	Cluster2 13	Cluster3 39	1.3	
SRR1792184	USA	SB0265	North America	Cluster 4	Cluster2 13	Cluster3 37	1.3	
SRR1792189	USA	SB0265	North America	Cluster 4	Cluster2 13	Cluster3 38	1.3	
SRR1792190	USA	SB0265	North America	Cluster 4	Cluster2 13	Cluster3 37	1.3	
SRR1792200	USA	SB0265	North America	Cluster 4	Cluster2 13	Cluster3 37	1.3	
SRR1792201	USA	SB0265	North America	Cluster 4	Cluster2 13	Cluster3 37	1.3	
SRR1792206	USA	SB0265	North America	Cluster 4	Cluster2 13	Cluster3 37	1.3	
SRR1792208	USA	SB0265	North America	Cluster 4	Cluster2 13	Cluster3 37	1.3	
SRR1792209	USA	SB0265	North America	Cluster 4	Cluster2 13	Cluster3 38	1.3	
SRR1792210	USA	SB0265	North America	Cluster 4	Cluster2 13	Cluster3 37	1.3	
SRR1792212	USA	SB0265	North America	Cluster 4	Cluster2 13	Cluster3 37	1.3	
SRR1792213	USA	SB0265	North America	Cluster 4	Cluster2 13	Cluster3 37	1.3	
SRR1792214	USA	SB0265	North America	Cluster 4	Cluster2 13	Cluster3 37	1.3	
SRR1792215	USA	SB0265	North America	Cluster 4	Cluster2 13	Cluster3 37	1.3	
SRR1792216	USA	SB0265	North America	Cluster 4	Cluster2 13	Cluster3 37	1.3	
SRR1792217	USA	SB0265	North America	Cluster 4	Cluster2 13	Cluster3 37	1.3	
SRR1792218	USA	SB0265	North America	Cluster 4	Cluster2 13	Cluster3 37	1.3	
SRR1792219	USA	SB0265	North America	Cluster 4	Cluster2 13	Cluster3 37	1.3	
SRR1792220	USA	SB0265	North America	Cluster 4	Cluster2 13	Cluster3 37	1.3	
SRR1792221	USA	SB0265	North America	Cluster 4	Cluster2 13	Cluster3 37	1.3	
SRR1792222	USA	SB0265	North America	Cluster 4	Cluster2 13	Cluster3 37	1.3	
SRR1792223	USA	SB0265	North America	Cluster 4	Cluster2 13	Cluster3 37	1.3	
SRR1792224	USA	SB0265	North America	Cluster 4	Cluster2 13	Cluster3 37	1.3	

SRR1792225	USA	SB0265	North America	Cluster 4	Cluster2 13	Cluster3 37	1.3	
SRR1792226	USA	SB0265	North America	Cluster 4	Cluster2 13	Cluster3 37	1.3	
SRR1792227	USA	SB0265	North America	Cluster 4	Cluster2 13	Cluster3 37	1.3	
SRR1792228	USA	SB0265	North America	Cluster 4	Cluster2 13	Cluster3 37	1.3	
SRR1792229	USA	SB0265	North America	Cluster 4	Cluster2 13	Cluster3 37	1.3	
SRR1792230	USA	SB0265	North America	Cluster 4	Cluster2 13	Cluster3 37	1.3	
SRR1792232	USA	SB0265	North America	Cluster 4	Cluster2 13	Cluster3 37	1.3	
SRR1792233	USA	SB0265	North America	Cluster 4	Cluster2 13	Cluster3 37	1.3	
SRR1792234	USA	SB0265	North America	Cluster 4	Cluster2 13	Cluster3 37	1.3	
SRR1792235	USA	SB0265	North America	Cluster 4	Cluster2 13	Cluster3 37	1.3	
SRR1792236	USA	SB0265	North America	Cluster 4	Cluster2 13	Cluster3 37	1.3	
SRR1792238	USA	SB0265	North America	Cluster 4	Cluster2 13	Cluster3 37	1.3	
SRR1792239	USA	SB0265	North America	Cluster 4	Cluster2 13	Cluster3 37	1.3	
SRR1792246	USA	SB0265	North America	Cluster 4	Cluster2 13	Cluster3 37	1.3	
SRR1792264	UNKNOWN	SB0265	UNKNOWN	Cluster 4	Cluster2 13	Cluster3 38	1.3	
SRR1792284	USA	SB0265	North America	Cluster 4	Cluster2 13	Cluster3 37	1.3	
SRR1792285	USA	SB0265	North America	Cluster 4	Cluster2 13	Cluster3 37	1.3	
SRR1792292	USA	SB0265	North America	Cluster 4	Cluster2 13	Cluster3 37	1.3	
SRR1792293	USA	SB0265	North America	Cluster 4	Cluster2 13	Cluster3 37	1.3	
SRR1792295	USA	SB0265	North America	Cluster 4	Cluster2 13	Cluster3 37	1.3	
SRR1792444	USA	SB0265	North America	Cluster 4	Cluster2 13	Cluster3 37	1.3	
SRR1792473	USA	SB0265	North America	Cluster 4	Cluster2 13	Cluster3 38	1.3	
SRR1792475	USA	Not_found	North America	Cluster 4	Cluster2 13	Cluster3 37	1.3	
SRR1792476	USA	SB0265	North America	Cluster 4	Cluster2 13	Cluster3 37	1.3	
SRR1792478	USA	SB0265	North America	Cluster 4	Cluster2 13	Cluster3 38	1.3	
SRR1792480	USA	SB0265	North America	Cluster 4	Cluster2 13	Cluster3 37	1.3	
SRR1792484	USA	SB0265	North America	Cluster 4	Cluster2 13	Cluster3 37	1.3	
SRR1792496	USA	SB0265	North America	Cluster 4	Cluster2 13	Cluster3 38	1.3	
SRR1792499	USA	SB0265	North America	Cluster 4	Cluster2 13	Cluster3 39	1.3	
SRR1792500	USA	SB0265	North America	Cluster 4	Cluster2 13	Cluster3 38	1.3	
SRR1792505	USA	SB0265	North America	Cluster 4	Cluster2 13	Cluster3 37	1.3	
SRR1173284	Uganda	Not_found	Africa	Cluster 4	Cluster2 15	Cluster3 41	1.4	Not shown in trees
Tb103.mbURU-010	Uruguay	SB0130	Latin America	Cluster 3	Cluster2 11	Cluster3 28		
Tb73.mbURU-002	Uruguay	SB0130	Latin America	Cluster 3	Cluster2 11	Cluster3 28		
Tb87.mbURU-007	Uruguay	SB0130	Latin America	Cluster 3	Cluster2 11	Cluster3 28		
mbURU-017.mbURU-017	Uruguay	SB0130	Latin America	Cluster 3	Cluster2 11	Cluster3 28		
mbURU-018.mbURU-018	Uruguay	SB0130	Latin America	Cluster 3	Cluster2 11	Cluster3 28		
SRR1791770	Mexico	SB0130	North America	Cluster 3	Cluster2 11	Cluster3 31		
SRR1791777	Mexico	SB0943	North America	Cluster 3	Cluster2 11	Cluster3 31		
SRR1791784	USA	SB1071	North America	Cluster 3	Cluster2 11	Cluster3 30		
SRR1792279	Mexico	SB0267	North America	Cluster 3	Cluster2 11	Cluster3 30		
ERR1815544		SB0130	Africa	Cluster 3	Cluster2 11	Cluster3 31		
ERR1815548		SB0267	Africa	Cluster 3	Cluster2 11	Cluster3 30		
SRR5216691	New_Zealand	SB1031	Oceania	Cluster 3	Cluster2 11	Cluster3 29		
SRR5216731	New_Zealand	SB1031	Oceania	Cluster 3	Cluster2 11	Cluster3 29		
SRR5216754	New_Zealand	SB1031	Oceania	Cluster 3	Cluster2 11	Cluster3 29		
SRR5216768	New_Zealand	SB1031	Oceania	Cluster 3	Cluster2 11	Cluster3 29		
SRR5216781	New_Zealand	SB1031	Oceania	Cluster 3	Cluster2 11	Cluster3 29		
SRR5216797	New_Zealand	SB1031	Oceania	Cluster 3	Cluster2 11	Cluster3 29		
SRR5216814	New_Zealand	SB1031	Oceania	Cluster 3	Cluster2 11	Cluster3 29		

SRR5216883	New_Zealand	SB1031	Oceania	Cluster 3	Cluster2 11	Cluster3 29		
SRR5216914	New_Zealand	SB1031	Oceania	Cluster 3	Cluster2 11	Cluster3 29		
SRR5216929	New_Zealand	SB1031	Oceania	Cluster 3	Cluster2 11	Cluster3 29		
SRR5216934	New_Zealand	SB1031	Oceania	Cluster 3	Cluster2 11	Cluster3 29		
SRR5216936	New_Zealand	SB1031	Oceania	Cluster 3	Cluster2 11	Cluster3 29		
SRR5216947	New_Zealand	SB1031	Oceania	Cluster 3	Cluster2 11	Cluster3 29		
SRR5216949	New_Zealand	SB1031	Oceania	Cluster 3	Cluster2 11	Cluster3 29		
SRR5216961	New_Zealand	SB1031	Oceania	Cluster 3	Cluster2 11	Cluster3 29		
SRR5216967	New_Zealand	SB1031	Oceania	Cluster 3	Cluster2 11	Cluster3 29		
SRR1791801	Mexico	SB2020	North America	Cluster 3	Cluster2 10	Cluster3 27	2.1	
SRR1791816	Mexico	SB2020	North America	Cluster 3	Cluster2 10	Cluster3 27	2.1	
SRR1791823	Mexico	Not_found	North America	Cluster 3	Cluster2 10	Cluster3 25	2.1	
SRR1791878	Mexico	SB0327	North America	Cluster 3	Cluster2 10	Cluster3 25	2.1	
SRR1791901	USA	SB0327	North America	Cluster 3	Cluster2 10	Cluster3 25	2.1	
SRR1791926	Mexico	SB0327	North America	Cluster 3	Cluster2 10	Cluster3 25	2.1	
SRR1792001	Mexico	SB0327	North America	Cluster 3	Cluster2 10	Cluster3 25	2.1	
SRR1792042	USA	SB0327	North America	Cluster 3	Cluster2 10	Cluster3 25	2.1	
SRR1792045	USA	SB0327	North America	Cluster 3	Cluster2 10	Cluster3 25	2.1	
SRR1792046	Mexico	SB1216	North America	Cluster 3	Cluster2 10	Cluster3 25	2.1	
SRR1792064	Mexico	SB0327	North America	Cluster 3	Cluster2 10	Cluster3 25	2.1	
SRR1792065	Mexico	SB1216	North America	Cluster 3	Cluster2 10	Cluster3 25	2.1	
SRR1792120	UNKNOWN	SB0327	UNKNOWN	Cluster 3	Cluster2 10	Cluster3 26	2.1	
SRR1792162	Mexico	SB0327	North America	Cluster 3	Cluster2 10	Cluster3 25	2.1	
SRR1792254	USA	SB2011	North America	Cluster 3	Cluster2 10	Cluster3 25	2.1	
SRR1792259	UNKNOWN	SB0327	UNKNOWN	Cluster 3	Cluster2 10	Cluster3 25	2.1	
SRR1792260	USA	SB2011	North America	Cluster 3	Cluster2 10	Cluster3 25	2.1	
SRR1792268	USA	SB2011	North America	Cluster 3	Cluster2 10	Cluster3 25	2.1	
SRR1792270	USA	SB2011	North America	Cluster 3	Cluster2 10	Cluster3 25	2.1	
SRR1792273	USA	SB2011	North America	Cluster 3	Cluster2 10	Cluster3 25	2.1	
SRR1792274	USA	SB2011	North America	Cluster 3	Cluster2 10	Cluster3 25	2.1	
SRR1792275	USA	SB2011	North America	Cluster 3	Cluster2 10	Cluster3 25	2.1	
SRR1792276	USA	SB2011	North America	Cluster 3	Cluster2 10	Cluster3 25	2.1	
SRR1792278	USA	SB2011	North America	Cluster 3	Cluster2 10	Cluster3 25	2.1	
SRR1792308	UNKNOWN	SB1216	UNKNOWN	Cluster 3	Cluster2 10	Cluster3 25	2.1	
SRR1792310	Mexico	Not_found	North America	Cluster 3	Cluster2 10	Cluster3 26	2.1	
SRR1792329	Mexico	SB0327	North America	Cluster 3	Cluster2 10	Cluster3 25	2.1	
SRR1792330	Mexico	SB0327	North America	Cluster 3	Cluster2 10	Cluster3 25	2.1	
SRR1792343	Mexico	SB0327	North America	Cluster 3	Cluster2 10	Cluster3 25	2.1	
SRR1792346	UNKNOWN	SB0327	UNKNOWN	Cluster 3	Cluster2 10	Cluster3 25	2.1	
SRR1792347	Mexico	SB0327	North America	Cluster 3	Cluster2 10	Cluster3 25	2.1	
SRR1792363	Mexico	SB0327	North America	Cluster 3	Cluster2 10	Cluster3 25	2.1	
SRR1792438	Mexico	SB0130	North America	Cluster 3	Cluster2 10	Cluster3 27	2.1	
SRR1791871	Mexico	SB0130	North America	Cluster 3	Cluster2 8	Cluster3 20	2.2	
SRR1791911	USA	SB1758	North America	Cluster 3	Cluster2 8	Cluster3 20	2.2	
SRR1792175	USA	SB1758	North America	Cluster 3	Cluster2 8	Cluster3 20	2.2	
SRR1792487	USA	SB0130	North America	Cluster 3	Cluster2 8	Cluster3 20	2.2	
ERR1815539		SB0130	Africa	Cluster 3	Cluster2 8	Cluster3 20	2.2	
ERR1815550		SB0130	Africa	Cluster 3	Cluster2 8	Cluster3 20	2.2	
SRR5216690	New_Zealand	SB0130	Oceania	Cluster 3	Cluster2 8	Cluster3 20	2.2	
SRR5216694	New_Zealand	SB0130	Oceania	Cluster 3	Cluster2 8	Cluster3 20	2.2	
SRR5216696	New_Zealand	SB0130	Oceania	Cluster 3	Cluster2 8	Cluster3 21	2.2	

SRR5216701	New_Zealand	SB0130	Oceania	Cluster 3	Cluster2 8	Cluster3 20	2.2	
SRR5216710	New_Zealand	SB0130	Oceania	Cluster 3	Cluster2 8	Cluster3 20	2.2	
SRR5216717	New_Zealand	SB0130	Oceania	Cluster 3	Cluster2 8	Cluster3 21	2.2	
SRR5216726	New_Zealand	SB0130	Oceania	Cluster 3	Cluster2 8	Cluster3 21	2.2	
SRR5216728	New_Zealand	SB0130	Oceania	Cluster 3	Cluster2 8	Cluster3 20	2.2	
SRR5216729	New_Zealand	SB0130	Oceania	Cluster 3	Cluster2 8	Cluster3 22	2.2	
SRR5216730	New_Zealand	SB0130	Oceania	Cluster 3	Cluster2 8	Cluster3 21	2.2	
SRR5216734	New_Zealand	SB0130	Oceania	Cluster 3	Cluster2 8	Cluster3 20	2.2	
SRR5216735	New_Zealand	SB0130	Oceania	Cluster 3	Cluster2 8	Cluster3 20	2.2	
SRR5216737	New_Zealand	SB1446	Oceania	Cluster 3	Cluster2 8	Cluster3 21	2.2	
SRR5216745	New_Zealand	SB0130	Oceania	Cluster 3	Cluster2 8	Cluster3 20	2.2	
SRR5216746	New_Zealand	SB0130	Oceania	Cluster 3	Cluster2 8	Cluster3 20	2.2	
SRR5216749	New_Zealand	SB0130	Oceania	Cluster 3	Cluster2 8	Cluster3 21	2.2	
SRR5216753	New_Zealand	SB0130	Oceania	Cluster 3	Cluster2 8	Cluster3 22	2.2	
SRR5216756	New_Zealand	SB0130	Oceania	Cluster 3	Cluster2 8	Cluster3 20	2.2	
SRR5216757	New_Zealand	SB0130	Oceania	Cluster 3	Cluster2 8	Cluster3 20	2.2	
SRR5216765	New_Zealand	SB0130	Oceania	Cluster 3	Cluster2 8	Cluster3 20	2.2	
SRR5216766	New_Zealand	SB0130	Oceania	Cluster 3	Cluster2 8	Cluster3 20	2.2	
SRR5216770	New_Zealand	SB0130	Oceania	Cluster 3	Cluster2 8	Cluster3 22	2.2	
SRR5216775	New_Zealand	SB0130	Oceania	Cluster 3	Cluster2 8	Cluster3 20	2.2	
SRR5216776	New_Zealand	SB0130	Oceania	Cluster 3	Cluster2 8	Cluster3 20	2.2	
SRR5216777	New_Zealand	SB0130	Oceania	Cluster 3	Cluster2 8	Cluster3 20	2.2	
SRR5216778	New_Zealand	SB0130	Oceania	Cluster 3	Cluster2 8	Cluster3 20	2.2	
SRR5216780	New_Zealand	SB0130	Oceania	Cluster 3	Cluster2 8	Cluster3 20	2.2	
SRR5216784	New_Zealand	SB0130	Oceania	Cluster 3	Cluster2 8	Cluster3 20	2.2	
SRR5216785	New_Zealand	SB0130	Oceania	Cluster 3	Cluster2 8	Cluster3 20	2.2	
SRR5216789	New_Zealand	SB0130	Oceania	Cluster 3	Cluster2 8	Cluster3 20	2.2	
SRR5216792	New_Zealand	SB0130	Oceania	Cluster 3	Cluster2 8	Cluster3 20	2.2	
SRR5216800	New_Zealand	SB0130	Oceania	Cluster 3	Cluster2 8	Cluster3 20	2.2	
SRR5216803	New_Zealand	SB0130	Oceania	Cluster 3	Cluster2 8	Cluster3 21	2.2	
SRR5216804	New_Zealand	Not_found	Oceania	Cluster 3	Cluster2 8	Cluster3 22	2.2	
SRR5216806	New_Zealand	SB0130	Oceania	Cluster 3	Cluster2 8	Cluster3 20	2.2	
SRR5216810	New_Zealand	SB0130	Oceania	Cluster 3	Cluster2 8	Cluster3 20	2.2	
SRR5216812	New_Zealand	SB0130	Oceania	Cluster 3	Cluster2 8	Cluster3 22	2.2	
SRR5216813	New_Zealand	SB0130	Oceania	Cluster 3	Cluster2 8	Cluster3 21	2.2	
SRR5216815	New_Zealand	SB0130	Oceania	Cluster 3	Cluster2 8	Cluster3 21	2.2	
SRR5216816	New_Zealand	SB0130	Oceania	Cluster 3	Cluster2 8	Cluster3 22	2.2	
SRR5216819	New_Zealand	SB0130	Oceania	Cluster 3	Cluster2 8	Cluster3 20	2.2	
SRR5216822	New_Zealand	SB0130	Oceania	Cluster 3	Cluster2 8	Cluster3 20	2.2	
SRR5216823	New_Zealand	SB0130	Oceania	Cluster 3	Cluster2 8	Cluster3 22	2.2	
SRR5216843	New_Zealand	SB0130	Oceania	Cluster 3	Cluster2 8	Cluster3 21	2.2	
SRR5216856	New_Zealand	SB0130	Oceania	Cluster 3	Cluster2 8	Cluster3 20	2.2	
SRR5216857	New_Zealand	SB0130	Oceania	Cluster 3	Cluster2 8	Cluster3 22	2.2	
SRR5216858	New_Zealand	SB0130	Oceania	Cluster 3	Cluster2 8	Cluster3 22	2.2	
SRR5216860	New_Zealand	SB0130	Oceania	Cluster 3	Cluster2 8	Cluster3 20	2.2	
SRR5216866	New_Zealand	SB0130	Oceania	Cluster 3	Cluster2 8	Cluster3 20	2.2	
SRR5216867	New_Zealand	SB0130	Oceania	Cluster 3	Cluster2 8	Cluster3 20	2.2	
SRR5216868	New_Zealand	SB0130	Oceania	Cluster 3	Cluster2 8	Cluster3 21	2.2	
SRR5216869	New_Zealand	SB0130	Oceania	Cluster 3	Cluster2 8	Cluster3 22	2.2	
SRR5216870	New_Zealand	SB0130	Oceania	Cluster 3	Cluster2 8	Cluster3 20	2.2	
SRR5216872	New_Zealand	Not_found	Oceania	Cluster 3	Cluster2 8	Cluster3 22	2.2	

SRR5216873	New_Zealand	SB0130	Oceania	Cluster 3	Cluster2 8	Cluster3 21	2.2	
SRR5216874	New_Zealand	SB0130	Oceania	Cluster 3	Cluster2 8	Cluster3 20	2.2	
SRR5216877	New_Zealand	SB0130	Oceania	Cluster 3	Cluster2 8	Cluster3 20	2.2	
SRR5216879	New_Zealand	SB0130	Oceania	Cluster 3	Cluster2 8	Cluster3 22	2.2	
SRR5216888	New_Zealand	SB0130	Oceania	Cluster 3	Cluster2 8	Cluster3 20	2.2	
SRR5216895	New_Zealand	SB0130	Oceania	Cluster 3	Cluster2 8	Cluster3 20	2.2	
SRR5216896	New_Zealand	SB0130	Oceania	Cluster 3	Cluster2 8	Cluster3 20	2.2	
SRR5216897	New_Zealand	SB0130	Oceania	Cluster 3	Cluster2 8	Cluster3 21	2.2	
SRR5216898	New_Zealand	SB0130	Oceania	Cluster 3	Cluster2 8	Cluster3 22	2.2	
SRR5216900	New_Zealand	SB0130	Oceania	Cluster 3	Cluster2 8	Cluster3 22	2.2	
SRR5216901	New_Zealand	SB0130	Oceania	Cluster 3	Cluster2 8	Cluster3 21	2.2	
SRR5216902	New_Zealand	SB0130	Oceania	Cluster 3	Cluster2 8	Cluster3 22	2.2	
SRR5216905	New_Zealand	SB0130	Oceania	Cluster 3	Cluster2 8	Cluster3 21	2.2	
SRR5216907	New_Zealand	SB0130	Oceania	Cluster 3	Cluster2 8	Cluster3 20	2.2	
SRR5216915	New_Zealand	SB0130	Oceania	Cluster 3	Cluster2 8	Cluster3 20	2.2	
SRR5216917	New_Zealand	SB0130	Oceania	Cluster 3	Cluster2 8	Cluster3 21	2.2	
SRR5216919	New_Zealand	SB0130	Oceania	Cluster 3	Cluster2 8	Cluster3 22	2.2	
SRR5216920	New_Zealand	SB0130	Oceania	Cluster 3	Cluster2 8	Cluster3 20	2.2	
SRR5216922	New_Zealand	SB0130	Oceania	Cluster 3	Cluster2 8	Cluster3 20	2.2	
SRR5216924	New_Zealand	SB0130	Oceania	Cluster 3	Cluster2 8	Cluster3 20	2.2	
SRR5216930	New_Zealand	SB0130	Oceania	Cluster 3	Cluster2 8	Cluster3 21	2.2	
SRR5216931	New_Zealand	SB0130	Oceania	Cluster 3	Cluster2 8	Cluster3 21	2.2	
SRR5216932	New_Zealand	SB0130	Oceania	Cluster 3	Cluster2 8	Cluster3 20	2.2	
SRR5216935	New_Zealand	SB0130	Oceania	Cluster 3	Cluster2 8	Cluster3 20	2.2	
SRR5216942	New_Zealand	SB0130	Oceania	Cluster 3	Cluster2 8	Cluster3 21	2.2	
SRR5216945	New_Zealand	SB0130	Oceania	Cluster 3	Cluster2 8	Cluster3 21	2.2	
SRR5216948	New_Zealand	SB0130	Oceania	Cluster 3	Cluster2 8	Cluster3 20	2.2	
SRR5216951	New_Zealand	SB0130	Oceania	Cluster 3	Cluster2 8	Cluster3 20	2.2	
SRR5216953	New_Zealand	SB0130	Oceania	Cluster 3	Cluster2 8	Cluster3 22	2.2	
SRR5216955	New_Zealand	SB0130	Oceania	Cluster 3	Cluster2 8	Cluster3 22	2.2	
SRR5216960	New_Zealand	SB0130	Oceania	Cluster 3	Cluster2 8	Cluster3 22	2.2	
SRR5216963	New_Zealand	SB0130	Oceania	Cluster 3	Cluster2 8	Cluster3 22	2.2	
SRR5216964	New_Zealand	SB0130	Oceania	Cluster 3	Cluster2 8	Cluster3 20	2.2	
SRR5216968	New_Zealand	SB0130	Oceania	Cluster 3	Cluster2 8	Cluster3 20	2.2	
SRR5216969	New_Zealand	SB0130	Oceania	Cluster 3	Cluster2 8	Cluster3 20	2.2	
SRR5216972	New_Zealand	SB0130	Oceania	Cluster 3	Cluster2 8	Cluster3 22	2.2	
SRR5216973	New_Zealand	SB0130	Oceania	Cluster 3	Cluster2 8	Cluster3 21	2.2	
SRR5216974	New_Zealand	SB0130	Oceania	Cluster 3	Cluster2 8	Cluster3 21	2.2	
SRR5216977	New_Zealand	SB0130	Oceania	Cluster 3	Cluster2 8	Cluster3 22	2.2	
SRR5216984	New_Zealand	SB0130	Oceania	Cluster 3	Cluster2 8	Cluster3 22	2.2	
SRR5216985	New_Zealand	SB0130	Oceania	Cluster 3	Cluster2 8	Cluster3 22	2.2	
SRR5216695	New_Zealand	Not_found	Oceania	Cluster 3	Cluster2 9	Cluster3 24	2.3	
SRR5216697	New_Zealand	SB1504	Oceania	Cluster 3	Cluster2 9	Cluster3 24	2.3	
SRR5216703	New_Zealand	SB1504	Oceania	Cluster 3	Cluster2 9	Cluster3 23	2.3	
SRR5216705	New_Zealand	SB1504	Oceania	Cluster 3	Cluster2 9	Cluster3 24	2.3	
SRR5216708	New_Zealand	SB1504	Oceania	Cluster 3	Cluster2 9	Cluster3 24	2.3	
SRR5216712	New_Zealand	SB1504	Oceania	Cluster 3	Cluster2 9	Cluster3 23	2.3	
SRR5216714	New_Zealand	SB1504	Oceania	Cluster 3	Cluster2 9	Cluster3 24	2.3	
SRR5216716	New_Zealand	SB1504	Oceania	Cluster 3	Cluster2 9	Cluster3 23	2.3	
SRR5216722	New_Zealand	SB1504	Oceania	Cluster 3	Cluster2 9	Cluster3 24	2.3	
SRR5216723	New_Zealand	SB1504	Oceania	Cluster 3	Cluster2 9	Cluster3 23	2.3	

SRR5216725	New_Zealand	SB1504	Oceania	Cluster 3	Cluster2 9	Cluster3 24	2.3	
SRR5216732	New_Zealand	SB1504	Oceania	Cluster 3	Cluster2 9	Cluster3 23	2.3	
SRR5216733	New_Zealand	SB1504	Oceania	Cluster 3	Cluster2 9	Cluster3 24	2.3	
SRR5216738	New_Zealand	SB1504	Oceania	Cluster 3	Cluster2 9	Cluster3 24	2.3	
SRR5216739	New_Zealand	SB1504	Oceania	Cluster 3	Cluster2 9	Cluster3 23	2.3	
SRR5216747	New_Zealand	SB1504	Oceania	Cluster 3	Cluster2 9	Cluster3 23	2.3	
SRR5216750	New_Zealand	SB1504	Oceania	Cluster 3	Cluster2 9	Cluster3 23	2.3	
SRR5216752	New_Zealand	SB1504	Oceania	Cluster 3	Cluster2 9	Cluster3 24	2.3	
SRR5216755	New_Zealand	SB1504	Oceania	Cluster 3	Cluster2 9	Cluster3 23	2.3	
SRR5216760	New_Zealand	SB1504	Oceania	Cluster 3	Cluster2 9	Cluster3 23	2.3	
SRR5216761	New_Zealand	SB1504	Oceania	Cluster 3	Cluster2 9	Cluster3 24	2.3	
SRR5216762	New_Zealand	SB1504	Oceania	Cluster 3	Cluster2 9	Cluster3 24	2.3	
SRR5216767	New_Zealand	SB1504	Oceania	Cluster 3	Cluster2 9	Cluster3 24	2.3	
SRR5216771	New_Zealand	SB1504	Oceania	Cluster 3	Cluster2 9	Cluster3 24	2.3	
SRR5216782	New_Zealand	SB1504	Oceania	Cluster 3	Cluster2 9	Cluster3 24	2.3	
SRR5216786	New_Zealand	SB1504	Oceania	Cluster 3	Cluster2 9	Cluster3 23	2.3	
SRR5216790	New_Zealand	SB1504	Oceania	Cluster 3	Cluster2 9	Cluster3 23	2.3	
SRR5216791	New_Zealand	SB1504	Oceania	Cluster 3	Cluster2 9	Cluster3 24	2.3	
SRR5216801	New_Zealand	SB1504	Oceania	Cluster 3	Cluster2 9	Cluster3 24	2.3	
SRR5216807	New_Zealand	SB1504	Oceania	Cluster 3	Cluster2 9	Cluster3 24	2.3	
SRR5216808	New_Zealand	SB1504	Oceania	Cluster 3	Cluster2 9	Cluster3 23	2.3	
SRR5216811	New_Zealand	SB1504	Oceania	Cluster 3	Cluster2 9	Cluster3 23	2.3	
SRR5216817	New_Zealand	SB1504	Oceania	Cluster 3	Cluster2 9	Cluster3 24	2.3	
SRR5216820	New_Zealand	SB1504	Oceania	Cluster 3	Cluster2 9	Cluster3 24	2.3	
SRR5216824	New_Zealand	SB1504	Oceania	Cluster 3	Cluster2 9	Cluster3 23	2.3	
SRR5216826	New_Zealand	Not_found	Oceania	Cluster 3	Cluster2 9	Cluster3 24	2.3	
SRR5216827	New_Zealand	SB1504	Oceania	Cluster 3	Cluster2 9	Cluster3 23	2.3	
SRR5216828	New_Zealand	SB1504	Oceania	Cluster 3	Cluster2 9	Cluster3 24	2.3	
SRR5216831	New_Zealand	SB0130	Oceania	Cluster 3	Cluster2 9	Cluster3 23	2.3	
SRR5216832	New_Zealand	SB1504	Oceania	Cluster 3	Cluster2 9	Cluster3 24	2.3	
SRR5216833	New_Zealand	SB1504	Oceania	Cluster 3	Cluster2 9	Cluster3 23	2.3	
SRR5216836	New_Zealand	SB1504	Oceania	Cluster 3	Cluster2 9	Cluster3 24	2.3	
SRR5216837	New_Zealand	SB1504	Oceania	Cluster 3	Cluster2 9	Cluster3 24	2.3	
SRR5216842	New_Zealand	SB1504	Oceania	Cluster 3	Cluster2 9	Cluster3 23	2.3	
SRR5216844	New_Zealand	SB1504	Oceania	Cluster 3	Cluster2 9	Cluster3 24	2.3	
SRR5216848	New_Zealand	SB1504	Oceania	Cluster 3	Cluster2 9	Cluster3 24	2.3	
SRR5216849	New_Zealand	SB1504	Oceania	Cluster 3	Cluster2 9	Cluster3 23	2.3	
SRR5216851	New_Zealand	SB1504	Oceania	Cluster 3	Cluster2 9	Cluster3 24	2.3	
SRR5216852	New_Zealand	SB1504	Oceania	Cluster 3	Cluster2 9	Cluster3 24	2.3	
SRR5216859	New_Zealand	SB1504	Oceania	Cluster 3	Cluster2 9	Cluster3 24	2.3	
SRR5216862	New_Zealand	SB1504	Oceania	Cluster 3	Cluster2 9	Cluster3 24	2.3	
SRR5216871	New_Zealand	SB1504	Oceania	Cluster 3	Cluster2 9	Cluster3 23	2.3	
SRR5216875	New_Zealand	SB1504	Oceania	Cluster 3	Cluster2 9	Cluster3 24	2.3	
SRR5216878	New_Zealand	SB1504	Oceania	Cluster 3	Cluster2 9	Cluster3 23	2.3	
SRR5216885	New_Zealand	SB1504	Oceania	Cluster 3	Cluster2 9	Cluster3 24	2.3	
SRR5216886	New_Zealand	SB1504	Oceania	Cluster 3	Cluster2 9	Cluster3 24	2.3	
SRR5216890	New_Zealand	SB1504	Oceania	Cluster 3	Cluster2 9	Cluster3 23	2.3	
SRR5216899	New_Zealand	SB1504	Oceania	Cluster 3	Cluster2 9	Cluster3 24	2.3	
SRR5216904	New_Zealand	Not_found	Oceania	Cluster 3	Cluster2 9	Cluster3 23	2.3	
SRR5216918	New_Zealand	SB1504	Oceania	Cluster 3	Cluster2 9	Cluster3 24	2.3	
SRR5216921	New_Zealand	SB1504	Oceania	Cluster 3	Cluster2 9	Cluster3 24	2.3	

SRR5216925	New_Zealand	SB1504	Oceania	Cluster 3	Cluster2 9	Cluster3 24	2.3	
SRR5216927	New_Zealand	SB1504	Oceania	Cluster 3	Cluster2 9	Cluster3 24	2.3	
SRR5216933	New_Zealand	SB1504	Oceania	Cluster 3	Cluster2 9	Cluster3 23	2.3	
SRR5216938	New_Zealand	SB1504	Oceania	Cluster 3	Cluster2 9	Cluster3 23	2.3	
SRR5216939	New_Zealand	SB1504	Oceania	Cluster 3	Cluster2 9	Cluster3 24	2.3	
SRR5216944	New_Zealand	SB1504	Oceania	Cluster 3	Cluster2 9	Cluster3 24	2.3	
SRR5216946	New_Zealand	SB1504	Oceania	Cluster 3	Cluster2 9	Cluster3 23	2.3	
SRR5216950	New_Zealand	SB1504	Oceania	Cluster 3	Cluster2 9	Cluster3 23	2.3	
SRR5216957	New_Zealand	SB1504	Oceania	Cluster 3	Cluster2 9	Cluster3 24	2.3	
SRR5216962	New_Zealand	SB1504	Oceania	Cluster 3	Cluster2 9	Cluster3 24	2.3	
SRR5216971	New_Zealand	SB1504	Oceania	Cluster 3	Cluster2 9	Cluster3 24	2.3	
SRR1791697	USA	SB0145	North America	Cluster 2	Cluster2 7	Cluster3 17	3.1	
SRR1791807	Mexico	SB0145	North America	Cluster 2	Cluster2 7	Cluster3 17	3.1	
SRR1791923	Mexico	SB0145	North America	Cluster 2	Cluster2 7	Cluster3 19	3.1	
SRR1791938	Mexico	SB0145	North America	Cluster 2	Cluster2 7	Cluster3 17	3.1	
SRR1792034	USA	SB0145	North America	Cluster 2	Cluster2 7	Cluster3 18	3.1	
SRR1792240	USA	SB0145	North America	Cluster 2	Cluster2 7	Cluster3 18	3.1	
SRR1792298	Mexico	SB0145	North America	Cluster 2	Cluster2 7	Cluster3 17	3.1	
SRR1792299	UNKNOWN	Not_found	UNKNOWN	Cluster 2	Cluster2 7	Cluster3 19	3.1	
SRR1792300	Mexico	SB0145	North America	Cluster 2	Cluster2 7	Cluster3 17	3.1	
SRR1792301	Mexico	SB0145	North America	Cluster 2	Cluster2 7	Cluster3 17	3.1	
SRR1792302	Mexico	SB0145	North America	Cluster 2	Cluster2 7	Cluster3 17	3.1	
SRR1792303	UNKNOWN	Not_found	UNKNOWN	Cluster 2	Cluster2 7	Cluster3 17	3.1	
SRR1792304	UNKNOWN	SB0145	UNKNOWN	Cluster 2	Cluster2 7	Cluster3 17	3.1	
SRR1792311	USA	SB0145	North America	Cluster 2	Cluster2 7	Cluster3 17	3.1	
SRR1792312	USA	SB0145	North America	Cluster 2	Cluster2 7	Cluster3 17	3.1	
SRR1792314	USA	SB0145	North America	Cluster 2	Cluster2 7	Cluster3 17	3.1	
SRR1792315	USA	SB0145	North America	Cluster 2	Cluster2 7	Cluster3 19	3.1	
SRR1792316	USA	SB0145	North America	Cluster 2	Cluster2 7	Cluster3 17	3.1	
SRR1792317	USA	SB0145	North America	Cluster 2	Cluster2 7	Cluster3 17	3.1	
SRR1792318	USA	SB0145	North America	Cluster 2	Cluster2 7	Cluster3 17	3.1	
SRR1792320	USA	SB0145	North America	Cluster 2	Cluster2 7	Cluster3 17	3.1	
SRR1792326	USA	SB0145	North America	Cluster 2	Cluster2 7	Cluster3 17	3.1	
SRR1792327	USA	SB0145	North America	Cluster 2	Cluster2 7	Cluster3 17	3.1	
SRR1792332	USA	SB0145	North America	Cluster 2	Cluster2 7	Cluster3 17	3.1	
SRR1792337	USA	SB0145	North America	Cluster 2	Cluster2 7	Cluster3 17	3.1	
SRR1792342	USA	SB0145	North America	Cluster 2	Cluster2 7	Cluster3 17	3.1	
SRR1792443	USA	SB0145	North America	Cluster 2	Cluster2 7	Cluster3 17	3.1	
6556.mbURU-001	Uruguay	SB0145	Latin America	Cluster 2	Cluster2 4	Cluster3 11		
Tb74.mbURU-003	Uruguay	SB0145	Latin America	Cluster 2	Cluster2 4	Cluster3 11		
Tb79.mbURU-005	Uruguay	SB0145	Latin America	Cluster 2	Cluster2 4	Cluster3 11		
mbURU-016.mbURU-016	Uruguay	SB0145	Latin America	Cluster 2	Cluster2 4	Cluster3 11		
mbURU-019.mbURU-019	Uruguay	SB0145	Latin America	Cluster 2	Cluster2 4	Cluster3 11		
SRR1791760	Mexico	SB1040	North America	Cluster 2	Cluster2 4	Cluster3 9		
SRR1791780	USA	SB0145	North America	Cluster 2	Cluster2 4	Cluster3 10		
SRR1791799	Mexico	SB0145	North America	Cluster 2	Cluster2 4	Cluster3 9		
SRR1791800	Mexico	SB0145	North America	Cluster 2	Cluster2 4	Cluster3 9		
SRR1791805	UNKNOWN	SB0145	UNKNOWN	Cluster 2	Cluster2 4	Cluster3 9		
SRR1791954	Mexico	SB1040	North America	Cluster 2	Cluster2 4	Cluster3 9		
SRR1792014	USA	SB0145	North America	Cluster 2	Cluster2 4	Cluster3 10		

SRR1792107	USA	SB0145	North America	Cluster 2	Cluster2 4	Cluster3 10		
SRR1792146	USA	SB0145	North America	Cluster 2	Cluster2 4	Cluster3 10		
SRR1792152	USA	SB0145	North America	Cluster 2	Cluster2 4	Cluster3 10		
SRR1792159	USA	SB0145	North America	Cluster 2	Cluster2 4	Cluster3 10		
SRR1792241	Mexico	SB0145	North America	Cluster 2	Cluster2 4	Cluster3 9		
SRR1792282	USA	SB0145	North America	Cluster 2	Cluster2 4	Cluster3 9		
SRR1792350	USA	SB0145	North America	Cluster 2	Cluster2 4	Cluster3 10		
SRR1791705	Mexico	SB0145	North America	Cluster 2	Cluster2 5	Cluster3 14	3.2	
SRR1791738	USA	SB0145	North America	Cluster 2	Cluster2 5	Cluster3 13	3.2	
SRR1791817	USA	SB0145	North America	Cluster 2	Cluster2 5	Cluster3 13	3.2	
SRR1791818	USA	SB0145	North America	Cluster 2	Cluster2 5	Cluster3 13	3.2	
SRR1791874	USA	SB0145	North America	Cluster 2	Cluster2 5	Cluster3 12	3.2	
SRR1792024	USA	SB0145	North America	Cluster 2	Cluster2 5	Cluster3 13	3.2	
SRR1792054	USA	SB0145	North America	Cluster 2	Cluster2 5	Cluster3 13	3.2	
SRR1792055	USA	SB0145	North America	Cluster 2	Cluster2 5	Cluster3 13	3.2	
SRR1792185	USA	SB0145	North America	Cluster 2	Cluster2 5	Cluster3 13	3.2	
SRR1792242	USA	SB0145	North America	Cluster 2	Cluster2 5	Cluster3 13	3.2	
SRR1792257	USA	SB0145	North America	Cluster 2	Cluster2 5	Cluster3 13	3.2	
SRR1792344	USA	SB0145	North America	Cluster 2	Cluster2 5	Cluster3 13	3.2	
SRR1792351	USA	SB0145	North America	Cluster 2	Cluster2 5	Cluster3 13	3.2	
SRR1792352	USA	SB0145	North America	Cluster 2	Cluster2 5	Cluster3 12	3.2	
SRR1792355	USA	SB2007	North America	Cluster 2	Cluster2 5	Cluster3 14	3.2	
SRR1792375	USA	SB0145	North America	Cluster 2	Cluster2 5	Cluster3 12	3.2	
SRR1792377	USA	SB0145	North America	Cluster 2	Cluster2 5	Cluster3 13	3.2	
SRR1792378	USA	SB0145	North America	Cluster 2	Cluster2 5	Cluster3 12	3.2	
SRR1792385	USA	SB0145	North America	Cluster 2	Cluster2 5	Cluster3 13	3.2	
SRR1792390	USA	SB0145	North America	Cluster 2	Cluster2 5	Cluster3 12	3.2	
SRR1792397	USA	SB0145	North America	Cluster 2	Cluster2 5	Cluster3 13	3.2	
SRR1792401	USA	SB0145	North America	Cluster 2	Cluster2 5	Cluster3 13	3.2	
SRR1792411	USA	Not_found	North America	Cluster 2	Cluster2 5	Cluster3 14	3.2	
SRR1792442	USA	Not_found	North America	Cluster 2	Cluster2 5	Cluster3 14	3.2	
SRR1792459	USA	SB1039	North America	Cluster 2	Cluster2 5	Cluster3 13	3.2	
SRR1792470	USA	SB0145	North America	Cluster 2	Cluster2 5	Cluster3 13	3.2	
SRR1792494	USA	SB0145	North America	Cluster 2	Cluster2 5	Cluster3 12	3.2	
SRR1792497	USA	SB0145	North America	Cluster 2	Cluster2 5	Cluster3 13	3.2	
SRR3091233	USA	SB0145	North America	Cluster 2	Cluster2 5	Cluster3 13	3.2	
SRR3091234	USA	SB0145	North America	Cluster 2	Cluster2 5	Cluster3 12	3.2	
SRR3091235	USA	SB0145	North America	Cluster 2	Cluster2 5	Cluster3 13	3.2	
SRR3091237	USA	Not_found	North America	Cluster 2	Cluster2 5	Cluster3 13	3.2	
SRR3091238	USA	SB0145	North America	Cluster 2	Cluster2 5	Cluster3 13	3.2	
SRR3091239	USA	Not_found	North America	Cluster 2	Cluster2 5	Cluster3 13	3.2	
SRR3091250	USA	Not_found	North America	Cluster 2	Cluster2 5	Cluster3 13	3.2	
SRR3091251	USA	Not_found	North America	Cluster 2	Cluster2 5	Cluster3 13	3.2	
SRR3091254	USA	Not_found	North America	Cluster 2	Cluster2 5	Cluster3 13	3.2	
SRR1791693	USA	SB0145	North America	Cluster 2	Cluster2 6	Cluster3 15	3.2	
SRR1791694	USA	SB0145	North America	Cluster 2	Cluster2 6	Cluster3 15	3.2	
SRR1791696	USA	SB0145	North America	Cluster 2	Cluster2 6	Cluster3 15	3.2	
SRR1791698	USA	SB0145	North America	Cluster 2	Cluster2 6	Cluster3 15	3.2	
SRR1791699	USA	SB0145	North America	Cluster 2	Cluster2 6	Cluster3 15	3.2	
SRR1791715	USA	SB0145	North America	Cluster 2	Cluster2 6	Cluster3 16	3.2	
SRR1791716	USA	SB0145	North America	Cluster 2	Cluster2 6	Cluster3 15	3.2	

SRR1791718	USA	SB0145	North America	Cluster 2	Cluster2 6	Cluster3 15	3.2	
SRR1791719	USA	SB0145	North America	Cluster 2	Cluster2 6	Cluster3 15	3.2	
SRR1791730	USA	Not_found	North America	Cluster 2	Cluster2 6	Cluster3 15	3.2	
SRR1791731	USA	SB0145	North America	Cluster 2	Cluster2 6	Cluster3 15	3.2	
SRR1791751	Mexico	SB0145	North America	Cluster 2	Cluster2 6	Cluster3 16	3.2	
SRR1791757	USA	SB0145	North America	Cluster 2	Cluster2 6	Cluster3 15	3.2	
SRR1791759	Mexico	SB0145	North America	Cluster 2	Cluster2 6	Cluster3 16	3.2	
SRR1791772	USA	SB0145	North America	Cluster 2	Cluster2 6	Cluster3 15	3.2	
SRR1791775	USA	SB0145	North America	Cluster 2	Cluster2 6	Cluster3 15	3.2	
SRR1791803	USA	SB0145	North America	Cluster 2	Cluster2 6	Cluster3 15	3.2	
SRR1791810	USA	SB0145	North America	Cluster 2	Cluster2 6	Cluster3 15	3.2	
SRR1791812	USA	SB0145	North America	Cluster 2	Cluster2 6	Cluster3 15	3.2	
SRR1791819	USA	SB0145	North America	Cluster 2	Cluster2 6	Cluster3 15	3.2	
SRR1791825	USA	SB0145	North America	Cluster 2	Cluster2 6	Cluster3 15	3.2	
SRR1791839	USA	SB0145	North America	Cluster 2	Cluster2 6	Cluster3 15	3.2	
SRR1791854	USA	SB0145	North America	Cluster 2	Cluster2 6	Cluster3 15	3.2	
SRR1791857	USA	SB0145	North America	Cluster 2	Cluster2 6	Cluster3 15	3.2	
SRR1791859	USA	SB0145	North America	Cluster 2	Cluster2 6	Cluster3 16	3.2	
SRR1791863	USA	SB0145	North America	Cluster 2	Cluster2 6	Cluster3 15	3.2	
SRR1791865	Mexico	SB1040	North America	Cluster 2	Cluster2 6	Cluster3 16	3.2	
SRR1791873	USA	SB0145	North America	Cluster 2	Cluster2 6	Cluster3 16	3.2	
SRR1791888	USA	SB0145	North America	Cluster 2	Cluster2 6	Cluster3 16	3.2	
SRR1791894	USA	SB0145	North America	Cluster 2	Cluster2 6	Cluster3 15	3.2	
SRR1791902	USA	SB0292	North America	Cluster 2	Cluster2 6	Cluster3 15	3.2	
SRR1791925	USA	SB0145	North America	Cluster 2	Cluster2 6	Cluster3 15	3.2	
SRR1791929	USA	SB0145	North America	Cluster 2	Cluster2 6	Cluster3 15	3.2	
SRR1791930	USA	SB0145	North America	Cluster 2	Cluster2 6	Cluster3 15	3.2	
SRR1791976	Mexico	SB0145	North America	Cluster 2	Cluster2 6	Cluster3 16	3.2	
SRR1791986	USA	SB0145	North America	Cluster 2	Cluster2 6	Cluster3 15	3.2	
SRR1792012	USA	SB0145	North America	Cluster 2	Cluster2 6	Cluster3 15	3.2	
SRR1792018	USA	SB0145	North America	Cluster 2	Cluster2 6	Cluster3 15	3.2	
SRR1792019	USA	SB0145	North America	Cluster 2	Cluster2 6	Cluster3 15	3.2	
SRR1792021	USA	SB0145	North America	Cluster 2	Cluster2 6	Cluster3 15	3.2	
SRR1792023	USA	SB0145	North America	Cluster 2	Cluster2 6	Cluster3 15	3.2	
SRR1792028	Mexico	SB1040	North America	Cluster 2	Cluster2 6	Cluster3 16	3.2	
SRR1792032	USA	SB0145	North America	Cluster 2	Cluster2 6	Cluster3 15	3.2	
SRR1792037	USA	SB0145	North America	Cluster 2	Cluster2 6	Cluster3 15	3.2	
SRR1792050	USA	SB0145	North America	Cluster 2	Cluster2 6	Cluster3 15	3.2	
SRR1792053	USA	SB0145	North America	Cluster 2	Cluster2 6	Cluster3 15	3.2	
SRR1792056	USA	SB0145	North America	Cluster 2	Cluster2 6	Cluster3 15	3.2	
SRR1792057	USA	SB0145	North America	Cluster 2	Cluster2 6	Cluster3 15	3.2	
SRR1792058	USA	SB0145	North America	Cluster 2	Cluster2 6	Cluster3 15	3.2	
SRR1792063	Mexico	SB0145	North America	Cluster 2	Cluster2 6	Cluster3 16	3.2	
SRR1792071	Mexico	SB0145	North America	Cluster 2	Cluster2 6	Cluster3 16	3.2	
SRR1792089	USA	SB0145	North America	Cluster 2	Cluster2 6	Cluster3 15	3.2	
SRR1792096	USA	SB0815	North America	Cluster 2	Cluster2 6	Cluster3 15	3.2	
SRR1792098	USA	SB0145	North America	Cluster 2	Cluster2 6	Cluster3 15	3.2	
SRR1792112	USA	SB0145	North America	Cluster 2	Cluster2 6	Cluster3 15	3.2	
SRR1792117	UNKNOWN	SB0145	UNKNOWN	Cluster 2	Cluster2 6	Cluster3 16	3.2	
SRR1792156	UNKNOWN	SB0145	UNKNOWN	Cluster 2	Cluster2 6	Cluster3 16	3.2	
SRR1792160	USA	SB0145	North America	Cluster 2	Cluster2 6	Cluster3 15	3.2	

SRR1792186	USA	SB0145	North America	Cluster 2	Cluster2 6	Cluster3 15	3.2	
SRR1792243	USA	SB0145	North America	Cluster 2	Cluster2 6	Cluster3 15	3.2	
SRR1792252	USA	SB0292	North America	Cluster 2	Cluster2 6	Cluster3 15	3.2	
SRR1792253	USA	SB0292	North America	Cluster 2	Cluster2 6	Cluster3 15	3.2	
SRR1792255	USA	SB0145	North America	Cluster 2	Cluster2 6	Cluster3 15	3.2	
SRR1792256	USA	SB0145	North America	Cluster 2	Cluster2 6	Cluster3 15	3.2	
SRR1792262	USA	SB0145	North America	Cluster 2	Cluster2 6	Cluster3 15	3.2	
SRR1792265	USA	SB0145	North America	Cluster 2	Cluster2 6	Cluster3 15	3.2	
SRR1792266	USA	SB0145	North America	Cluster 2	Cluster2 6	Cluster3 15	3.2	
SRR1792267	USA	SB0145	North America	Cluster 2	Cluster2 6	Cluster3 15	3.2	
SRR1792271	USA	SB0145	North America	Cluster 2	Cluster2 6	Cluster3 15	3.2	
SRR1792272	USA	SB0145	North America	Cluster 2	Cluster2 6	Cluster3 15	3.2	
SRR1792286	Mexico	SB0145	North America	Cluster 2	Cluster2 6	Cluster3 15	3.2	
SRR1792289	USA	SB0145	North America	Cluster 2	Cluster2 6	Cluster3 15	3.2	
SRR1792290	USA	SB0145	North America	Cluster 2	Cluster2 6	Cluster3 15	3.2	
SRR1792291	USA	SB0145	North America	Cluster 2	Cluster2 6	Cluster3 15	3.2	
SRR1792294	USA	SB0145	North America	Cluster 2	Cluster2 6	Cluster3 15	3.2	
SRR1792306	USA	SB0145	North America	Cluster 2	Cluster2 6	Cluster3 15	3.2	
SRR1792323	USA	SB0145	North America	Cluster 2	Cluster2 6	Cluster3 15	3.2	
SRR1792324	USA	SB0145	North America	Cluster 2	Cluster2 6	Cluster3 15	3.2	
SRR1792325	USA	SB0145	North America	Cluster 2	Cluster2 6	Cluster3 15	3.2	
SRR1792331	USA	Not_found	North America	Cluster 2	Cluster2 6	Cluster3 15	3.2	
SRR1792335	USA	SB0145	North America	Cluster 2	Cluster2 6	Cluster3 15	3.2	
SRR1792370	USA	SB0145	North America	Cluster 2	Cluster2 6	Cluster3 15	3.2	
SRR1792373	USA	SB0145	North America	Cluster 2	Cluster2 6	Cluster3 15	3.2	
SRR1792374	USA	SB0145	North America	Cluster 2	Cluster2 6	Cluster3 15	3.2	
SRR1792376	USA	SB0145	North America	Cluster 2	Cluster2 6	Cluster3 15	3.2	
SRR1792379	USA	SB0145	North America	Cluster 2	Cluster2 6	Cluster3 15	3.2	
SRR1792380	USA	SB0145	North America	Cluster 2	Cluster2 6	Cluster3 15	3.2	
SRR1792381	USA	SB0145	North America	Cluster 2	Cluster2 6	Cluster3 15	3.2	
SRR1792382	USA	SB0145	North America	Cluster 2	Cluster2 6	Cluster3 15	3.2	
SRR1792383	USA	SB0145	North America	Cluster 2	Cluster2 6	Cluster3 15	3.2	
SRR1792386	USA	SB0145	North America	Cluster 2	Cluster2 6	Cluster3 15	3.2	
SRR1792387	USA	SB0145	North America	Cluster 2	Cluster2 6	Cluster3 15	3.2	
SRR1792393	USA	SB0145	North America	Cluster 2	Cluster2 6	Cluster3 15	3.2	
SRR1792394	USA	SB0145	North America	Cluster 2	Cluster2 6	Cluster3 15	3.2	
SRR1792395	USA	SB0145	North America	Cluster 2	Cluster2 6	Cluster3 15	3.2	
SRR1792396	USA	SB0145	North America	Cluster 2	Cluster2 6	Cluster3 15	3.2	
SRR1792398	USA	SB0145	North America	Cluster 2	Cluster2 6	Cluster3 15	3.2	
SRR1792399	USA	SB0145	North America	Cluster 2	Cluster2 6	Cluster3 15	3.2	
SRR1792400	USA	SB0145	North America	Cluster 2	Cluster2 6	Cluster3 15	3.2	
SRR1792402	USA	SB0145	North America	Cluster 2	Cluster2 6	Cluster3 15	3.2	
SRR1792403	USA	SB0145	North America	Cluster 2	Cluster2 6	Cluster3 15	3.2	
SRR1792404	USA	SB0145	North America	Cluster 2	Cluster2 6	Cluster3 15	3.2	
SRR1792405	USA	SB0145	North America	Cluster 2	Cluster2 6	Cluster3 15	3.2	
SRR1792406	USA	SB0145	North America	Cluster 2	Cluster2 6	Cluster3 15	3.2	
SRR1792407	USA	SB0145	North America	Cluster 2	Cluster2 6	Cluster3 15	3.2	
SRR1792408	USA	SB0145	North America	Cluster 2	Cluster2 6	Cluster3 15	3.2	
SRR1792412	USA	SB0145	North America	Cluster 2	Cluster2 6	Cluster3 15	3.2	
SRR1792418	USA	SB0145	North America	Cluster 2	Cluster2 6	Cluster3 15	3.2	
SRR1792419	USA	SB0145	North America	Cluster 2	Cluster2 6	Cluster3 15	3.2	

SRR1792420	USA	SB0145	North America	Cluster 2	Cluster2 6	Cluster3 15	3.2	
SRR1792423	USA	SB0145	North America	Cluster 2	Cluster2 6	Cluster3 15	3.2	
SRR1792424	USA	SB0145	North America	Cluster 2	Cluster2 6	Cluster3 15	3.2	
SRR1792425	USA	SB0145	North America	Cluster 2	Cluster2 6	Cluster3 15	3.2	
SRR1792426	USA	SB0145	North America	Cluster 2	Cluster2 6	Cluster3 15	3.2	
SRR1792427	USA	SB0145	North America	Cluster 2	Cluster2 6	Cluster3 15	3.2	
SRR1792428	USA	SB0145	North America	Cluster 2	Cluster2 6	Cluster3 15	3.2	
SRR1792430	USA	SB0145	North America	Cluster 2	Cluster2 6	Cluster3 15	3.2	
SRR1792431	USA	SB0145	North America	Cluster 2	Cluster2 6	Cluster3 15	3.2	
SRR1792432	UNKNOWN	SB1040	UNKNOWN	Cluster 2	Cluster2 6	Cluster3 16	3.2	
SRR1792433	USA	SB0145	North America	Cluster 2	Cluster2 6	Cluster3 15	3.2	
SRR1792434	USA	SB0145	North America	Cluster 2	Cluster2 6	Cluster3 15	3.2	
SRR1792436	USA	SB0145	North America	Cluster 2	Cluster2 6	Cluster3 15	3.2	
SRR1792437	USA	SB0145	North America	Cluster 2	Cluster2 6	Cluster3 15	3.2	
SRR1792439	USA	SB0145	North America	Cluster 2	Cluster2 6	Cluster3 15	3.2	
SRR1792440	USA	SB0145	North America	Cluster 2	Cluster2 6	Cluster3 15	3.2	
SRR1792441	Mexico	SB0145	North America	Cluster 2	Cluster2 6	Cluster3 16	3.2	
SRR1792448	USA	SB0145	North America	Cluster 2	Cluster2 6	Cluster3 16	3.2	
SRR1792451	USA	SB0145	North America	Cluster 2	Cluster2 6	Cluster3 15	3.2	
SRR1792452	USA	Not_found	North America	Cluster 2	Cluster2 6	Cluster3 15	3.2	
SRR1792453	USA	SB0145	North America	Cluster 2	Cluster2 6	Cluster3 15	3.2	
SRR1792455	USA	SB0145	North America	Cluster 2	Cluster2 6	Cluster3 16	3.2	
SRR1792456	USA	SB0145	North America	Cluster 2	Cluster2 6	Cluster3 15	3.2	
SRR1792457	USA	SB1039	North America	Cluster 2	Cluster2 6	Cluster3 15	3.2	
SRR1792458	USA	SB0145	North America	Cluster 2	Cluster2 6	Cluster3 15	3.2	
SRR1792460	USA	SB0145	North America	Cluster 2	Cluster2 6	Cluster3 15	3.2	
SRR1792461	USA	SB0145	North America	Cluster 2	Cluster2 6	Cluster3 15	3.2	
SRR1792462	USA	SB0145	North America	Cluster 2	Cluster2 6	Cluster3 15	3.2	
SRR1792463	USA	SB0145	North America	Cluster 2	Cluster2 6	Cluster3 15	3.2	
SRR1792464	USA	SB0145	North America	Cluster 2	Cluster2 6	Cluster3 15	3.2	
SRR1792465	USA	SB0145	North America	Cluster 2	Cluster2 6	Cluster3 15	3.2	
SRR1792466	USA	SB0145	North America	Cluster 2	Cluster2 6	Cluster3 15	3.2	
SRR1792467	USA	SB0145	North America	Cluster 2	Cluster2 6	Cluster3 15	3.2	
SRR1792468	USA	SB0145	North America	Cluster 2	Cluster2 6	Cluster3 15	3.2	
SRR1792469	USA	SB0145	North America	Cluster 2	Cluster2 6	Cluster3 15	3.2	
SRR1792471	USA	SB0145	North America	Cluster 2	Cluster2 6	Cluster3 15	3.2	
SRR1792483	USA	SB0145	North America	Cluster 2	Cluster2 6	Cluster3 15	3.2	
SRR1792501	USA	SB0145	North America	Cluster 2	Cluster2 6	Cluster3 15	3.2	
SRR1792503	USA	SB0145	North America	Cluster 2	Cluster2 6	Cluster3 15	3.2	
SRR1792507	USA	SB0145	North America	Cluster 2	Cluster2 6	Cluster3 15	3.2	
SRR1792508	USA	SB0145	North America	Cluster 2	Cluster2 6	Cluster3 15	3.2	
SRR1792509	USA	SB0145	North America	Cluster 2	Cluster2 6	Cluster3 15	3.2	
SRR1792510	USA	SB0145	North America	Cluster 2	Cluster2 6	Cluster3 15	3.2	
SRR3091232	USA	SB0145	North America	Cluster 2	Cluster2 6	Cluster3 15	3.2	
SRR3091236	USA	SB0145	North America	Cluster 2	Cluster2 6	Cluster3 15	3.2	
SRR3091240	USA	SB0145	North America	Cluster 2	Cluster2 6	Cluster3 15	3.2	
SRR3091241	USA	SB0145	North America	Cluster 2	Cluster2 6	Cluster3 15	3.2	
SRR3091242	USA	Not_found	North America	Cluster 2	Cluster2 6	Cluster3 15	3.2	
SRR3091245	USA	SB2100	North America	Cluster 2	Cluster2 6	Cluster3 15	3.2	
SRR3091246	USA	SB0145	North America	Cluster 2	Cluster2 6	Cluster3 15	3.2	
SRR3091247	USA	SB0145	North America	Cluster 2	Cluster2 6	Cluster3 15	3.2	

SRR3091248	USA	SB0145	North America	Cluster 2	Cluster2 6	Cluster3 15	3.2	
SRR3091249	USA	SB0145	North America	Cluster 2	Cluster2 6	Cluster3 15	3.2	
SRR3091252	USA	SB0145	North America	Cluster 2	Cluster2 6	Cluster3 15	3.2	
SRR3091253	USA	SB0145	North America	Cluster 2	Cluster2 6	Cluster3 15	3.2	
SRR3091255	USA	SB0145	North America	Cluster 2	Cluster2 6	Cluster3 15	3.2	
SRR1791700	Mexico	SB1499	North America	Cluster 1	Cluster2 3	Cluster3 6	4.1	
SRR1791779	UNKNOWN	Not_found	UNKNOWN	Cluster 1	Cluster2 3	Cluster3 8	4.1	
SRR1791834	Mexico	SB1499	North America	Cluster 1	Cluster2 3	Cluster3 6	4.1	
SRR1791942	Mexico	SB1499	North America	Cluster 1	Cluster2 3	Cluster3 6	4.1	
SRR1791963	UNKNOWN	SB1499	UNKNOWN	Cluster 1	Cluster2 3	Cluster3 6	4.1	
SRR1791964	UNKNOWN	SB1499	UNKNOWN	Cluster 1	Cluster2 3	Cluster3 6	4.1	
SRR1791965	UNKNOWN	SB1499	UNKNOWN	Cluster 1	Cluster2 3	Cluster3 6	4.1	
SRR1791968	UNKNOWN	SB1499	UNKNOWN	Cluster 1	Cluster2 3	Cluster3 6	4.1	
SRR1791971	UNKNOWN	SB1499	UNKNOWN	Cluster 1	Cluster2 3	Cluster3 6	4.1	
SRR1791972	UNKNOWN	SB1499	UNKNOWN	Cluster 1	Cluster2 3	Cluster3 6	4.1	
SRR1792356	Mexico	Not_found	North America	Cluster 1	Cluster2 3	Cluster3 7	4.1	
SRR1792372	Mexico	Not_found	North America	Cluster 1	Cluster2 3	Cluster3 7	4.1	
Tb75.mbURU-004	Uruguay	SB0274	Latin America	Cluster 1	Cluster2 2	Cluster3 4	4.2	
Tb86.mbURU-006	Uruguay	SB1072	Latin America	Cluster 1	Cluster2 2	Cluster3 4	4.2	
Tb89.mbURU-008	Uruguay	SB0274	Latin America	Cluster 1	Cluster2 2	Cluster3 4	4.2	
Tb90.mbURU-009	Uruguay	SB0274	Latin America	Cluster 1	Cluster2 2	Cluster3 4	4.2	
mbURU-011.mbURU-011	Uruguay	SB0274	Latin America	Cluster 1	Cluster2 2	Cluster3 4	4.2	
mbURU-013.mbURU-013	Uruguay	SB0274	Latin America	Cluster 1	Cluster2 2	Cluster3 4	4.2	
mbURU-015.mbURU-015	Uruguay	SB0274	Latin America	Cluster 1	Cluster2 2	Cluster3 4	4.2	
mbURU-021.mbURU-021	Uruguay	SB0274	Latin America	Cluster 1	Cluster2 2	Cluster3 4	4.2	
mbURU-022.mbURU-022	Uruguay	SB0140	Latin America	Cluster 1	Cluster2 2	Cluster3 4	4.2	
mbURU-023.mbURU-023	Uruguay	SB0274	Latin America	Cluster 1	Cluster2 2	Cluster3 4	4.2	
SRR1791707	USA	SB0140	North America	Cluster 1	Cluster2 2	Cluster3 4	4.2	
SRR1791752	Mexico	Not_found	North America	Cluster 1	Cluster2 2	Cluster3 4	4.2	
SRR1791762	Mexico	SB0140	North America	Cluster 1	Cluster2 2	Cluster3 3	4.2	
SRR1791763	Mexico	Not_found	North America	Cluster 1	Cluster2 2	Cluster3 3	4.2	
SRR1791764	UNKNOWN	SB0140	UNKNOWN	Cluster 1	Cluster2 2	Cluster3 3	4.2	
SRR1791765	Mexico	SB1757	North America	Cluster 1	Cluster2 2	Cluster3 3	4.2	
SRR1791766	Mexico	SB0307	North America	Cluster 1	Cluster2 2	Cluster3 4	4.2	
SRR1791773	Mexico	SB0140	North America	Cluster 1	Cluster2 2	Cluster3 5	4.2	
SRR1791785	Mexico	SB0140	North America	Cluster 1	Cluster2 2	Cluster3 4	4.2	
SRR1791788	Mexico	SB1033	North America	Cluster 1	Cluster2 2	Cluster3 3	4.2	
SRR1791809	Mexico	SB0140	North America	Cluster 1	Cluster2 2	Cluster3 3	4.2	
SRR1791820	UNKNOWN	SB0140	UNKNOWN	Cluster 1	Cluster2 2	Cluster3 5	4.2	
SRR1791821	UNKNOWN	SB0140	UNKNOWN	Cluster 1	Cluster2 2	Cluster3 3	4.2	
SRR1791822	Mexico	SB0140	North America	Cluster 1	Cluster2 2	Cluster3 5	4.2	
SRR1791826	Mexico	SB0140	North America	Cluster 1	Cluster2 2	Cluster3 4	4.2	
SRR1791829	Mexico	SB0140	North America	Cluster 1	Cluster2 2	Cluster3 3	4.2	
SRR1791847	Mexico	SB0971	North America	Cluster 1	Cluster2 2	Cluster3 3	4.2	
SRR1791861	Mexico	SB0140	North America	Cluster 1	Cluster2 2	Cluster3 5	4.2	
SRR1791864	Mexico	SB0140	North America	Cluster 1	Cluster2 2	Cluster3 4	4.2	
SRR1791872	UNKNOWN	SB0140	UNKNOWN	Cluster 1	Cluster2 2	Cluster3 5	4.2	
SRR1791881	USA	SB0140	North America	Cluster 1	Cluster2 2	Cluster3 4	4.2	
SRR1791884	UNKNOWN	SB0971	UNKNOWN	Cluster 1	Cluster2 2	Cluster3 3	4.2	

SRR1791885	USA	SB0140	North America	Cluster 1	Cluster2 2	Cluster3 4	4.2	
SRR1791907	Mexico	SB1751	North America	Cluster 1	Cluster2 2	Cluster3 3	4.2	
SRR1791927	Mexico	SB0971	North America	Cluster 1	Cluster2 2	Cluster3 3	4.2	
SRR1791928	Mexico	SB1502	North America	Cluster 1	Cluster2 2	Cluster3 3	4.2	
SRR1791933	UNKNOWN	SB2014	UNKNOWN	Cluster 1	Cluster2 2	Cluster3 3	4.2	
SRR1791941	UNKNOWN	SB0971	UNKNOWN	Cluster 1	Cluster2 2	Cluster3 3	4.2	
SRR1791945	Mexico	Not_found	North America	Cluster 1	Cluster2 2	Cluster3 3	4.2	
SRR1791947	USA	SB0271	North America	Cluster 1	Cluster2 2	Cluster3 5	4.2	
SRR1791949	USA	SB0971	North America	Cluster 1	Cluster2 2	Cluster3 3	4.2	
SRR1791951	UNKNOWN	SB0140	UNKNOWN	Cluster 1	Cluster2 2	Cluster3 5	4.2	
SRR1791952	Mexico	SB0971	North America	Cluster 1	Cluster2 2	Cluster3 3	4.2	
SRR1791956	Mexico	SB0971	North America	Cluster 1	Cluster2 2	Cluster3 3	4.2	
SRR1791958	Mexico	SB0140	North America	Cluster 1	Cluster2 2	Cluster3 4	4.2	
SRR1791967	USA	SB0271	North America	Cluster 1	Cluster2 2	Cluster3 5	4.2	
SRR1791975	Mexico	SB0140	North America	Cluster 1	Cluster2 2	Cluster3 4	4.2	
SRR1791977	Mexico	SB0140	North America	Cluster 1	Cluster2 2	Cluster3 5	4.2	
SRR1791978	UNKNOWN	SB0140	UNKNOWN	Cluster 1	Cluster2 2	Cluster3 4	4.2	
SRR1791979	USA	SB0271	North America	Cluster 1	Cluster2 2	Cluster3 5	4.2	
SRR1791980	USA	SB0271	North America	Cluster 1	Cluster2 2	Cluster3 5	4.2	
SRR1791981	USA	SB0271	North America	Cluster 1	Cluster2 2	Cluster3 5	4.2	
SRR1791982	USA	SB0271	North America	Cluster 1	Cluster2 2	Cluster3 5	4.2	
SRR1791983	USA	SB0271	North America	Cluster 1	Cluster2 2	Cluster3 5	4.2	
SRR1791984	USA	SB0271	North America	Cluster 1	Cluster2 2	Cluster3 5	4.2	
SRR1791994	USA	SB0271	North America	Cluster 1	Cluster2 2	Cluster3 5	4.2	
SRR1791995	USA	SB0271	North America	Cluster 1	Cluster2 2	Cluster3 5	4.2	
SRR1791996	USA	SB0271	North America	Cluster 1	Cluster2 2	Cluster3 5	4.2	
SRR1791997	USA	SB0271	North America	Cluster 1	Cluster2 2	Cluster3 5	4.2	
SRR1791998	Mexico	SB0140	North America	Cluster 1	Cluster2 2	Cluster3 4	4.2	
SRR1792006	USA	SB0271	North America	Cluster 1	Cluster2 2	Cluster3 5	4.2	
SRR1792007	USA	SB0271	North America	Cluster 1	Cluster2 2	Cluster3 5	4.2	
SRR1792009	USA	SB0271	North America	Cluster 1	Cluster2 2	Cluster3 5	4.2	
SRR1792010	USA	SB0271	North America	Cluster 1	Cluster2 2	Cluster3 5	4.2	
SRR1792017	UNKNOWN	SB1757	UNKNOWN	Cluster 1	Cluster2 2	Cluster3 3	4.2	
SRR1792020	Mexico	SB0140	North America	Cluster 1	Cluster2 2	Cluster3 3	4.2	
SRR1792036	Mexico	SB0971	North America	Cluster 1	Cluster2 2	Cluster3 3	4.2	
SRR1792038	USA	SB0271	North America	Cluster 1	Cluster2 2	Cluster3 5	4.2	
SRR1792043	USA	SB0271	North America	Cluster 1	Cluster2 2	Cluster3 5	4.2	
SRR1792044	USA	SB0271	North America	Cluster 1	Cluster2 2	Cluster3 5	4.2	
SRR1792048	USA	SB0271	North America	Cluster 1	Cluster2 2	Cluster3 5	4.2	
SRR1792049	USA	SB0271	North America	Cluster 1	Cluster2 2	Cluster3 5	4.2	
SRR1792051	USA	SB0140	North America	Cluster 1	Cluster2 2	Cluster3 4	4.2	
SRR1792060	USA	SB0140	North America	Cluster 1	Cluster2 2	Cluster3 4	4.2	
SRR1792061	USA	SB0271	North America	Cluster 1	Cluster2 2	Cluster3 5	4.2	
SRR1792062	USA	SB0271	North America	Cluster 1	Cluster2 2	Cluster3 5	4.2	
SRR1792066	Mexico	SB0140	North America	Cluster 1	Cluster2 2	Cluster3 3	4.2	
SRR1792067	USA	SB0271	North America	Cluster 1	Cluster2 2	Cluster3 5	4.2	
SRR1792070	USA	SB0271	North America	Cluster 1	Cluster2 2	Cluster3 5	4.2	
SRR1792072	Mexico	SB0140	North America	Cluster 1	Cluster2 2	Cluster3 3	4.2	
SRR1792076	USA	SB0271	North America	Cluster 1	Cluster2 2	Cluster3 5	4.2	
SRR1792090	USA	SB0271	North America	Cluster 1	Cluster2 2	Cluster3 5	4.2	
SRR1792094	USA	SB0271	North America	Cluster 1	Cluster2 2	Cluster3 5	4.2	

SRR1792095	USA	SB0271	North America	Cluster 1	Cluster2 2	Cluster3 5	4.2	
SRR1792097	USA	SB0271	North America	Cluster 1	Cluster2 2	Cluster3 5	4.2	
SRR1792100	USA	SB0271	North America	Cluster 1	Cluster2 2	Cluster3 5	4.2	
SRR1792102	USA	SB0271	North America	Cluster 1	Cluster2 2	Cluster3 5	4.2	
SRR1792103	USA	SB0271	North America	Cluster 1	Cluster2 2	Cluster3 5	4.2	
SRR1792104	USA	SB0271	North America	Cluster 1	Cluster2 2	Cluster3 5	4.2	
SRR1792105	USA	SB0271	North America	Cluster 1	Cluster2 2	Cluster3 5	4.2	
SRR1792114	USA	SB0271	North America	Cluster 1	Cluster2 2	Cluster3 5	4.2	
SRR1792115	USA	SB0271	North America	Cluster 1	Cluster2 2	Cluster3 5	4.2	
SRR1792116	USA	SB0271	North America	Cluster 1	Cluster2 2	Cluster3 5	4.2	
SRR1792119	USA	SB0271	North America	Cluster 1	Cluster2 2	Cluster3 5	4.2	
SRR1792121	USA	SB0271	North America	Cluster 1	Cluster2 2	Cluster3 5	4.2	
SRR1792123	USA	SB0271	North America	Cluster 1	Cluster2 2	Cluster3 5	4.2	
SRR1792124	USA	SB0271	North America	Cluster 1	Cluster2 2	Cluster3 5	4.2	
SRR1792125	UNKNOWN	SB0971	UNKNOWN	Cluster 1	Cluster2 2	Cluster3 3	4.2	
SRR1792126	UNKNOWN	SB0971	UNKNOWN	Cluster 1	Cluster2 2	Cluster3 3	4.2	
SRR1792127	UNKNOWN	SB0971	UNKNOWN	Cluster 1	Cluster2 2	Cluster3 3	4.2	
SRR1792129	UNKNOWN	SB0971	UNKNOWN	Cluster 1	Cluster2 2	Cluster3 3	4.2	
SRR1792130	UNKNOWN	SB0971	UNKNOWN	Cluster 1	Cluster2 2	Cluster3 3	4.2	
SRR1792131	UNKNOWN	SB0971	UNKNOWN	Cluster 1	Cluster2 2	Cluster3 3	4.2	
SRR1792133	UNKNOWN	SB0971	UNKNOWN	Cluster 1	Cluster2 2	Cluster3 3	4.2	
SRR1792134	UNKNOWN	SB0971	UNKNOWN	Cluster 1	Cluster2 2	Cluster3 3	4.2	
SRR1792135	UNKNOWN	SB0971	UNKNOWN	Cluster 1	Cluster2 2	Cluster3 3	4.2	
SRR1792136	UNKNOWN	SB0971	UNKNOWN	Cluster 1	Cluster2 2	Cluster3 3	4.2	
SRR1792137	UNKNOWN	SB0971	UNKNOWN	Cluster 1	Cluster2 2	Cluster3 3	4.2	
SRR1792138	UNKNOWN	SB0971	UNKNOWN	Cluster 1	Cluster2 2	Cluster3 3	4.2	
SRR1792139	UNKNOWN	SB0971	UNKNOWN	Cluster 1	Cluster2 2	Cluster3 3	4.2	
SRR1792142	USA	SB0271	North America	Cluster 1	Cluster2 2	Cluster3 5	4.2	
SRR1792143	USA	SB0271	North America	Cluster 1	Cluster2 2	Cluster3 5	4.2	
SRR1792144	USA	SB0271	North America	Cluster 1	Cluster2 2	Cluster3 5	4.2	
SRR1792145	USA	SB0271	North America	Cluster 1	Cluster2 2	Cluster3 5	4.2	
SRR1792149	USA	SB0271	North America	Cluster 1	Cluster2 2	Cluster3 5	4.2	
SRR1792153	UNKNOWN	SB0971	UNKNOWN	Cluster 1	Cluster2 2	Cluster3 3	4.2	
SRR1792154	UNKNOWN	SB0971	UNKNOWN	Cluster 1	Cluster2 2	Cluster3 3	4.2	
SRR1792166	USA	SB0271	North America	Cluster 1	Cluster2 2	Cluster3 5	4.2	
SRR1792167	USA	SB0271	North America	Cluster 1	Cluster2 2	Cluster3 5	4.2	
SRR1792168	USA	SB0271	North America	Cluster 1	Cluster2 2	Cluster3 5	4.2	
SRR1792172	Mexico	SB1348	North America	Cluster 1	Cluster2 2	Cluster3 3	4.2	
SRR1792177	USA	SB0271	North America	Cluster 1	Cluster2 2	Cluster3 5	4.2	
SRR1792178	USA	SB0271	North America	Cluster 1	Cluster2 2	Cluster3 5	4.2	
SRR1792179	USA	SB0271	North America	Cluster 1	Cluster2 2	Cluster3 5	4.2	
SRR1792180	USA	SB0271	North America	Cluster 1	Cluster2 2	Cluster3 5	4.2	
SRR1792181	USA	SB0271	North America	Cluster 1	Cluster2 2	Cluster3 5	4.2	
SRR1792182	USA	SB0271	North America	Cluster 1	Cluster2 2	Cluster3 5	4.2	
SRR1792183	USA	SB0271	North America	Cluster 1	Cluster2 2	Cluster3 5	4.2	
SRR1792191	USA	SB0271	North America	Cluster 1	Cluster2 2	Cluster3 5	4.2	
SRR1792192	USA	SB0271	North America	Cluster 1	Cluster2 2	Cluster3 5	4.2	
SRR1792194	USA	SB0971	North America	Cluster 1	Cluster2 2	Cluster3 3	4.2	
SRR1792195	USA	SB0971	North America	Cluster 1	Cluster2 2	Cluster3 3	4.2	
SRR1792196	USA	SB0971	North America	Cluster 1	Cluster2 2	Cluster3 3	4.2	
SRR1792198	USA	SB0971	North America	Cluster 1	Cluster2 2	Cluster3 3	4.2	

SRR1792199	USA	SB0271	North America	Cluster 1	Cluster2 2	Cluster3 5	4.2	
SRR1792231	USA	SB0271	North America	Cluster 1	Cluster2 2	Cluster3 5	4.2	
SRR1792244	USA	SB0271	North America	Cluster 1	Cluster2 2	Cluster3 5	4.2	
SRR1792269	UNKNOWN	SB2014	UNKNOWN	Cluster 1	Cluster2 2	Cluster3 3	4.2	
SRR1792277	UNKNOWN	SB0971	UNKNOWN	Cluster 1	Cluster2 2	Cluster3 3	4.2	
SRR1792280	UNKNOWN	SB0140	UNKNOWN	Cluster 1	Cluster2 2	Cluster3 5	4.2	
SRR1792281	Mexico	SB0140	North America	Cluster 1	Cluster2 2	Cluster3 5	4.2	
SRR1792283	USA	SB0971	North America	Cluster 1	Cluster2 2	Cluster3 3	4.2	
SRR1792305	USA	SB0971	North America	Cluster 1	Cluster2 2	Cluster3 3	4.2	
SRR1792307	Mexico	SB0971	North America	Cluster 1	Cluster2 2	Cluster3 3	4.2	
SRR1792328	Mexico	SB0971	North America	Cluster 1	Cluster2 2	Cluster3 3	4.2	
SRR1792334	USA	SB0271	North America	Cluster 1	Cluster2 2	Cluster3 5	4.2	
SRR1792338	USA	SB0271	North America	Cluster 1	Cluster2 2	Cluster3 5	4.2	
SRR1792339	USA	SB0271	North America	Cluster 1	Cluster2 2	Cluster3 5	4.2	
SRR1792340	USA	SB0271	North America	Cluster 1	Cluster2 2	Cluster3 5	4.2	
SRR1792341	USA	SB0271	North America	Cluster 1	Cluster2 2	Cluster3 5	4.2	
SRR1792358	Mexico	SB0140	North America	Cluster 1	Cluster2 2	Cluster3 4	4.2	
SRR1792429	USA	SB0271	North America	Cluster 1	Cluster2 2	Cluster3 5	4.2	
SRR1792435	UNKNOWN	SB0140	UNKNOWN	Cluster 1	Cluster2 2	Cluster3 4	4.2	
SRR1792447	Mexico	Not_found	North America	Cluster 1	Cluster2 2	Cluster3 3	4.2	
SRR5216711	New_Zealand	SB0140	Oceania	Cluster 1	Cluster2 2	Cluster3 4	4.2	
SRR5216719	New_Zealand	SB0140	Oceania	Cluster 1	Cluster2 2	Cluster3 4	4.2	
SRR5216758	New_Zealand	SB0140	Oceania	Cluster 1	Cluster2 2	Cluster3 4	4.2	
SRR5216809	New_Zealand	SB0140	Oceania	Cluster 1	Cluster2 2	Cluster3 4	4.2	
SRR5216841	New_Zealand	SB0140	Oceania	Cluster 1	Cluster2 2	Cluster3 4	4.2	
SRR5216850	New_Zealand	SB0484	Oceania	Cluster 1	Cluster2 2	Cluster3 4	4.2	
SRR5216855	New_Zealand	SB0140	Oceania	Cluster 1	Cluster2 2	Cluster3 4	4.2	
SRR5216910	New_Zealand	SB0980	Oceania	Cluster 1	Cluster2 2	Cluster3 4	4.2	
SRR5216937	New_Zealand	SB0140	Oceania	Cluster 1	Cluster2 2	Cluster3 4	4.2	
SRR5216954	New_Zealand	SB0980	Oceania	Cluster 1	Cluster2 2	Cluster3 4	4.2	
SRR5216958	New_Zealand	SB0140	Oceania	Cluster 1	Cluster2 2	Cluster3 4	4.2	
SRR5216978	New_Zealand	SB0140	Oceania	Cluster 1	Cluster2 2	Cluster3 4	4.2	
SRR5216979	New_Zealand	SB0140	Oceania	Cluster 1	Cluster2 2	Cluster3 4	4.2	
SRR1791695	USA	SB0673	North America	Cluster 1	Cluster2 1	Cluster3 1	4.3	
SRR1791701	UNKNOWN	SB0673	North America	Cluster 1	Cluster2 1	Cluster3 1	4.3	
SRR1791702	Mexico	SB0673	North America	Cluster 1	Cluster2 1	Cluster3 1	4.3	
SRR1791703	Mexico	SB0673	North America	Cluster 1	Cluster2 1	Cluster3 1	4.3	
SRR1791704	Mexico	SB0673	North America	Cluster 1	Cluster2 1	Cluster3 2	4.3	
SRR1791706	Mexico	SB0673	North America	Cluster 1	Cluster2 1	Cluster3 2	4.3	
SRR1791709	Mexico	SB0673	North America	Cluster 1	Cluster2 1	Cluster3 1	4.3	
SRR1791713	UNKNOWN	SB0673	UNKNOWN	Cluster 1	Cluster2 1	Cluster3 1	4.3	
SRR1791717	UNKNOWN	Not_found	UNKNOWN	Cluster 1	Cluster2 1	Cluster3 2	4.3	
SRR1791720	Mexico	SB0673	North America	Cluster 1	Cluster2 1	Cluster3 2	4.3	
SRR1791724	Mexico	SB0673	North America	Cluster 1	Cluster2 1	Cluster3 2	4.3	
SRR1791725	Mexico	SB0673	North America	Cluster 1	Cluster2 1	Cluster3 1	4.3	
SRR1791726	UNKNOWN	SB0673	UNKNOWN	Cluster 1	Cluster2 1	Cluster3 1	4.3	
SRR1791729	Mexico	SB0673	North America	Cluster 1	Cluster2 1	Cluster3 2	4.3	
SRR1791732	Mexico	Not_found	North America	Cluster 1	Cluster2 1	Cluster3 1	4.3	
SRR1791734	USA	SB0673	North America	Cluster 1	Cluster2 1	Cluster3 2	4.3	
SRR1791735	Mexico	SB0673	North America	Cluster 1	Cluster2 1	Cluster3 2	4.3	
SRR1791741	UNKNOWN	SB0673	UNKNOWN	Cluster 1	Cluster2 1	Cluster3 2	4.3	

SRR1791742	Mexico	SB1750	North America	Cluster 1	Cluster2 1	Cluster3 2	4.3	
SRR1791743	Mexico	SB0673	North America	Cluster 1	Cluster2 1	Cluster3 2	4.3	
SRR1791744	Mexico	SB0673	North America	Cluster 1	Cluster2 1	Cluster3 1	4.3	
SRR1791748	Mexico	SB0673	North America	Cluster 1	Cluster2 1	Cluster3 2	4.3	
SRR1791749	Mexico	SB0673	North America	Cluster 1	Cluster2 1	Cluster3 2	4.3	
SRR1791750	Mexico	SB0673	North America	Cluster 1	Cluster2 1	Cluster3 2	4.3	
SRR1791754	USA	Not_found	North America	Cluster 1	Cluster2 1	Cluster3 1	4.3	
SRR1791755	Mexico	SB0673	North America	Cluster 1	Cluster2 1	Cluster3 2	4.3	
SRR1791758	Mexico	SB0673	North America	Cluster 1	Cluster2 1	Cluster3 1	4.3	
SRR1791761	Mexico	SB0673	North America	Cluster 1	Cluster2 1	Cluster3 1	4.3	
SRR1791767	Mexico	SB0986	North America	Cluster 1	Cluster2 1	Cluster3 2	4.3	
SRR1791769	USA	SB0673	North America	Cluster 1	Cluster2 1	Cluster3 2	4.3	
SRR1791774	UNKNOWN	SB1812	UNKNOWN	Cluster 1	Cluster2 1	Cluster3 2	4.3	
SRR1791778	Mexico	SB0673	North America	Cluster 1	Cluster2 1	Cluster3 1	4.3	
SRR1791781	Mexico	SB0673	North America	Cluster 1	Cluster2 1	Cluster3 2	4.3	
SRR1791782	Mexico	SB0673	North America	Cluster 1	Cluster2 1	Cluster3 2	4.3	
SRR1791783	Mexico	SB0673	North America	Cluster 1	Cluster2 1	Cluster3 2	4.3	
SRR1791786	Mexico	SB0673	North America	Cluster 1	Cluster2 1	Cluster3 2	4.3	
SRR1791787	Mexico	Not_found	North America	Cluster 1	Cluster2 1	Cluster3 2	4.3	
SRR1791789	UNKNOWN	SB0673	UNKNOWN	Cluster 1	Cluster2 1	Cluster3 2	4.3	
SRR1791790	Mexico	Not_found	North America	Cluster 1	Cluster2 1	Cluster3 1	4.3	
SRR1791791	Mexico	Not_found	North America	Cluster 1	Cluster2 1	Cluster3 2	4.3	
SRR1791794	USA	SB0673	North America	Cluster 1	Cluster2 1	Cluster3 1	4.3	
SRR1791796	Mexico	SB0673	North America	Cluster 1	Cluster2 1	Cluster3 2	4.3	
SRR1791798	Mexico	Not_found	North America	Cluster 1	Cluster2 1	Cluster3 1	4.3	
SRR1791806	Mexico	SB0673	North America	Cluster 1	Cluster2 1	Cluster3 2	4.3	
SRR1791808	USA	SB0673	North America	Cluster 1	Cluster2 1	Cluster3 1	4.3	
SRR1791811	Mexico	SB0673	North America	Cluster 1	Cluster2 1	Cluster3 2	4.3	
SRR1791813	USA	SB0673	North America	Cluster 1	Cluster2 1	Cluster3 1	4.3	
SRR1791815	UNKNOWN	SB0673	UNKNOWN	Cluster 1	Cluster2 1	Cluster3 1	4.3	
SRR1791827	UNKNOWN	SB0673	UNKNOWN	Cluster 1	Cluster2 1	Cluster3 2	4.3	
SRR1791831	Mexico	SB0673	North America	Cluster 1	Cluster2 1	Cluster3 2	4.3	
SRR1791835	Mexico	SB0673	North America	Cluster 1	Cluster2 1	Cluster3 2	4.3	
SRR1791836	Mexico	SB0673	North America	Cluster 1	Cluster2 1	Cluster3 2	4.3	
SRR1791837	Mexico	SB0673	North America	Cluster 1	Cluster2 1	Cluster3 2	4.3	
SRR1791838	USA	SB0673	North America	Cluster 1	Cluster2 1	Cluster3 2	4.3	
SRR1791840	USA	SB0673	North America	Cluster 1	Cluster2 1	Cluster3 1	4.3	
SRR1791842	UNKNOWN	Not_found	UNKNOWN	Cluster 1	Cluster2 1	Cluster3 2	4.3	
SRR1791843	Mexico	SB0673	North America	Cluster 1	Cluster2 1	Cluster3 2	4.3	
SRR1791846	Mexico	Not_found	North America	Cluster 1	Cluster2 1	Cluster3 2	4.3	
SRR1791848	UNKNOWN	Not_found	UNKNOWN	Cluster 1	Cluster2 1	Cluster3 2	4.3	
SRR1791849	USA	SB0673	North America	Cluster 1	Cluster2 1	Cluster3 1	4.3	
SRR1791850	Mexico	Not_found	North America	Cluster 1	Cluster2 1	Cluster3 2	4.3	
SRR1791851	Mexico	SB0673	North America	Cluster 1	Cluster2 1	Cluster3 2	4.3	
SRR1791852	UNKNOWN	SB0673	UNKNOWN	Cluster 1	Cluster2 1	Cluster3 2	4.3	
SRR1791853	Mexico	SB1812	North America	Cluster 1	Cluster2 1	Cluster3 2	4.3	
SRR1791855	UNKNOWN	SB1750	UNKNOWN	Cluster 1	Cluster2 1	Cluster3 2	4.3	
SRR1791856	USA	SB0673	North America	Cluster 1	Cluster2 1	Cluster3 1	4.3	
SRR1791877	Mexico	SB0673	North America	Cluster 1	Cluster2 1	Cluster3 2	4.3	
SRR1791879	Mexico	SB0484	North America	Cluster 1	Cluster2 1	Cluster3 2	4.3	
SRR1791886	Mexico	SB0673	North America	Cluster 1	Cluster2 1	Cluster3 2	4.3	

SRR1791887	Mexico	SB0673	North America	Cluster 1	Cluster2 1	Cluster3 2	4.3	
SRR1791890	Mexico	SB0673	North America	Cluster 1	Cluster2 1	Cluster3 2	4.3	
SRR1791892	Mexico	SB0673	North America	Cluster 1	Cluster2 1	Cluster3 2	4.3	
SRR1791893	Mexico	SB0673	North America	Cluster 1	Cluster2 1	Cluster3 2	4.3	
SRR1791895	Mexico	SB0673	North America	Cluster 1	Cluster2 1	Cluster3 1	4.3	
SRR1791896	Mexico	SB0673	North America	Cluster 1	Cluster2 1	Cluster3 1	4.3	
SRR1791897	Mexico	SB0673	North America	Cluster 1	Cluster2 1	Cluster3 1	4.3	
SRR1791898	Mexico	SB0673	North America	Cluster 1	Cluster2 1	Cluster3 1	4.3	
SRR1791900	Mexico	SB0673	North America	Cluster 1	Cluster2 1	Cluster3 2	4.3	
SRR1791903	Mexico	SB0673	North America	Cluster 1	Cluster2 1	Cluster3 2	4.3	
SRR1791904	Mexico	SB0673	North America	Cluster 1	Cluster2 1	Cluster3 1	4.3	
SRR1791905	USA	SB0673	North America	Cluster 1	Cluster2 1	Cluster3 2	4.3	
SRR1791906	USA	SB0673	North America	Cluster 1	Cluster2 1	Cluster3 2	4.3	
SRR1791914	USA	SB0673	North America	Cluster 1	Cluster2 1	Cluster3 2	4.3	
SRR1791915	Mexico	SB0673	North America	Cluster 1	Cluster2 1	Cluster3 2	4.3	
SRR1791916	Mexico	SB0673	North America	Cluster 1	Cluster2 1	Cluster3 1	4.3	
SRR1791917	Mexico	SB0673	North America	Cluster 1	Cluster2 1	Cluster3 2	4.3	
SRR1791919	USA	SB0673	North America	Cluster 1	Cluster2 1	Cluster3 2	4.3	
SRR1791920	USA	SB0673	North America	Cluster 1	Cluster2 1	Cluster3 2	4.3	
SRR1791922	USA	SB0673	North America	Cluster 1	Cluster2 1	Cluster3 2	4.3	
SRR1791924	Mexico	SB0673	North America	Cluster 1	Cluster2 1	Cluster3 2	4.3	
SRR1791934	USA	SB0673	North America	Cluster 1	Cluster2 1	Cluster3 2	4.3	
SRR1791936	UNKNOWN	SB1745	UNKNOWN	Cluster 1	Cluster2 1	Cluster3 2	4.3	
SRR1791943	Mexico	SB0673	North America	Cluster 1	Cluster2 1	Cluster3 2	4.3	
SRR1791944	USA	SB0673	North America	Cluster 1	Cluster2 1	Cluster3 2	4.3	
SRR1791948	UNKNOWN	SB0673	UNKNOWN	Cluster 1	Cluster2 1	Cluster3 2	4.3	
SRR1791953	Mexico	SB0673	North America	Cluster 1	Cluster2 1	Cluster3 1	4.3	
SRR1791961	UNKNOWN	SB0673	UNKNOWN	Cluster 1	Cluster2 1	Cluster3 2	4.3	
SRR1791962	Mexico	SB0673	North America	Cluster 1	Cluster2 1	Cluster3 2	4.3	
SRR1791966	Mexico	SB0673	North America	Cluster 1	Cluster2 1	Cluster3 2	4.3	
SRR1791973	Mexico	SB0673	North America	Cluster 1	Cluster2 1	Cluster3 1	4.3	
SRR1791989	Mexico	SB0673	North America	Cluster 1	Cluster2 1	Cluster3 2	4.3	
SRR1791990	Mexico	SB0673	North America	Cluster 1	Cluster2 1	Cluster3 1	4.3	
SRR1791991	Mexico	SB0673	North America	Cluster 1	Cluster2 1	Cluster3 1	4.3	
SRR1791992	Mexico	SB0673	North America	Cluster 1	Cluster2 1	Cluster3 1	4.3	
SRR1791993	Mexico	SB0673	North America	Cluster 1	Cluster2 1	Cluster3 1	4.3	
SRR1792013	Mexico	SB0673	North America	Cluster 1	Cluster2 1	Cluster3 2	4.3	
SRR1792016	Mexico	SB0673	North America	Cluster 1	Cluster2 1	Cluster3 2	4.3	
SRR1792026	Mexico	SB1750	North America	Cluster 1	Cluster2 1	Cluster3 2	4.3	
SRR1792029	Mexico	SB0673	North America	Cluster 1	Cluster2 1	Cluster3 2	4.3	
SRR1792030	Mexico	SB0673	North America	Cluster 1	Cluster2 1	Cluster3 2	4.3	
SRR1792031	Mexico	SB0673	North America	Cluster 1	Cluster2 1	Cluster3 2	4.3	
SRR1792033	Mexico	SB0673	North America	Cluster 1	Cluster2 1	Cluster3 1	4.3	
SRR1792035	Mexico	SB0673	North America	Cluster 1	Cluster2 1	Cluster3 2	4.3	
SRR1792039	Mexico	SB0986	North America	Cluster 1	Cluster2 1	Cluster3 2	4.3	
SRR1792040	Mexico	SB0673	North America	Cluster 1	Cluster2 1	Cluster3 2	4.3	
SRR1792059	USA	SB0673	North America	Cluster 1	Cluster2 1	Cluster3 2	4.3	
SRR1792074	Mexico	SB0673	North America	Cluster 1	Cluster2 1	Cluster3 2	4.3	
SRR1792075	USA	SB0673	North America	Cluster 1	Cluster2 1	Cluster3 2	4.3	
SRR1792079	USA	SB0673	North America	Cluster 1	Cluster2 1	Cluster3 2	4.3	
SRR1792080	USA	SB0673	North America	Cluster 1	Cluster2 1	Cluster3 2	4.3	

SRR1792081	USA	SB0673	North America	Cluster 1	Cluster2 1	Cluster3 2	4.3	
SRR1792082	USA	SB0673	North America	Cluster 1	Cluster2 1	Cluster3 2	4.3	
SRR1792083	USA	SB0673	North America	Cluster 1	Cluster2 1	Cluster3 2	4.3	
SRR1792084	USA	SB0673	North America	Cluster 1	Cluster2 1	Cluster3 2	4.3	
SRR1792085	USA	SB0673	North America	Cluster 1	Cluster2 1	Cluster3 2	4.3	
SRR1792086	USA	SB0673	North America	Cluster 1	Cluster2 1	Cluster3 2	4.3	
SRR1792087	USA	SB0673	North America	Cluster 1	Cluster2 1	Cluster3 2	4.3	
SRR1792088	Mexico	SB0673	North America	Cluster 1	Cluster2 1	Cluster3 2	4.3	
SRR1792091	USA	SB0673	North America	Cluster 1	Cluster2 1	Cluster3 2	4.3	
SRR1792092	USA	SB0673	North America	Cluster 1	Cluster2 1	Cluster3 2	4.3	
SRR1792093	USA	SB0673	North America	Cluster 1	Cluster2 1	Cluster3 2	4.3	
SRR1792099	Mexico	SB0673	North America	Cluster 1	Cluster2 1	Cluster3 1	4.3	
SRR1792106	USA	SB0673	North America	Cluster 1	Cluster2 1	Cluster3 2	4.3	
SRR1792113	Mexico	SB0673	North America	Cluster 1	Cluster2 1	Cluster3 1	4.3	
SRR1792148	UNKNOWN	SB0673	UNKNOWN	Cluster 1	Cluster2 1	Cluster3 2	4.3	
SRR1792150	Mexico	SB0673	North America	Cluster 1	Cluster2 1	Cluster3 2	4.3	
SRR1792151	Mexico	Not_found	North America	Cluster 1	Cluster2 1	Cluster3 2	4.3	
SRR1792158	Mexico	SB0673	North America	Cluster 1	Cluster2 1	Cluster3 2	4.3	
SRR1792169	USA	SB0673	North America	Cluster 1	Cluster2 1	Cluster3 2	4.3	
SRR1792170	UNKNOWN	SB0673	UNKNOWN	Cluster 1	Cluster2 1	Cluster3 1	4.3	
SRR1792174	UNKNOWN	SB0673	UNKNOWN	Cluster 1	Cluster2 1	Cluster3 1	4.3	
SRR1792202	UNKNOWN	SB0673	UNKNOWN	Cluster 1	Cluster2 1	Cluster3 2	4.3	
SRR1792248	UNKNOWN	SB0673	UNKNOWN	Cluster 1	Cluster2 1	Cluster3 1	4.3	
SRR1792250	USA	SB0673	North America	Cluster 1	Cluster2 1	Cluster3 1	4.3	
SRR1792287	UNKNOWN	SB0673	UNKNOWN	Cluster 1	Cluster2 1	Cluster3 2	4.3	
SRR1792288	USA	SB0673	North America	Cluster 1	Cluster2 1	Cluster3 2	4.3	
SRR1792296	USA	SB0673	North America	Cluster 1	Cluster2 1	Cluster3 2	4.3	
SRR1792309	Mexico	SB0673	North America	Cluster 1	Cluster2 1	Cluster3 1	4.3	
SRR1792313	Mexico	SB0673	North America	Cluster 1	Cluster2 1	Cluster3 2	4.3	
SRR1792319	USA	SB0673	North America	Cluster 1	Cluster2 1	Cluster3 2	4.3	
SRR1792321	USA	SB0673	North America	Cluster 1	Cluster2 1	Cluster3 2	4.3	
SRR1792322	USA	SB0673	North America	Cluster 1	Cluster2 1	Cluster3 2	4.3	
SRR1792333	Mexico	SB0673	North America	Cluster 1	Cluster2 1	Cluster3 2	4.3	
SRR1792336	UNKNOWN	SB0673	UNKNOWN	Cluster 1	Cluster2 1	Cluster3 1	4.3	
SRR1792345	Mexico	SB0673	North America	Cluster 1	Cluster2 1	Cluster3 2	4.3	
SRR1792348	USA	SB0673	North America	Cluster 1	Cluster2 1	Cluster3 1	4.3	
SRR1792349	Mexico	SB0673	North America	Cluster 1	Cluster2 1	Cluster3 1	4.3	
SRR1792353	Mexico	SB1750	North America	Cluster 1	Cluster2 1	Cluster3 2	4.3	
SRR1792354	Mexico	SB0673	North America	Cluster 1	Cluster2 1	Cluster3 2	4.3	
SRR1792357	Mexico	SB0673	North America	Cluster 1	Cluster2 1	Cluster3 1	4.3	
SRR1792359	USA	SB0673	North America	Cluster 1	Cluster2 1	Cluster3 1	4.3	
SRR1792360	USA	SB0673	North America	Cluster 1	Cluster2 1	Cluster3 1	4.3	
SRR1792361	USA	SB0673	North America	Cluster 1	Cluster2 1	Cluster3 1	4.3	
SRR1792362	USA	SB0673	North America	Cluster 1	Cluster2 1	Cluster3 1	4.3	
SRR1792364	USA	SB0673	North America	Cluster 1	Cluster2 1	Cluster3 1	4.3	
SRR1792365	USA	SB0673	North America	Cluster 1	Cluster2 1	Cluster3 1	4.3	
SRR1792366	USA	SB0673	North America	Cluster 1	Cluster2 1	Cluster3 1	4.3	
SRR1792367	USA	SB0673	North America	Cluster 1	Cluster2 1	Cluster3 1	4.3	
SRR1792368	USA	SB0673	North America	Cluster 1	Cluster2 1	Cluster3 1	4.3	
SRR1792369	Mexico	SB0673	North America	Cluster 1	Cluster2 1	Cluster3 1	4.3	
SRR1792371	USA	SB0673	North America	Cluster 1	Cluster2 1	Cluster3 1	4.3	

SRR1792384	Mexico	SB0673	North America	Cluster 1	Cluster2 1	Cluster3 2	4.3	
SRR1792413	USA	SB0673	North America	Cluster 1	Cluster2 1	Cluster3 1	4.3	
SRR1792414	USA	SB0673	North America	Cluster 1	Cluster2 1	Cluster3 1	4.3	
SRR1792415	USA	SB0673	North America	Cluster 1	Cluster2 1	Cluster3 1	4.3	
SRR1792416	USA	SB0673	North America	Cluster 1	Cluster2 1	Cluster3 1	4.3	
SRR1792417	USA	SB0673	North America	Cluster 1	Cluster2 1	Cluster3 1	4.3	
SRR1792421	UNKNOWN	SB0673	UNKNOWN	Cluster 1	Cluster2 1	Cluster3 1	4.3	
SRR1792422	Mexico	Not_found	North America	Cluster 1	Cluster2 1	Cluster3 2	4.3	
SRR1792446	USA	SB0673	North America	Cluster 1	Cluster2 1	Cluster3 2	4.3	
SRR1792450	USA	SB0673	North America	Cluster 1	Cluster2 1	Cluster3 2	4.3	
SRR1792489	USA	SB0673	North America	Cluster 1	Cluster2 1	Cluster3 1	4.3	
SRR1792504	USA	SB0673	North America	Cluster 1	Cluster2 1	Cluster3 1	4.3	
SRR1792506	USA	SB0673	North America	Cluster 1	Cluster2 1	Cluster3 1	4.3	
AF2122/97 (no SRA)	United Kingdom	SB0140	Europe	Cluster 5	Cluster2 17	Cluster3 44	5.1	Reference strain
ERR841874	Northern_Ireland	SB0140	Europe	Cluster 5	Cluster2 17	Cluster3 44	5.1	
SRR1657068	Panama	SB1041	Latin America	Cluster 5	Cluster2 17	Cluster3 45	5.1	
SRR1791737	Mexico	SB0140	North America	Cluster 5	Cluster2 17	Cluster3 44	5.1	
SRR1791740	Mexico	Not_found	North America	Cluster 5	Cluster2 17	Cluster3 44	5.1	
SRR1791883	Mexico	SB0140	North America	Cluster 5	Cluster2 17	Cluster3 44	5.1	
SRR1791970	Mexico	SB0140	North America	Cluster 5	Cluster2 17	Cluster3 44	5.1	
SRR1792078	UNKNOWN	SB0140	UNKNOWN	Cluster 5	Cluster2 17	Cluster3 44	5.1	
SRR1792247	USA	Not_found	North America	Cluster 5	Cluster2 17	Cluster3 44	5.1	
ERR1815543	South Africa	SB0140	Africa	Cluster 5	Cluster2 17	Cluster3 44	5.1	
SRR5216698	New_Zealand	SB0140	Oceania	Cluster 5	Cluster2 17	Cluster3 44	5.1	
SRR5216706	New_Zealand	SB0140	Oceania	Cluster 5	Cluster2 17	Cluster3 44	5.1	
SRR5216709	New_Zealand	SB0140	Oceania	Cluster 5	Cluster2 17	Cluster3 44	5.1	
SRR5216724	New_Zealand	SB0140	Oceania	Cluster 5	Cluster2 17	Cluster3 44	5.1	
SRR5216727	New_Zealand	SB0140	Oceania	Cluster 5	Cluster2 17	Cluster3 44	5.1	
SRR5216741	New_Zealand	SB0140	Oceania	Cluster 5	Cluster2 17	Cluster3 44	5.1	
SRR5216759	New_Zealand	SB0140	Oceania	Cluster 5	Cluster2 17	Cluster3 44	5.1	
SRR5216763	New_Zealand	SB0140	Oceania	Cluster 5	Cluster2 17	Cluster3 44	5.1	
SRR5216774	New_Zealand	SB0140	Oceania	Cluster 5	Cluster2 17	Cluster3 44	5.1	
SRR5216779	New_Zealand	SB0140	Oceania	Cluster 5	Cluster2 17	Cluster3 44	5.1	
SRR5216799	New_Zealand	SB0140	Oceania	Cluster 5	Cluster2 17	Cluster3 44	5.1	
SRR5216835	New_Zealand	SB0140	Oceania	Cluster 5	Cluster2 17	Cluster3 44	5.1	
SRR5216854	New_Zealand	SB0140	Oceania	Cluster 5	Cluster2 17	Cluster3 44	5.1	
SRR5216884	New_Zealand	SB0140	Oceania	Cluster 5	Cluster2 17	Cluster3 44	5.1	
SRR5216891	New_Zealand	SB0140	Oceania	Cluster 5	Cluster2 17	Cluster3 44	5.1	
SRR5216893	New_Zealand	SB0140	Oceania	Cluster 5	Cluster2 17	Cluster3 44	5.1	
SRR5216906	New_Zealand	SB0140	Oceania	Cluster 5	Cluster2 17	Cluster3 44	5.1	
SRR5216928	New_Zealand	SB0140	Oceania	Cluster 5	Cluster2 17	Cluster3 44	5.1	
SRR5216970	New_Zealand	SB0140	Oceania	Cluster 5	Cluster2 17	Cluster3 44	5.1	
SRR5216692	New_Zealand	SB0140	Oceania	Cluster 5	Cluster2 18	Cluster3 47	5.2	
SRR5216693	New_Zealand	SB0140	Oceania	Cluster 5	Cluster2 18	Cluster3 46	5.2	
SRR5216699	New_Zealand	SB0140	Oceania	Cluster 5	Cluster2 18	Cluster3 46	5.2	
SRR5216700	New_Zealand	SB0140	Oceania	Cluster 5	Cluster2 18	Cluster3 46	5.2	
SRR5216702	New_Zealand	SB0140	Oceania	Cluster 5	Cluster2 18	Cluster3 46	5.2	
SRR5216704	New_Zealand	SB0140	Oceania	Cluster 5	Cluster2 18	Cluster3 46	5.2	
SRR5216707	New_Zealand	SB0140	Oceania	Cluster 5	Cluster2 18	Cluster3 46	5.2	
SRR5216713	New_Zealand	SB0140	Oceania	Cluster 5	Cluster2 18	Cluster3 46	5.2	

SRR5216715	New_Zealand	SB0140	Oceania	Cluster 5	Cluster2 18	Cluster3 46	5.2	
SRR5216718	New_Zealand	SB0140	Oceania	Cluster 5	Cluster2 18	Cluster3 46	5.2	
SRR5216720	New_Zealand	SB0140	Oceania	Cluster 5	Cluster2 18	Cluster3 46	5.2	
SRR5216721	New_Zealand	SB0140	Oceania	Cluster 5	Cluster2 18	Cluster3 46	5.2	
SRR5216736	New_Zealand	SB0140	Oceania	Cluster 5	Cluster2 18	Cluster3 46	5.2	
SRR5216740	New_Zealand	SB0140	Oceania	Cluster 5	Cluster2 18	Cluster3 46	5.2	
SRR5216742	New_Zealand	SB0140	Oceania	Cluster 5	Cluster2 18	Cluster3 46	5.2	
SRR5216743	New_Zealand	SB0140	Oceania	Cluster 5	Cluster2 18	Cluster3 46	5.2	
SRR5216744	New_Zealand	SB0140	Oceania	Cluster 5	Cluster2 18	Cluster3 46	5.2	
SRR5216748	New_Zealand	SB0140	Oceania	Cluster 5	Cluster2 18	Cluster3 46	5.2	
SRR5216751	New_Zealand	SB0140	Oceania	Cluster 5	Cluster2 18	Cluster3 46	5.2	
SRR5216764	New_Zealand	SB0140	Oceania	Cluster 5	Cluster2 18	Cluster3 46	5.2	
SRR5216769	New_Zealand	SB0140	Oceania	Cluster 5	Cluster2 18	Cluster3 46	5.2	
SRR5216772	New_Zealand	SB0140	Oceania	Cluster 5	Cluster2 18	Cluster3 46	5.2	
SRR5216773	New_Zealand	SB0140	Oceania	Cluster 5	Cluster2 18	Cluster3 46	5.2	
SRR5216783	New_Zealand	SB0140	Oceania	Cluster 5	Cluster2 18	Cluster3 46	5.2	
SRR5216787	New_Zealand	SB0140	Oceania	Cluster 5	Cluster2 18	Cluster3 46	5.2	
SRR5216788	New_Zealand	SB0140	Oceania	Cluster 5	Cluster2 18	Cluster3 46	5.2	
SRR5216793	New_Zealand	SB0140	Oceania	Cluster 5	Cluster2 18	Cluster3 46	5.2	
SRR5216794	New_Zealand	SB0140	Oceania	Cluster 5	Cluster2 18	Cluster3 46	5.2	
SRR5216795	New_Zealand	SB0140	Oceania	Cluster 5	Cluster2 18	Cluster3 46	5.2	
SRR5216796	New_Zealand	SB0140	Oceania	Cluster 5	Cluster2 18	Cluster3 46	5.2	
SRR5216798	New_Zealand	SB0140	Oceania	Cluster 5	Cluster2 18	Cluster3 46	5.2	
SRR5216802	New_Zealand	SB0140	Oceania	Cluster 5	Cluster2 18	Cluster3 46	5.2	
SRR5216805	New_Zealand	SB0140	Oceania	Cluster 5	Cluster2 18	Cluster3 46	5.2	
SRR5216818	New_Zealand	SB0140	Oceania	Cluster 5	Cluster2 18	Cluster3 46	5.2	
SRR5216821	New_Zealand	SB0140	Oceania	Cluster 5	Cluster2 18	Cluster3 46	5.2	
SRR5216825	New_Zealand	SB0140	Oceania	Cluster 5	Cluster2 18	Cluster3 46	5.2	
SRR5216829	New_Zealand	SB0140	Oceania	Cluster 5	Cluster2 18	Cluster3 46	5.2	
SRR5216830	New_Zealand	SB0140	Oceania	Cluster 5	Cluster2 18	Cluster3 46	5.2	
SRR5216834	New_Zealand	SB0140	Oceania	Cluster 5	Cluster2 18	Cluster3 46	5.2	
SRR5216838	New_Zealand	SB0140	Oceania	Cluster 5	Cluster2 18	Cluster3 46	5.2	
SRR5216839	New_Zealand	SB0140	Oceania	Cluster 5	Cluster2 18	Cluster3 46	5.2	
SRR5216840	New_Zealand	SB0140	Oceania	Cluster 5	Cluster2 18	Cluster3 46	5.2	
SRR5216845	New_Zealand	SB0140	Oceania	Cluster 5	Cluster2 18	Cluster3 46	5.2	
SRR5216846	New_Zealand	SB0140	Oceania	Cluster 5	Cluster2 18	Cluster3 46	5.2	
SRR5216847	New_Zealand	SB0273	Oceania	Cluster 5	Cluster2 18	Cluster3 46	5.2	
SRR5216853	New_Zealand	SB0140	Oceania	Cluster 5	Cluster2 18	Cluster3 46	5.2	
SRR5216861	New_Zealand	SB0140	Oceania	Cluster 5	Cluster2 18	Cluster3 46	5.2	
SRR5216863	New_Zealand	SB0140	Oceania	Cluster 5	Cluster2 18	Cluster3 46	5.2	
SRR5216864	New_Zealand	SB0140	Oceania	Cluster 5	Cluster2 18	Cluster3 46	5.2	
SRR5216865	New_Zealand	SB0140	Oceania	Cluster 5	Cluster2 18	Cluster3 46	5.2	
SRR5216876	New_Zealand	SB0140	Oceania	Cluster 5	Cluster2 18	Cluster3 46	5.2	
SRR5216880	New_Zealand	SB0140	Oceania	Cluster 5	Cluster2 18	Cluster3 46	5.2	
SRR5216881	New_Zealand	SB0140	Oceania	Cluster 5	Cluster2 18	Cluster3 46	5.2	
SRR5216882	New_Zealand	SB0140	Oceania	Cluster 5	Cluster2 18	Cluster3 46	5.2	
SRR5216887	New_Zealand	SB0140	Oceania	Cluster 5	Cluster2 18	Cluster3 46	5.2	
SRR5216889	New_Zealand	SB0140	Oceania	Cluster 5	Cluster2 18	Cluster3 46	5.2	
SRR5216892	New_Zealand	SB0140	Oceania	Cluster 5	Cluster2 18	Cluster3 46	5.2	
SRR5216894	New_Zealand	SB0140	Oceania	Cluster 5	Cluster2 18	Cluster3 46	5.2	
SRR5216903	New_Zealand	SB0140	Oceania	Cluster 5	Cluster2 18	Cluster3 46	5.2	

SRR5216908	New_Zealand	SB0140	Oceania	Cluster 5	Cluster2 18	Cluster3 46	5.2	
SRR5216909	New_Zealand	SB0140	Oceania	Cluster 5	Cluster2 18	Cluster3 46	5.2	
SRR5216911	New_Zealand	SB0140	Oceania	Cluster 5	Cluster2 18	Cluster3 46	5.2	
SRR5216912	New_Zealand	SB0140	Oceania	Cluster 5	Cluster2 18	Cluster3 46	5.2	
SRR5216913	New_Zealand	SB0273	Oceania	Cluster 5	Cluster2 18	Cluster3 46	5.2	
SRR5216916	New_Zealand	SB0140	Oceania	Cluster 5	Cluster2 18	Cluster3 46	5.2	
SRR5216923	New_Zealand	SB0140	Oceania	Cluster 5	Cluster2 18	Cluster3 46	5.2	
SRR5216926	New_Zealand	SB0140	Oceania	Cluster 5	Cluster2 18	Cluster3 46	5.2	
SRR5216940	New_Zealand	SB0140	Oceania	Cluster 5	Cluster2 18	Cluster3 46	5.2	
SRR5216941	New_Zealand	SB0140	Oceania	Cluster 5	Cluster2 18	Cluster3 46	5.2	
SRR5216943	New_Zealand	SB0140	Oceania	Cluster 5	Cluster2 18	Cluster3 46	5.2	
SRR5216952	New_Zealand	SB0140	Oceania	Cluster 5	Cluster2 18	Cluster3 46	5.2	
SRR5216956	New_Zealand	SB0140	Oceania	Cluster 5	Cluster2 18	Cluster3 46	5.2	
SRR5216959	New_Zealand	SB0140	Oceania	Cluster 5	Cluster2 18	Cluster3 46	5.2	
SRR5216965	New_Zealand	SB0140	Oceania	Cluster 5	Cluster2 18	Cluster3 46	5.2	
SRR5216966	New_Zealand	SB0140	Oceania	Cluster 5	Cluster2 18	Cluster3 46	5.2	
SRR5216976	New_Zealand	SB0140	Oceania	Cluster 5	Cluster2 18	Cluster3 46	5.2	
SRR5216980	New_Zealand	SB0140	Oceania	Cluster 5	Cluster2 18	Cluster3 46	5.2	
SRR5216981	New_Zealand	SB0140	Oceania	Cluster 5	Cluster2 18	Cluster3 46	5.2	
SRR5216982	New_Zealand	SB0140	Oceania	Cluster 5	Cluster2 18	Cluster3 46	5.2	
SRR5216983	New_Zealand	SB0140	Oceania	Cluster 5	Cluster2 18	Cluster3 46	5.2	
ERR125598	Northern_Ireland	SB0140	Europe	Cluster 5	Cluster2 19	Cluster3 49	5.3	
ERR125599	Northern_Ireland	SB0140	Europe	Cluster 5	Cluster2 19	Cluster3 49	5.3	
ERR125600	Northern_Ireland	SB0140	Europe	Cluster 5	Cluster2 19	Cluster3 49	5.3	
ERR125601	Northern_Ireland	SB0140	Europe	Cluster 5	Cluster2 19	Cluster3 49	5.3	
ERR125602	Northern_Ireland	SB0140	Europe	Cluster 5	Cluster2 19	Cluster3 49	5.3	
ERR125603	Northern_Ireland	SB0140	Europe	Cluster 5	Cluster2 19	Cluster3 49	5.3	
ERR125604	Northern_Ireland	SB0140	Europe	Cluster 5	Cluster2 19	Cluster3 49	5.3	
ERR125605	Northern_Ireland	SB0140	Europe	Cluster 5	Cluster2 19	Cluster3 49	5.3	
ERR125606	Northern_Ireland	SB0140	Europe	Cluster 5	Cluster2 19	Cluster3 49	5.3	
ERR125607	Northern_Ireland	SB0140	Europe	Cluster 5	Cluster2 19	Cluster3 49	5.3	
ERR125608	Northern_Ireland	SB0140	Europe	Cluster 5	Cluster2 19	Cluster3 49	5.3	
ERR125609	Northern_Ireland	SB0140	Europe	Cluster 5	Cluster2 19	Cluster3 49	5.3	
ERR125610	Northern_Ireland	SB0140	Europe	Cluster 5	Cluster2 19	Cluster3 49	5.3	
ERR125611	Northern_Ireland	SB0140	Europe	Cluster 5	Cluster2 19	Cluster3 49	5.3	
ERR125612	Northern_Ireland	SB0140	Europe	Cluster 5	Cluster2 19	Cluster3 49	5.3	
ERR125613	Northern_Ireland	SB0140	Europe	Cluster 5	Cluster2 19	Cluster3 49	5.3	
ERR125614	Northern_Ireland	SB0140	Europe	Cluster 5	Cluster2 19	Cluster3 49	5.3	
ERR125615	Northern_Ireland	SB0140	Europe	Cluster 5	Cluster2 19	Cluster3 49	5.3	
ERR125616	Northern_Ireland	SB0140	Europe	Cluster 5	Cluster2 19	Cluster3 49	5.3	
ERR125617	Northern_Ireland	SB0140	Europe	Cluster 5	Cluster2 19	Cluster3 49	5.3	
ERR125618	Northern_Ireland	SB0140	Europe	Cluster 5	Cluster2 19	Cluster3 49	5.3	
ERR125619	Northern_Ireland	SB0140	Europe	Cluster 5	Cluster2 19	Cluster3 49	5.3	
ERR125620	Northern_Ireland	SB0140	Europe	Cluster 5	Cluster2 19	Cluster3 49	5.3	
ERR125621	Northern_Ireland	SB0140	Europe	Cluster 5	Cluster2 19	Cluster3 49	5.3	
ERR125622	Northern_Ireland	SB0140	Europe	Cluster 5	Cluster2 19	Cluster3 49	5.3	
ERR125623	Northern_Ireland	SB0140	Europe	Cluster 5	Cluster2 19	Cluster3 49	5.3	
ERR125624	Northern_Ireland	SB0140	Europe	Cluster 5	Cluster2 19	Cluster3 49	5.3	
ERR125625	Northern_Ireland	SB0140	Europe	Cluster 5	Cluster2 19	Cluster3 49	5.3	
ERR125626	Northern_Ireland	SB0140	Europe	Cluster 5	Cluster2 19	Cluster3 49	5.3	
ERR125627	Northern_Ireland	SB0140	Europe	Cluster 5	Cluster2 19	Cluster3 49	5.3	

ERR125628	Northern_Ireland	SB0140	Europe	Cluster 5	Cluster2 19	Cluster3 49	5.3	
ERR841787	Northern_Ireland	SB0140	Europe	Cluster 5	Cluster2 19	Cluster3 48	5.3	
ERR841788	Northern_Ireland	SB0140	Europe	Cluster 5	Cluster2 19	Cluster3 48	5.3	
ERR841789	Northern_Ireland	SB0140	Europe	Cluster 5	Cluster2 19	Cluster3 48	5.3	
ERR841790	Northern_Ireland	SB0140	Europe	Cluster 5	Cluster2 19	Cluster3 48	5.3	
ERR841791	Northern_Ireland	SB0140	Europe	Cluster 5	Cluster2 19	Cluster3 48	5.3	
ERR841792	Northern_Ireland	SB0140	Europe	Cluster 5	Cluster2 19	Cluster3 48	5.3	
ERR841793	Northern_Ireland	SB0140	Europe	Cluster 5	Cluster2 19	Cluster3 48	5.3	
ERR841794	Northern_Ireland	SB0140	Europe	Cluster 5	Cluster2 19	Cluster3 48	5.3	
ERR841796	Northern_Ireland	SB0140	Europe	Cluster 5	Cluster2 19	Cluster3 48	5.3	
ERR841797	Northern_Ireland	SB0140	Europe	Cluster 5	Cluster2 19	Cluster3 48	5.3	
ERR841799	Northern_Ireland	SB0140	Europe	Cluster 5	Cluster2 19	Cluster3 48	5.3	
ERR841802	Northern_Ireland	SB0140	Europe	Cluster 5	Cluster2 19	Cluster3 48	5.3	
ERR841803	Northern_Ireland	SB0140	Europe	Cluster 5	Cluster2 19	Cluster3 48	5.3	
ERR841804	Northern_Ireland	SB0140	Europe	Cluster 5	Cluster2 19	Cluster3 48	5.3	
ERR841805	Northern_Ireland	SB0140	Europe	Cluster 5	Cluster2 19	Cluster3 48	5.3	
ERR841806	Northern_Ireland	SB0140	Europe	Cluster 5	Cluster2 19	Cluster3 48	5.3	
ERR841807	Northern_Ireland	SB0140	Europe	Cluster 5	Cluster2 19	Cluster3 48	5.3	
ERR841808	Northern_Ireland	SB0140	Europe	Cluster 5	Cluster2 19	Cluster3 48	5.3	
ERR841809	Northern_Ireland	SB0140	Europe	Cluster 5	Cluster2 19	Cluster3 49	5.3	
ERR841810	Northern_Ireland	SB0140	Europe	Cluster 5	Cluster2 19	Cluster3 50	5.3	
ERR841811	Northern_Ireland	SB0140	Europe	Cluster 5	Cluster2 19	Cluster3 48	5.3	
ERR841812	Northern_Ireland	SB0140	Europe	Cluster 5	Cluster2 19	Cluster3 48	5.3	
ERR841813	Northern_Ireland	SB0140	Europe	Cluster 5	Cluster2 19	Cluster3 48	5.3	
ERR841814	Northern_Ireland	SB0140	Europe	Cluster 5	Cluster2 19	Cluster3 48	5.3	
ERR841815	Northern_Ireland	SB0140	Europe	Cluster 5	Cluster2 19	Cluster3 48	5.3	
ERR841816	Northern_Ireland	SB0140	Europe	Cluster 5	Cluster2 19	Cluster3 48	5.3	
ERR841817	Northern_Ireland	SB0140	Europe	Cluster 5	Cluster2 19	Cluster3 48	5.3	
ERR841818	Northern_Ireland	SB0140	Europe	Cluster 5	Cluster2 19	Cluster3 48	5.3	
ERR841819	Northern_Ireland	SB0140	Europe	Cluster 5	Cluster2 19	Cluster3 48	5.3	
ERR841820	Northern_Ireland	SB0140	Europe	Cluster 5	Cluster2 19	Cluster3 48	5.3	
ERR841821	Northern_Ireland	SB0140	Europe	Cluster 5	Cluster2 19	Cluster3 48	5.3	
ERR841822	Northern_Ireland	SB0140	Europe	Cluster 5	Cluster2 19	Cluster3 49	5.3	
ERR841823	Northern_Ireland	SB0140	Europe	Cluster 5	Cluster2 19	Cluster3 48	5.3	
ERR841824	Northern_Ireland	SB0140	Europe	Cluster 5	Cluster2 19	Cluster3 48	5.3	
ERR841825	Northern_Ireland	SB0140	Europe	Cluster 5	Cluster2 19	Cluster3 48	5.3	
ERR841826	Northern_Ireland	SB0140	Europe	Cluster 5	Cluster2 19	Cluster3 48	5.3	
ERR841827	Northern_Ireland	SB0140	Europe	Cluster 5	Cluster2 19	Cluster3 48	5.3	
ERR841828	Northern_Ireland	SB0140	Europe	Cluster 5	Cluster2 19	Cluster3 48	5.3	
ERR841829	Northern_Ireland	SB0140	Europe	Cluster 5	Cluster2 19	Cluster3 48	5.3	
ERR841830	Northern_Ireland	SB0140	Europe	Cluster 5	Cluster2 19	Cluster3 48	5.3	
ERR841831	Northern_Ireland	SB0140	Europe	Cluster 5	Cluster2 19	Cluster3 48	5.3	
ERR841832	Northern_Ireland	SB0140	Europe	Cluster 5	Cluster2 19	Cluster3 48	5.3	
ERR841833	Northern_Ireland	SB0140	Europe	Cluster 5	Cluster2 19	Cluster3 48	5.3	
ERR841834	Northern_Ireland	SB0140	Europe	Cluster 5	Cluster2 19	Cluster3 50	5.3	
ERR841835	Northern_Ireland	SB0140	Europe	Cluster 5	Cluster2 19	Cluster3 48	5.3	
ERR841836	Northern_Ireland	SB0140	Europe	Cluster 5	Cluster2 19	Cluster3 50	5.3	
ERR841837	Northern_Ireland	SB0140	Europe	Cluster 5	Cluster2 19	Cluster3 50	5.3	
ERR841838	Northern_Ireland	SB0140	Europe	Cluster 5	Cluster2 19	Cluster3 50	5.3	
ERR841839	Northern_Ireland	SB0140	Europe	Cluster 5	Cluster2 19	Cluster3 50	5.3	
ERR841840	Northern_Ireland	SB0140	Europe	Cluster 5	Cluster2 19	Cluster3 50	5.3	

ERR841841	Northern_Ireland	SB0140	Europe	Cluster 5	Cluster2 19	Cluster3 48	5.3	
ERR841842	Northern_Ireland	SB0140	Europe	Cluster 5	Cluster2 19	Cluster3 48	5.3	
ERR841843	Northern_Ireland	SB0140	Europe	Cluster 5	Cluster2 19	Cluster3 49	5.3	
ERR841844	Northern_Ireland	SB0140	Europe	Cluster 5	Cluster2 19	Cluster3 50	5.3	
ERR841845	Northern_Ireland	SB0140	Europe	Cluster 5	Cluster2 19	Cluster3 50	5.3	
ERR841846	Northern_Ireland	SB0140	Europe	Cluster 5	Cluster2 19	Cluster3 50	5.3	
ERR841847	Northern_Ireland	SB0140	Europe	Cluster 5	Cluster2 19	Cluster3 50	5.3	
ERR841848	Northern_Ireland	SB0140	Europe	Cluster 5	Cluster2 19	Cluster3 50	5.3	
ERR841849	Northern_Ireland	SB0140	Europe	Cluster 5	Cluster2 19	Cluster3 48	5.3	
ERR841850	Northern_Ireland	SB0140	Europe	Cluster 5	Cluster2 19	Cluster3 50	5.3	
ERR841851	Northern_Ireland	SB0140	Europe	Cluster 5	Cluster2 19	Cluster3 50	5.3	
ERR841852	Northern_Ireland	SB0140	Europe	Cluster 5	Cluster2 19	Cluster3 48	5.3	
ERR841853	Northern_Ireland	SB0140	Europe	Cluster 5	Cluster2 19	Cluster3 48	5.3	
ERR841855	Northern_Ireland	SB0140	Europe	Cluster 5	Cluster2 19	Cluster3 48	5.3	
ERR841856	Northern_Ireland	SB0140	Europe	Cluster 5	Cluster2 19	Cluster3 48	5.3	
ERR841857	Northern_Ireland	SB0140	Europe	Cluster 5	Cluster2 19	Cluster3 48	5.3	
ERR841858	Northern_Ireland	SB0140	Europe	Cluster 5	Cluster2 19	Cluster3 48	5.3	
ERR841859	Northern_Ireland	SB0140	Europe	Cluster 5	Cluster2 19	Cluster3 48	5.3	
ERR841860	Northern_Ireland	SB0140	Europe	Cluster 5	Cluster2 19	Cluster3 50	5.3	
ERR841861	Northern_Ireland	SB0140	Europe	Cluster 5	Cluster2 19	Cluster3 48	5.3	
ERR841862	Northern_Ireland	SB0140	Europe	Cluster 5	Cluster2 19	Cluster3 48	5.3	
ERR841863	Northern_Ireland	SB0140	Europe	Cluster 5	Cluster2 19	Cluster3 48	5.3	
ERR841864	Northern_Ireland	SB0140	Europe	Cluster 5	Cluster2 19	Cluster3 48	5.3	
ERR841865	Northern_Ireland	SB0140	Europe	Cluster 5	Cluster2 19	Cluster3 48	5.3	
ERR841867	Northern_Ireland	SB0140	Europe	Cluster 5	Cluster2 19	Cluster3 48	5.3	
ERR841868	Northern_Ireland	SB0140	Europe	Cluster 5	Cluster2 19	Cluster3 48	5.3	
ERR841869	Northern_Ireland	SB0140	Europe	Cluster 5	Cluster2 19	Cluster3 48	5.3	
ERR841870	Northern_Ireland	SB0140	Europe	Cluster 5	Cluster2 19	Cluster3 48	5.3	
ERR841871	Northern_Ireland	SB0140	Europe	Cluster 5	Cluster2 19	Cluster3 49	5.3	
ERR841872	Northern_Ireland	SB0140	Europe	Cluster 5	Cluster2 19	Cluster3 48	5.3	
ERR841873	Northern_Ireland	SB0140	Europe	Cluster 5	Cluster2 19	Cluster3 49	5.3	
ERR841875	Northern_Ireland	SB0140	Europe	Cluster 5	Cluster2 19	Cluster3 48	5.3	
ERR841877	Northern_Ireland	SB0140	Europe	Cluster 5	Cluster2 19	Cluster3 50	5.3	
ERR841878	Northern_Ireland	SB0140	Europe	Cluster 5	Cluster2 19	Cluster3 48	5.3	
ERR841879	Northern_Ireland	SB0140	Europe	Cluster 5	Cluster2 19	Cluster3 50	5.3	
ERR841880	Northern_Ireland	SB0140	Europe	Cluster 5	Cluster2 19	Cluster3 48	5.3	
ERR841881	Northern_Ireland	SB0140	Europe	Cluster 5	Cluster2 19	Cluster3 48	5.3	
ERR841882	Northern_Ireland	SB0140	Europe	Cluster 5	Cluster2 19	Cluster3 48	5.3	
ERR841883	Northern_Ireland	SB0140	Europe	Cluster 5	Cluster2 19	Cluster3 48	5.3	
ERR841884	Northern_Ireland	SB0140	Europe	Cluster 5	Cluster2 19	Cluster3 48	5.3	
ERR841885	Northern_Ireland	SB0140	Europe	Cluster 5	Cluster2 19	Cluster3 49	5.3	
ERR841886	Northern_Ireland	SB0140	Europe	Cluster 5	Cluster2 19	Cluster3 48	5.3	
ERR841887	Northern_Ireland	SB0140	Europe	Cluster 5	Cluster2 19	Cluster3 48	5.3	
ERR841888	Northern_Ireland	SB0140	Europe	Cluster 5	Cluster2 19	Cluster3 48	5.3	
ERR841889	Northern_Ireland	SB0140	Europe	Cluster 5	Cluster2 19	Cluster3 48	5.3	
ERR841890	Northern_Ireland	SB0140	Europe	Cluster 5	Cluster2 19	Cluster3 48	5.3	
ERR841891	Northern_Ireland	SB0140	Europe	Cluster 5	Cluster2 19	Cluster3 48	5.3	
ERR841892	Northern_Ireland	SB0140	Europe	Cluster 5	Cluster2 19	Cluster3 48	5.3	
ERR841893	Northern_Ireland	SB0140	Europe	Cluster 5	Cluster2 19	Cluster3 48	5.3	
ERR841894	Northern_Ireland	SB0140	Europe	Cluster 5	Cluster2 19	Cluster3 48	5.3	
ERR841895	Northern_Ireland	SB0140	Europe	Cluster 5	Cluster2 19	Cluster3 48	5.3	

ERR841896	Northern_Ireland	SB0140	Europe	Cluster 5	Cluster2 19	Cluster3 48	5.3	
ERR841898	Northern_Ireland	SB0140	Europe	Cluster 5	Cluster2 19	Cluster3 48	5.3	
ERR841899	Northern_Ireland	SB0140	Europe	Cluster 5	Cluster2 19	Cluster3 48	5.3	
ERR841900	Northern_Ireland	SB0140	Europe	Cluster 5	Cluster2 19	Cluster3 48	5.3	
ERR841902	Northern_Ireland	SB0140	Europe	Cluster 5	Cluster2 19	Cluster3 48	5.3	

Cepas utilizadas para control

Strain (SRA accession)	Country	Sub lineage
SRR5485729	USA	1.1
SRR1791776	USA	1.3
SRR1792163	USA	1.3
SRR1792173	USA	1.3
SRR1792188	USA	1.3
SRR1792205	USA	1.3
SRR1792207	USA	1.3
SRR1792249	USA	1.3
SRR1792251	USA	1.3
SRR1792258	USA	1.3
SRR1792263	Unknown	1.3
SRR1792486	USA	1.3
SRR1792495	USA	1.3
SRR1791795	Unknown	2.1
SRR1792077	USA	2.1
SRR1792176	USA	2.2
SRR1791828	USA	3.2
SRR1792069	USA	3.2
SRR1791940	USA	3.2
SRR5817715	USA	3.2
SRR1791771	Mexico	4.3

ANEXO 2 – Características de las variantes identificadas

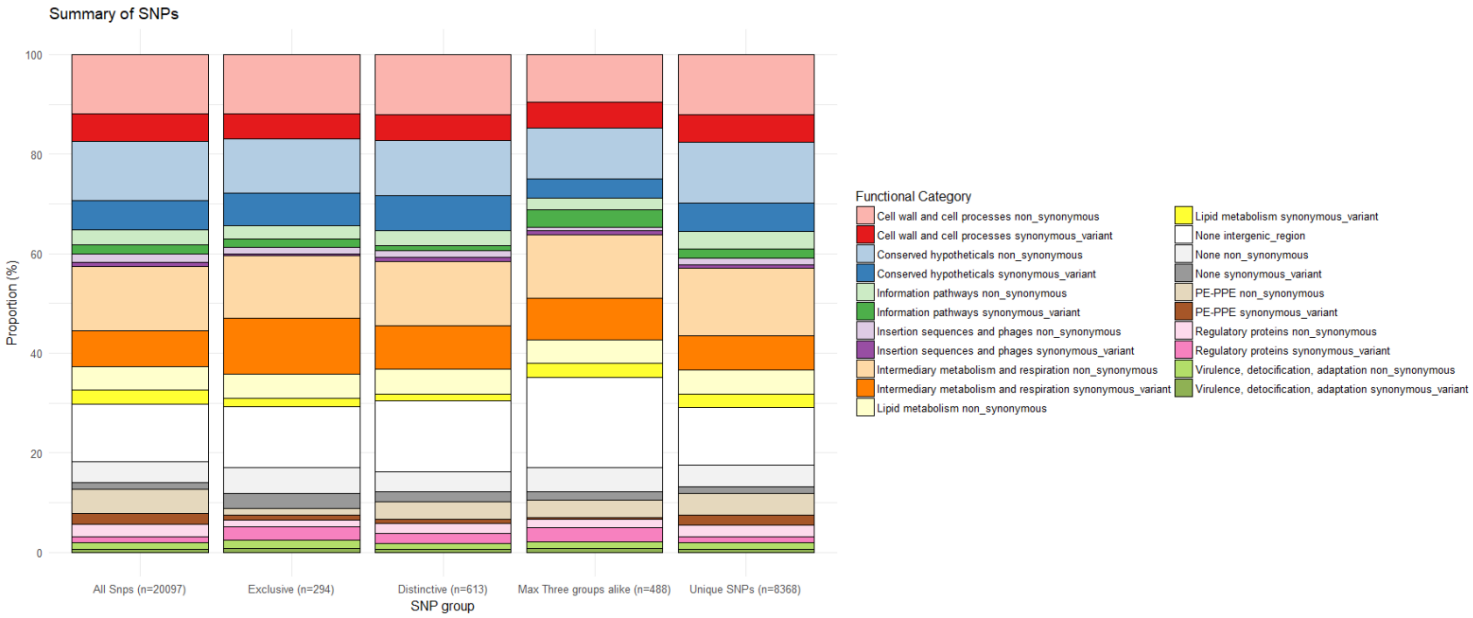

Figura Suplementaria 1.- División entre categorías funcionales y efecto de las variantes identificadas en las cepas analizadas. Datos mostrados para la totalidad de las variantes (n=20.097, all SNPs), para las variantes que son exclusivas de los (sub)linajes identificados (n=294, exclusive), para las variantes distintivas de sub linajes (n=613, distinctive), para las variantes comunes en dos o tres sub linajes completos (n=488, max three groups alike) y para las variantes encontradas en sólo una cepa (n=8.368, unique).

ANEXO 3 – Características de los spoligotipos identificados

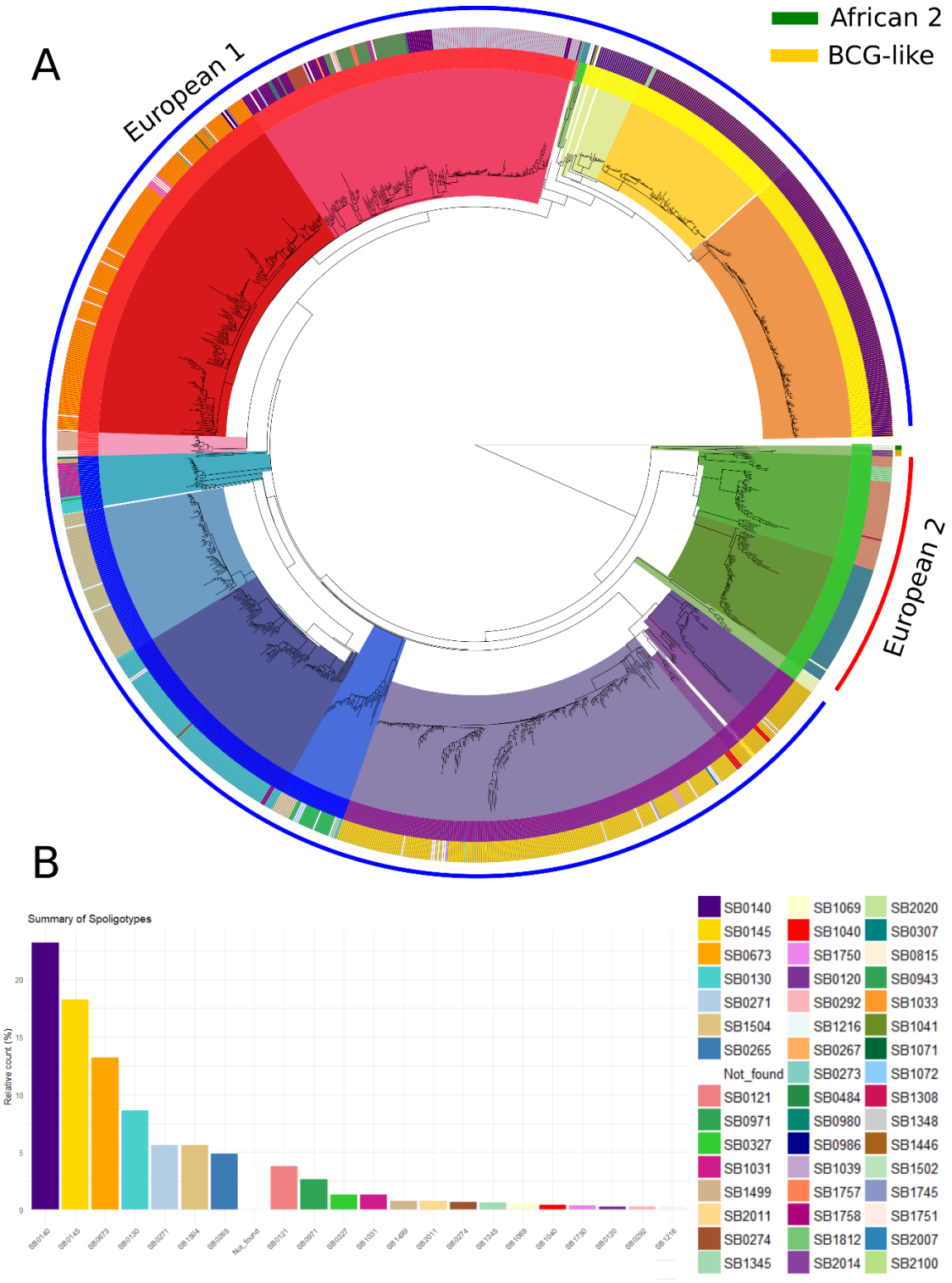

Figura Suplementaria 2.- (A) Localización de los spoligotipos identificados dentro de la filogenia mundial de *M. bovis* (segundo anillo más interno). En el anillo interno se identifican los cinco linajes principales, y en color de fondo los sub linajes que contiene cada uno, como en la **Figura 1**. Los anillos externos corresponden al complejo clonal al que las cepas pertenecen, de acuerdo con las características que los definen (espaciadores y RDs o SNPs). Se puede notar que hay un conjunto de cepas sin complejo clonal asignado. **(B)** Cantidades relativas de los patrones identificados en el total de cepas utilizadas. Los colores de spoligotipos indicados se corresponden tanto para (A) como para (B).

ANEXO 4 – Características de las Regiones de Diferencia identificadas

	ID	# Cepas	Largo *	Comienzo	Fin	Asociación	Especificidad
Determinantes de (sub)linajes	RD1009519	221	1399	1009519	1010918	3.2	100%
	RD1531687	34	12043	1531687	1543730	1.2	100%
	RD3/RD1766212	527	7272	1766212	1773484	3, 2.1, 2.2, 2.3	91%
	RD2317445	24	1622	2317445	2319067	3.1	100%
	RD3688270	245	1055	3688270	3689325	4.1, 4.2	83%
	RD3890969	121	3050	3890969	3894019	1.2, 1.3	94%
Otras RD	RD485420	16	1040	485420	486460	SB1031, New Zealand	100%
	RDAf2/RD681589	2	14079	681589	695668	China	100%
	RD1209296	9	1627	1209296	1210923	SB2011	100%
	RD1312977	2	1713	1312977	1314689	China	100%
	RD1319619	7	4793	1319619	1324412	SB2345, sub cluster in 1.2	100%
	RD1330942	11	2035	1330942	1332977	New Zealand in 4.3	100%
	RD1863310	12	6642	1863310	1869952	New Zealand in 4.3	100%
	RD2128199	2	1102	2128199	2129301	China	100%
	RD2176783	42	5040	2176783	2181823	NA	NA
	RD3379943	2	6724	3379943	3386666	China	100%
	RD3404007	7	3771	3404007	3407778	SB1069, subcluster in 1.1	100%
	RD4233584	6	5022	4233584	4238604	NA	NA

* El largo es una aproximación *in silico* del largo real que se puede obtener experimentalmente.

NA - Sin una asociación clara encontrada

Tabla Suplementaria 2.- 18 RDs principales identificadas entre todas las cepas estudiadas. El término Asociación refiere a los clados, spoligotipos o ubicaciones geográficas asociadas a la ausencia de estas regiones. La Especificidad es un porcentaje representante de la cantidad de cepas que carecen de una región y se encuentran dentro de la Asociación señalada sobre el total de cepas que carecen de la región dada.

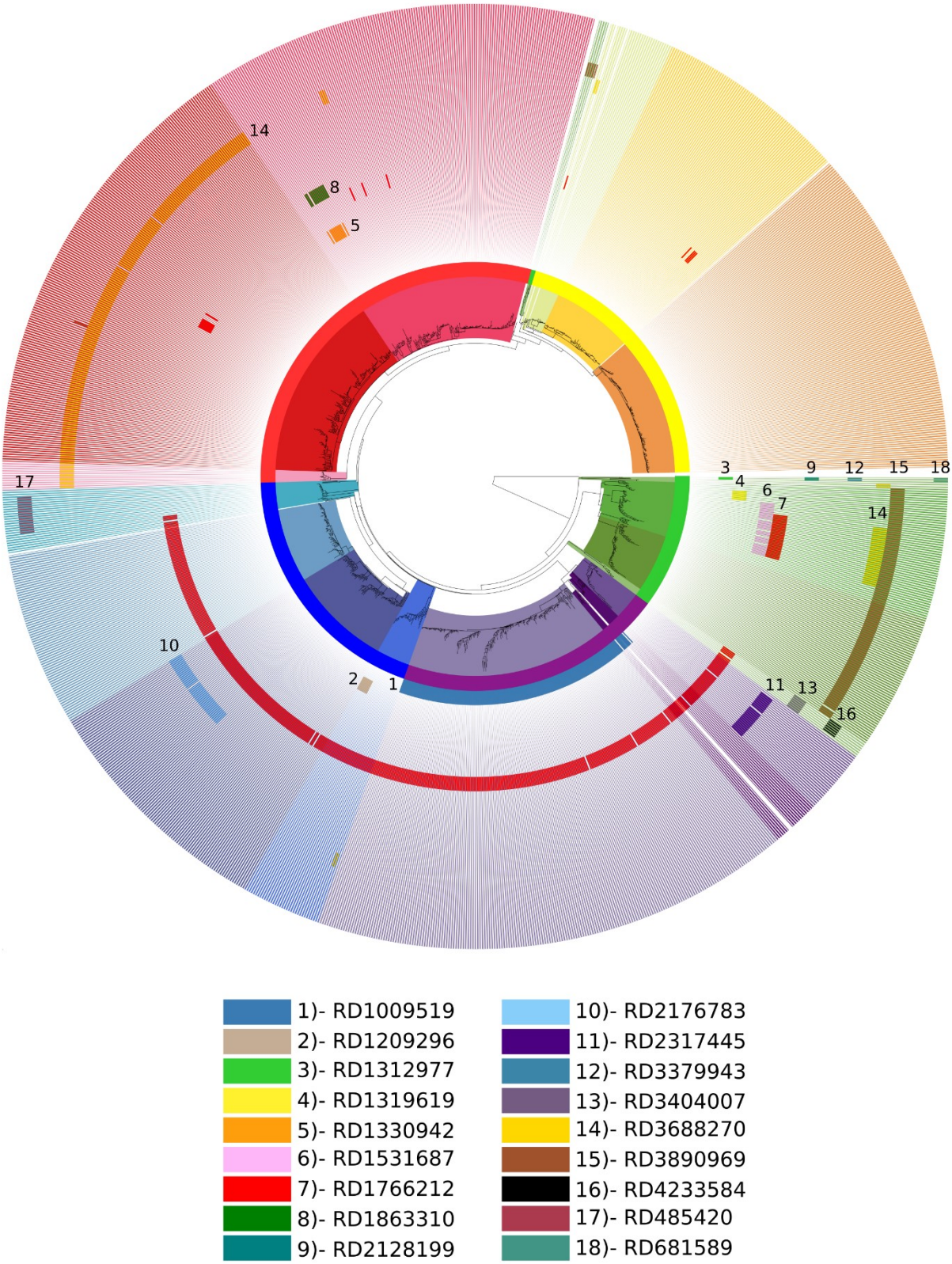

Figura Suplementaria 3.- Distribución de las 18 RDs identificadas dentro de la filogenia mundial de *M. bovis*. En el anillo interno se identifican los cinco linajes principales, y en color de fondo los sub linajes que contiene cada uno, como en la **Figura 1**.

ANEXO 5 – Reconstrucción filogenética original

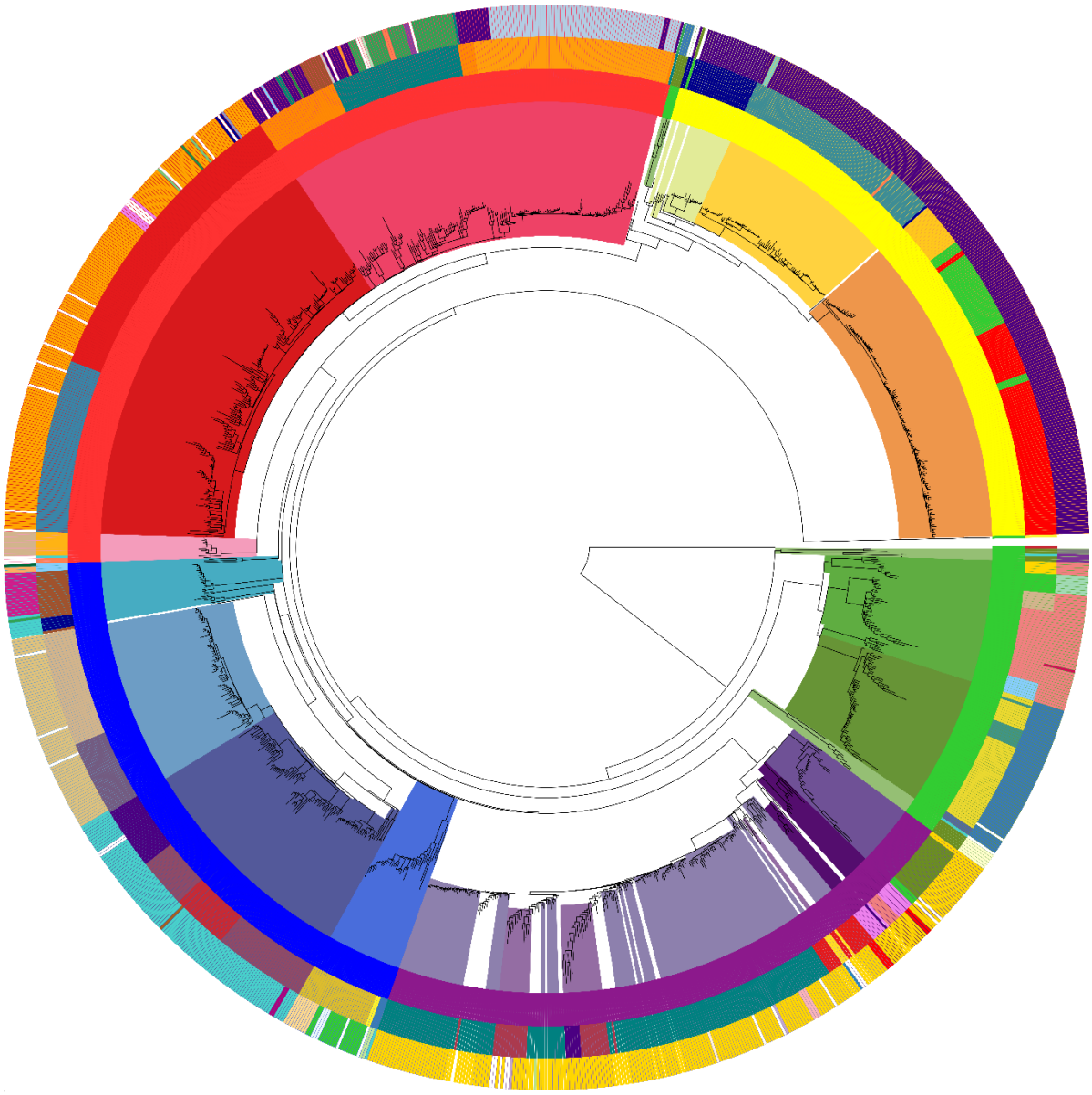

Figura Suplementaria 4.- Representación filogenética de la diversidad mundial de *M. bovis*, incluyendo los clusters originales de BAPS en el primer, segundo y tercer nivel de jerarquía (primer anillo, color de fondo y segundo anillo, respectivamente). El tercer anillo corresponde a los diferentes spoligotipos encontrados. Los colores en el segundo y tercer anillo fueron asignados al azar.

ANEXO 6 – SNPs seleccionados para el genotipado de *M. bovis*

Grupo	Tipo*	Cambio	#	Posición**	Ref	Alt	Producto	Locus ID	Gen
1	Exclusive	Syn	130	4077846	C	A	hypothetical protein	Mb3724	Mb3724
1	Exclusive	Syn	130	1048294	T	C	ATP-dependent DNA ligase	Mb0963	ligd
1	Exclusive	Syn	130	3467191	C	A	NADH dehydrogenase subunit B	Mb3170	nuoB
1.2	Distinctive	Syn	115	1779047	C	T	hypothetical protein	Mb1618c	Mb1618c
1.2	Distinctive	Syn	109	222939	A	G	hypothetical protein	Mb0197	Mb0197
1.2	Distinctive	Syn	112	2670247	G	A	GTP-binding protein LepA	Mb2427c	lepA
1.2	Distinctive	Syn	110	2973577	C	T	polyphosphate glucokinase	Mb2721	ppgK
1.2	Distinctive	Syn	118	566174	A	C	transcriptional regulator	Mb0484	Mb0484
1.2	Distinctive	Syn	124	321009	T	C	integral membrane nitrite extrusion protein NARU	Mb0273	narU
1.3	Exclusive	Syn	57	359494	C	T	trans-aconitate 2-methyltransferase	Mb0302	tam
1.3	Exclusive	Syn	57	4043145	G	A	periplasmic dipeptide-binding lipoprotein DppA	Mb3690c	dppA
1, 2	Shared	Syn	1210	3265769	A	G	hypothetical protein	Mb2980	Mb2980
2	Exclusive	Syn	248	1513457	C	T	drugs-transport transmembrane ATP-binding	Mb1383	irta
2	Exclusive	Syn	248	4227214	G	A	integral membrane transport protein	Mb3853c	mmpL8
2	Exclusive	Syn	248	299636	A	G	succinate dehydrogenase	Mb0255c	Mb0255c
2	Exclusive	Syn	248	3448807	C	T	hypothetical protein	Mb3154c	tgs1
2.1	Exclusive	Syn	33	802338	C	T	30S ribosomal protein S10	Mb0720	rpsJ
2.1	Exclusive	Syn	33	904074	C	T	amidophosphoribosyltransferase	Mb0831	purF
2.1	Exclusive	Syn	33	2662983	G	A	sulfate-transport ABC transporter ATP-binding	Mb2419c	cysA1
2.3	Shared	Syn	802	232188	G	C	transcriptional regulator	Mb0202	Mb0202
2.2	Distinctive	Missense	174	4299984	A	G	protease	Mb3913c	mycp1
2.3	Exclusive	Syn	73	2363358	G	A	5-methyltetrahydrofolate-homocysteine methyltransferase	Mb2148c	metH
2.3	Exclusive	Syn	73	3856175	T	C	Mce associated protein	Mb3522c	Mb3522c
2.3	Exclusive	Syn	73	699173	C	T	two component sensor kinase	Mb0616c	Mb0616c
2.3	Distinctive	Syn	73	1998304	G	A	oxidoreductase	Mb1803	Mb1803
2.3	Distinctive	Syn	73	824300	T	C	D-xylulose kinase XylB	Mb0750	xylB
2.3	Distinctive	Intergenic	567	1475059	A	AC	- Intergenic region -	rrf-ogt	rrf-ogt
2.1, 2.3	Shared	Syn	344	402355	T	C	hypothetical protein	Mb0343	Mb0343
2.1, 2.3	Shared	Syn	943	2278824	G	A	polyketide synthase	Mb2074c	pks12
2.2, 2.3	Shared	Syn	181	265675	T	C	aldehyde dehydrogenase	Mb0228c	Mb0228c
2.2, 2.3	Shared	Syn	175	547559	A	G	enoyl-CoA hydratase	Mb0464c	echA2
2.2, 2.3	Shared	Syn	175	2471336	G	A	bifunctional glutamine-synthetase adenylyltransferase	Mb2245c	glnE
3	Distinctive	Syn	991	1828921	A	G	Hypothetical protein	Mb1663c	Mb1663c
3	Distinctive	Syn	1205	130237	T	C	cation-transporter ATPase I	Mb0111c	ctpI
3	Distinctive	Syn	1208	2373913	C	T	hypothetical protein	Mb2156	Mb2156
3.1	Exclusive	Syn	27	240616	C	T	transmembrane transport protein MmpL11	Mb0208c	mmpL11
3.1	Exclusive	Syn	27	2742170	G	A	alpha-glucosidase	Mb2498	aglA

3.2	Distinctive	Syn	214	3618489	C	T	hypothetical protein	Mb3309	acce5
3.2	Distinctive	Syn	214	854043	G	A	two component system response transcriptional	Mb0780	phoP
3.2	Distinctive	Syn	214	483845	T	C	transmembrane transport protein MmpL1A	Mb0409c	mmpL1a
3.2	Distinctive	Syn	214	1295429	C	T	respiratory nitrate reductase subunit gamma NarI	Mb1196	narI
4.1	Exclusive	Syn	12	1226368	C	T	glycosyl hydrolase	Mb1126	Mb1126
4.1	Exclusive	Syn	12	4077342	G	A	hypothetical protein	Mb3723c	vapc48
4.1	Exclusive	Syn	12	3579879	G	C	two component sensory transduction histidine kinase MTRB	Mb3273c	mtrB
4.1	Exclusive	Syn	12	3901971	G	A	lipid-transfer protein	Mb3552	ltp4
4.1	Distinctive	Syn	1213	804997	T	C	50S ribosomal protein L2	Mb0724	rplB
4.1	Distinctive	Syn	12	3619791	G	A	thiosulfate sulfurtransferase	Mb3311	sseA
4.2	Distinctive	Syn	350	1493708	G	A	glycogen phosphorylase GlgP	Mb1363	glgP
4.2	Exclusive	Syn	184	1527223	A	G	transcriptional regulator	Mb1393	Mb1393
4.2	Distinctive	Syn	350	3278284	G	A	pyruvate carboxylase	Mb2991c	pca
4.3	Distinctive	Missense	355	1530862	A	T	PPE family protein	Mb1396c	PPE19
4.2, 4.3	Shared	Syn	351	1597426	G	A	esterase	Mb1461c	lipO
5.2	Exclusive	Syn	80	236518	C	T	zinc metalloprotease	Mb0204c	zmp1
5.2	Exclusive	Syn	80	2989417	C	T	hypothetical protein	Mb2740c	Mb2740c
5.3	Exclusive	Syn	137	2202679	G	A	hypothetical protein	Mb2002c	mpt64
5.3	Exclusive	Syn	137	3210746	G	A	phenolpthiocerol synthesis type-I polyketide	Mb2957	ppsB

Tabla Suplementaria 3.- Conjunto de 56 SNPs filogenéticamente informativos seleccionados para el genotipado de *M. bovis*.

* Correspondiente a las categorías de grupos de variantes identificadas, detalladas en el texto principal.
** Posiciones basadas en el genoma de referencia, AF2122/97.