IDENTIFICACIÓN DE LA PRESENCIA/AUSENCIA DE GENES DE LA VÍA DEL ÁCIDO SIÁLICO EN LEPTOSPIRAS Y ESPIROQUETAS

Tesis de Grado de la Licenciatura en Bioquímica

ESTUDIANTE: Gabriel Martínez

Facultad de Ciencias, UdelaR

INTRODUCCION

En el presente trabajo nos proponemos comprender la evolución de la vía del ácido N-acetilneurámico (ácido siálico) en *Leptospiras* y *Espiroquetas*. Utilizamos para este fin herramientas de genómica comparativa. La importancia científica de esta investigación cobra relevancia al considerar la patogenicidad que presentan algunas de estas especies, siendo responsables de la enfermedad conocida como leptospirosis.

La leptospirosis es una antropozoonosis de distribución mundial, presentándose como una enfermedad febril aguda con múltiples manifestaciones causadas por vasculitis generalizada de variada gravedad. La misma abarca, desde enfermedades sistemáticas auto limitadas, hasta condiciones potencialmente mortales por ictericia e insuficiencia renal. El agente etiológico de la leptospirosis, como se mencionó más arriba, se resume a un orden específico de bacterias llamadas espiroquetas, las cuales se caracterizan por ser microorganismos gram-negativos anaerobios helicoidales, móviles, comunes tanto en el medio ambiente (cursos de agua) como parasitando diversos animales (Pumarola et al., 2002). La dificultad principal para definir el papel de las espiroquetas en diferentes enfermedades ha sido el poder distinguir las especies y cepas individuales de dicho microorganismo, esto último debido en gran parte a la dificultad de su cultivo en el laboratorio.

También nos concentraremos en el estudio del género *Leptospira*, las cuales presentan formas de vida libre y parasitaria. Hasta el momento sólo se encuentra públicamente disponible el genoma completo de tres especies (aunque hay genomas de distintas cepas): *Leptospira interrogans, Leptospira borgpetersenii y Leptospira biflexa* (Cespedes et al., 2005). Las dos primeras son agentes patogénicos y de ellas, L. *interrogans* es la principal especie asociada con leptospirosis animal y humana. La última especie es un organismo de vida libre no patogénico.

Las *Leptospiras* patógenas incluyen alrededor de 80 tipos serológicos definidos, que antes eran designados como especies y ahora se conocen como serovariedades o serotipos de *L. interrogans*. El microorganismo puede infectar a la mayoría de los mamíferos del mundo, así como a reptiles, anfibios, peces, aves e invertebrados. El mecanismo por el cual se mantiene en la naturaleza involucra la colonización persistente de los túbulos renales de los animales portadores y los seres humanos se infectan por contacto directo o indirecto con la orina o la sangre de los animales infectados.

Si bien la anotación genómica en estos microorganismos ha mejorado sustancialmente desde los primeros reportes, existen muchas incógnitas aún particularmente a nivel de las diferentes vías metabólicas. Algunas de estas vías juegan un papel importante en la evasión inmune y presentan por lo tanto un interés sustantivo.

En particular nos interesa concentrarnos en el estudio de la síntesis, procesamiento y funcionalidad del ácido N-acetilneurámico (ac. siálico) en todas las *Leptospiras* patógenas y no-patógenas, comparando su genoma con el material genético de la mencionada vía estudiada con antelación en organismos modelos. La virulencia de las bacterias extracelulares depende de diferentes mecanismos que favorecen la invasión y colonización de los tejidos, que incluyen entre otros, mecanismos anti-fagocitosis e inhibición del complemento o de las proteínas activadas por el complemento. Algunas bacterias gram-negativas contienen en su pared celular ácido siálico, el cual es un inhibidor de la activación del complemento por la vía alternativa, por fijar al factor H. (Hakobyan et al., 2008)

El sistema de complemento está constituido por moléculas implicadas principalmente en la defensa del organismo frente a infecciones y células tumorales. Parte de los factores del complemento potencian la inflamación y la fagocitosis, produciendo la lisis de células infectadas y microorganismos. La mayor parte de los factores del complemento son proteínas plasmáticas (C2, C3, C4, C6, C7, C8, Factor B y Factor I), siendo una pequeña proporción de ellos proteínas de membrana. En la activación del complemento se ponen en marcha una serie de reacciones consecutivas en cascada, de forma tal que a partir de cada una de ellas se genera un producto activo, que además de determinar que la reacción consecutiva prosiga, puede tener diferentes acciones biológicas importantes en la defensa del organismo. Algunos de los factores del complemento son enzimas con carácter proteolítico, por lo cual durante el proceso de activación algunas moléculas son rotas en fragmentos, a los que para identificarlos se les añade letras minúsculas (Ej. C3a, C3b). Estos fragmentos poseen importantes funciones biológicas y son mediadores de la inflamación. El factor H (FH) es una proteína plasmática homóloga a la C4BP (C4 binding protein), que tiene la capacidad de unir C3b en la fase fluida y promueve la disociación del complejo C3bBb, factor necesario en el proceso lisogénico del microorganismo invasor.

Sólo el material genómico de las 7 especies de *Leptospiras* que vamos a analizar está secuenciado completamente. El genoma del resto de las especies de *Espiroquetas* que investigaremos, no está secuenciado en su totalidad, trabajaremos con taxones de diversos tamaños. Está información es relevante ya que la ausencia en determinada especie de *Espiroqueta* (salvo las *Leptospiras*) de ortólogos de alguno de los genes de la vía del ácido siálico, por cualquiera de los métodos de análisis empleado, no sería razón suficiente para afirmar la ausencia definitiva de ese gen en esa especie en particular.

Utilizando el modelo de *Escherichia coli* 07:k1 CE10, se ha identificado al operón de la vía del ácido N-acetilneurámico presente en bacterias gram-negativas, el mismo está conformado por cuatro genes principales (nanA, nanE, nanK, nanT). Por otra parte, es sabido que en *Leptospiras* dicho operón no presenta todos los genes, hecho que se

presume es debido a la ausencia de una enzima que participa en el catabolismo del ácido siálico (codificada por nanA).

Este resultado es algo sorprendente ya que los procesos selectivos en procariotas tienden a eliminar rápidamente aquellas secuencias que no presentan ventajas evolutivas. La existencia del resto de genes de la vía del ácido siálico podría justificarse por al menos tres razones a saber: a) el gen que aparece como faltante (identificado por similaridad de secuencia) se encuentra realmente presente y difiere fuertemente en la secuencia de aminoácidos, b) la función de este gen en la vía es realizada por otra proteína diferente a la de las gram-negativas, c) el resto de las proteínas de la vía cumplen algún rol en otras vías.

Para la identificación de los genes de esta vía (genes ortólogos) presentes en los distintos géneros de espiroquetas, se diseñó una estrategia basada en diferentes enfoques del problema, mayoritariamente en homología de secuencias. Este concepto suele utilizarse para predecir la función de un gen (anotación genómica): si la secuencia del gen *A* cuya función es conocida es homóloga a la secuencia del gen *B* cuya función es desconocida, entonces puede inferirse que B podría compartir la función con A. Con esta idea en mente, una primer estrategia consistirá en bajar de la "KEGG orthology database" las secuencias génicas ortólogas conocidas de los genes de la vía (operón nan), para realizar luego un BLAST contra el genoma de las *Leptospiras* de interés (genomas secuenciados completamente). De esta forma se expande considerablemente la diversidad de secuencias en la búsqueda de genes ortólogos.

La segunda estrategia estará fundamentada en la creación de una base de datos de secuencias ortólogas entre las espiroquetas conocidas, buscando identificar el patrón de presencia/ausencia en organismos cercanos filogenéticamente. A su vez, se utilizará un organismo modelo externo, como *E. coli* 07:k1 CE10, a fin de identificar aquellos clusters de genes bien conservados, particularmente los que corresponden a la vía de interés. Existen varios algoritmos (conjunto establecido de instrucciones o reglas bien definidas, ordenadas y finitas, que permite realizar una actividad mediante pasos sucesivos) para agrupar secuencias basadas en similitud. En nuestro caso se utilizará OrthoMCL (algoritmo de clústeres de Markov)(Van Dongen et al., 2000).

A partir de los resultados obtenidos en esta etapa se podrá determinar si el fenómeno de presencia/ausencia de genes de la vía es generalizado entre Espiroquetas o particular de *Leptospiras*.

OBJETIVO

El objetivo del presente trabajo consiste en la identificación de secuencias homólogas de genes de la vía del ácido siálico (operón nan) en los distintos géneros de espiroquetas, con particular interés en *Leptospiras*. Para eso, buscaremos patrones de presencia/ausencia de dichos genes en organismos filogenéticamente cercanos, utilizando herramientas genómicas como el software OrthoMCL, y la base de datos "KEGG orthology".

Principales características de las Leptospiras

Las *Leptospiras* (del griego *lepto*: delgado; y del latín *spira*: espiral) se agrupan también en serotipos, de acuerdo a sus relaciones antigénicas. Actualmente existen más de 200 serotipos reconocidos. En la reunión del año 2002 del Comité de Taxonomía de Leptospira de la Unión Internacional de Sociedades Microbiológicas, se aprobó la siguiente nomenclatura para *Leptospira*: el género y la especie deben ser escritas en cursiva, con el nombre del serotipo sin cursiva y la primera letra en mayúscula.

Las *Leptospiras* son microorganismos muy delgados, de 6 a 20 µm de largo y 0,1 a 0,2 µm de ancho, flexibles, helicoidales, con distancias promedio entre crestas consecutivas del orden de los 0,5 µm, presentando extremos generalmente curvados en forma de gancho (fig. 1). Son extraordinariamente móviles por el uso de filamentos axiales (axostilos), dividiéndose por fisión binaria. En virtud de su estructura en espiral alrededor de su eje axial, pueden presentarse hasta 20 enrollamientos en función de su longitud total. Son organismos aerobios estrictos, cultivándose con facilidad en medios artificiales (como el medio Ellinghausen-McCullough-Johnson-Harris) suplementados con 0,2-1% de suero de conejo, lo cual favorece el crecimiento de cepas muy exigentes (Johnson & Harris, 1967). Pueden sobrevivir largo tiempo en el agua o ambiente húmedo y templado, con pH neutro o ligeramente alcalino, aunque el pH óptimo para su crecimiento se encuentra comprendido entre 7,2-7,6.

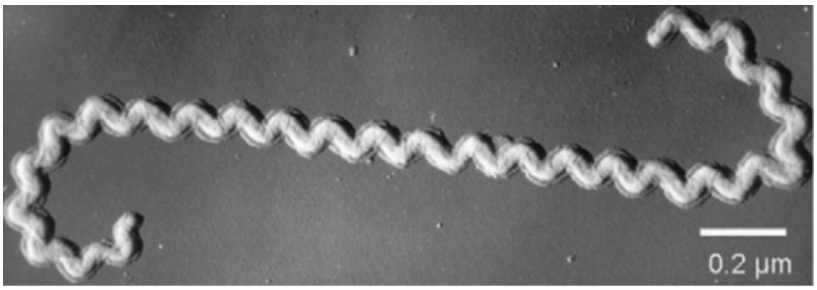


Figura 1 - Microscopía electrónica de Barrido de *Leptospira spp.* (Picardeau, 2013)

Las *Leptospiras* poseen una envoltura celular similar a las bacterias gram-negativas, consistiendo en dos membranas, una citoplasmática y otra externa. Sin embargo, la

capa de peptidoglicano (también llamada "mureína", es un copolímero formado por una secuencia alternante de N-acetilglucosamina y ácido N-acetilmurámico unidos mediante enlaces β[1,4]) está asociada con la membrana citoplasmática en vez de la membrana externa, algo que es único en las espiroquetas (Picardeau et al., 2001). La membrana externa contiene una variedad de lipoproteínas, proteínas transmembranas y periféricas. Se ha demostrado que varias proteínas de la membrana exterior se adhieren a la matriz extracelular y al factor H (una proteína de control del complemento), pudiendo ser de importancia en la adhesión de las *Leptospiras* a los tejidos del huésped y en la resistencia a las acciones del sistema del complemento (Verma et al., 2006).

Su capacidad de invadir tejidos está facilitada por la producción de la enzima hialuronidasa, la cual produce alteraciones en la permeabilidad del tejido conjuntivo del organismo huésped al hidrolizar el ácido hialurónico (Díaz et al., 2005). El genoma de las *Leptospiras* patogénicas consiste en dos cromosomas. El tamaño de los genomas de las especies *L. interrogans* Copenhageni y *L. interrogans* Lai es de aproximadamente 4.6 Mb, mientras que el de las especies *L. borgpetersenii* JB197 y H550, es de 3.9 Mb.

Una diferencia importante entre las especies *L. borgpetersenii* y *L. interrogans* es el mecanismo invasivo utilizado por cada especie, donde *L. borgpetersenii* es transmitida usualmente por la exposición directa a tejidos infectados, mientras que *L. interrogans* lo hace por lo general, a través del agua y el suelo contaminados con la orina de animales portadores (en sus riñones).

Características	[¥] *L. borgpetersenii*		[¥] *L. interrogans*		[¥] *L. biflexa*			
	CI	CII	CI	CII	CI	CII	P74	LE-1 prophage[b]
Tamaño (pb)	3,614,456	317,335	4,277,185	350,181	3,603,977	277,995	74,116	73,623
Contenido G-C (%)	41.0	41.2	35.1	35.0	38.9	39.3	37.5	38.5
Codificación protéica (%)	80	80	74.9	75.5	92.3	93.3	90.9	93.4
Secuencias codificantes protéicas								
CDS[a]	2,607	237	3,105	274	3,268	266	56	82
Con función asignada	1,644	135	1,817	159	2,042	141	31	19
Hipotéticamente conservado	373	32	484	34	464	43	5	2
Hipotéticamente único	590	70	804	81	762	82	20	61
Transposones	215	26	26	0	8	1	1	0
Pseudogenes	340	28	38	3	32	1	0	0
Genes ARN transferencia	37	0	37	0	35	0	0	0
Genes ARN ribosomales								
23S	2	0	2	0	2	0	0	0
16S	2	0	2	0	2	0	0	0
5S	1	0	1	0	2	0	0	0

[¥] *L. borgpetersenii* serovar Hardjo strain L550, *L. interrogans* serovar Copenhageni strain Fiocruz, *L. biflexa* serovar Patoc strain Ames
[a] excluyendo a transposones y pseudogenes
[b] [11,38]

Tabla 1- Modificada de (Picardeau et al., 2008) Comparación de las tres especies de *Leptospiras* investigadas a nivel genómico, considerando el tamaño del mismo medido en pares de bases, porcentaje del contenido de CG y el porcentaje del genoma que codifica para proteínas. Donde las siglas CI, CII y P74, significan respectivamente Cromosoma I, Cromosoma II y Plásmido 74 .

El gran número de genes defectuosos y de genes de inserción que presenta *L. borgpetersenii,* en comparación con *L. interrogans,* sugiere que las primeras están atravesando un proceso de degradación genómica, con una continua pérdida de genes necesarios para la supervivencia fuera del huésped animal (Bulach et al., 2006).

PATOGÉNESIS

Ciclo infeccioso de la Leptospirosis

Las *Leptospiras* invaden el cuerpo humano por las membranas, las mucosas o cortes en la piel. Si están en número suficiente como para vencer la resistencia del huésped, se multiplican y producen una infección clínica generalizada (etapa leptospirémica), localizándose la infección en el riñón. En ese momento, los microorganismos aparecen en la orina y se depositan en el medio ambiente con cada infección (etapa leptospirúrica). Existe también la tendencia a localizarse en el útero gestante, por lo que en esta forma puede causar el aborto.

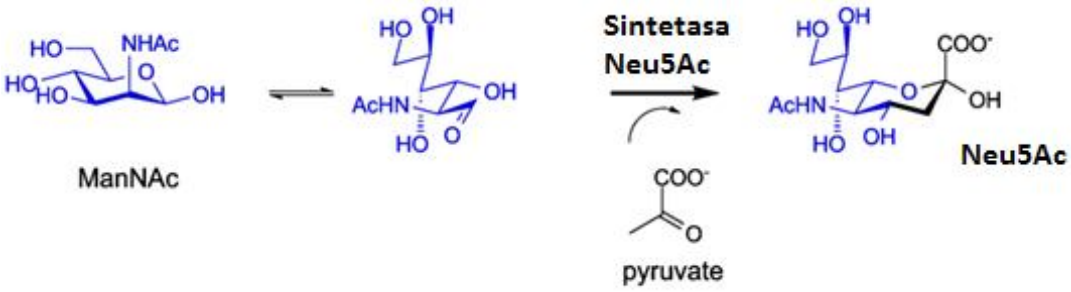
Según Blodd y colaboradores (1982) y (1988), después de penetrar por la piel (abrasiones) o mucosas, u al consumir alimentos u agua contaminada, los microorganismos se multiplican rápidamente en el torrente sanguíneo, provocando varios días de fiebre al paciente (fase leptospirémica, que puede durar hasta 7 días). En la fase siguiente, septicémica, puede haber casos clínicos con muerte subsecuente por la producción de algunos serotipos de hemolisinas, las que provocan la hemólisis grave, anoxia anémica con nefrosis hemoglobinuria (particularmente en animales jóvenes), o simplemente cursar de forma asintomática, frecuente en adultos. Luego que declina la fiebre aparecen anticuerpos en el torrente circulatorio (fase de formación de anticuerpos, se inicia al final de la primera semana hasta el final de la segunda) y microorganismos en la orina.

Younes y colaboradores (1998) evaluaron una citotoxina que inhibe la K-Na ATPasa, y que se encuentra en fracciones de glicoproteínas de *L. interrogans.* Dicha citotoxina contenía un inhibidor específico de la K-Na ATPasa que provocaba las disfunciones celulares responsables de los síntomas, en particular con los desórdenes electrolíticos, siendo el posible mecanismo de la fisiopatología de la leptospirosis.
Blodd y colaboradores (1982) plantean además que es frecuente la localización de leptospiras en el sistema nervioso de ovinos y caprinos provocando síntomas de encefalitis.

IMPORTANCIA DEL ÁCIDO NEURÁMICO EN LOS PROCESOS BIOLÓGICOS

ÁCIDO NEURÁMICO

El ácido neurámico es un monosacárido constituido por nueve átomos de carbono, derivado de una cetononosa. En teoría, el ácido neurámico podría considerarse como el producto de la condensación aldólica del ácido pirúvico y la D-manosamina (2-amino-2-desoximanosa) por acción de la ácido neurámico sintetasa.



Aunque no es posible encontrar a esta molécula como tal en la naturaleza, la misma aparece ampliamente distribuida bajo la forma de sus múltiples productos derivados. Presentándose tanto en tejidos animales como en bacterias, en especial como constituyentes de glicoproteínas y gangliósidos (lípido formado por dos ácidos grasos, una molécula de esfingosina, una cabeza de polisacáridos y uno o más grupos de ácido siálico). Los gangliósidos se encuentran en las membranas celulares de la mayoría de los tejidos, especialmente del tejido nervioso. El nombre de *"ácido neurámico"* fue introducido por el científico alemán Ernst Klenk en 1941, en referencia a los lípidos de origen cerebral desde los cuales se aisló por primera vez como producto derivado del procesamiento de dichos lípidos (Klenk, 1941). El símbolo utilizado para el ácido neurámico es Neu.

Los radicales *N* u *O* que dan lugar a derivados del ácido neurámico son denominados en su conjunto como "ácidos siálicos". El más común es el ácido N-acetilneurámico (también denominado Neu5Ac o NANA). El mismo es un monosacárido que se encuadra tanto en el grupo de los amino azúcares como en el de azúcares ácidos.

Los ácidos siálicos son cetoácidos que se descubrieron en mamíferos, aunque se encuentran en la mayor parte de los celomados, en protostomados (por ejemplo, artrópodos) y deuterostomados (por ejemplo, cordados y equinodermos) (Reglero et al., 2007).

En referencia a la etimología del nombre cabe mencionar que el término «ácido siálico» proviene del griego «sialos», cuyo significado es saliva. El compuesto fue descubierto por Gunnar Blix en 1936 (Blix, 1936), cuando lo aisló de la mucina bovina

de las glándulas submaxilares, pero recién en 1952 fue introducido dicho nombre (Blix et al.,1952).

La ruta biosintética del Neu5Ac tiene lugar a través de las siguientes reacciones:

a) síntesis de ManAc-6-P (ácido N-acetilmanosamina-6-fosfato)

b) síntesis de ManNAc (ácido N-acetilmanosamina)

c) síntesis de Neu5Ac (ácido N-acetilneurámico)

d) activación del monómero por CMP-β-Neu5Ac

e) transferencia del Neu5Ac a un aceptor

En el catabolismo de estos compuestos juegan un importante papel las enzimas neuraminidasa y la liasa del ácido N-acetilneurámico (codificada por el gen nanA). En animales, los ácidos siálicos están involucrados en interacciones célula-célula, interviniendo en la regulación de procesos de reconocimiento celular, en cambio en microorganismos se encuentran presentes en un número escaso de bacterias y hongos. Estos microorganismos utilizan los ácidos siálicos para establecer relaciones simbióticas o parasitarias con animales, o como potenciales fuentes de nitrógeno y carbono (Almagro et al., 2009).

ESTRUCTURA

Estructuralmente los ácidos siálicos constan de un anillo piranósico, constituido por cinco carbonos, y un oxígeno formado por un hemiacetal en conformación en silla. De los glicoconjugados que existen en la naturaleza, los ácidos siálicos se presentan sólo en la configuración α excepto en un compuesto potencialmente rico en energía, el CMP-ácido siálico donde el carbono anomérico se encuentra en configuración β (Mathews et al., 2003).

En 1957 Barry y Goebel descubrieron la presencia de ácidos siálicos en *Escherichia coli* K-235, al aislar un polímero de Neu5Ac, el cual fue nombrado posteriormente como ácido colomínico (Nunziato et al., 1995). Pese a que la mayoría de las bacterias no presentan la capacidad de producir ácidos siálicos se ha encontrado que varias cepas patógenas, como *E. coli* K12 son capaces de biosintetizarlos. También se han encontrado en las estructuras de otras bacterias patógenas que no pueden producirlos debido a que lo adquieren desde el organismo hospedador. Además algunos patógenos pueden catabolizar ácidos siálicos del medio como fuente de carbono, nitrógeno y energía, o como fuente de amino azúcares para biosintetizar la pared celular (Plumbridge et al., 1999).

En general, los ácidos siálicos se encuentran formando parte de la estructura capsular de la bacteria habiéndose detectado en forma de homopolímeros (unidades repetidas de un mismo polímero) o de heteropolímeros (estructuras polisacáridas formadas por diferentes tipos de monómeros) (Fernández et al., 2007).

Los lipopolisacáridos (LPS) o endotoxinas son los mayores componentes de la membrana externa de las bacterias gram-negativas, desempeñando una importante función en la activación de su sistema inmune al constituir el antígeno superficial más importante de este tipo de bacterias. Los LPS están compuestos por una región lipídica y una glicosídica con funciones separadas y/o sinérgicas lo que convierte a estas moléculas en los factores de virulencia más complejos de comprender. Siendo los principales responsables del shock endotóxico inducido por bacterias gram-negativas.

Algunas bacterias sintetizan diferentes tipos de ácidos siálicos, como el ácido 3-desoxi-2-ceto-nonulosónico, derivado de los ácidos legionámicos (originalmente descubierto en el LPS de *Legionella pneumophila*) y pseudámicos (originalmente descubierto en el LPS de *P. aeruginosa*). Ambos ácidos son similares en su estructura a los ácidos siálicos, diferenciándose en algunos pasos de su biosíntesis (Ricaldi et al., 2012).

Desde el punto de vista biosintético todos los ácidos siálicos, a excepción de algunos azúcares sialilados específicos de ciertas bacterias, derivan del ácido 2-ceto-3-desoxi-D-glicero-D-galacto-nonulosónico (KDN) y del ácido N-acetilneurámico (Neu5Ac) (Tanner et al., 2005).

METABOLISMO DE LOS ÁCIDOS SIÁLICOS EN BACTERIAS

La biosíntesis de los ácidos siálicos en bacterias se inicia por medio de los precursores comunes del metabolismo intermediario: a partir de D-glucosa se obtiene D-fructosa-6-fosfato (Fru-6P) que posteriormente se convierte en D-glucosamina-6-fosfato (GlcN-6P), la cual es acetilada para generar N-acetil-D-glucosamina-6-fosfato (GlcNAc-6P). Luego por acción de la enzima GlcNAc-6P-2-epimerasa, el compuesto GlcNAc-6P es convertido en un epímero, el N-acetilmanosamina-6-fosfato (ManNAc-6P). Posteriormente, ManNAc-6P es desfosforilado a ácido N-acetilmanosamina (ManNAc) (Zapata et al., 1992 ; Petersen et al.,2000).

OPERON NAN

Un operon está formado por un grupo de genes o regiones genómicas que codifican determinadas proteínas o enzimas, las cuales trabajarán en forma secuencial en una vía metabólica específica. La expresión del operon está regulada por elementos de control constitutivos, como un promotor, un operador y ciertos genes reguladores.

El principal aspecto en la regulación de la expresión génica ocurre en el control de la iniciación de la transcripción. La decisión de transcribir o no un gen, decisión que se toma y se regula en el paso de iniciación, es casi siempre el evento principal y frecuentemente es el único evento en el control de la expresión de los genes. La iniciación de la transcripción es un proceso inherentemente asimétrico, ya que solo una de las dos bandas del ADN se transcribe, la banda llamada con sentido o también banda codificadora.

El operon nan está formado por un conjunto de genes poli-cistrónicos (fig. 2), genes que al ser transcriptos van a dar origen a una molécula de ARNm, la cual codifica para varias cadenas polipetídicas diferentes. Este operon está constituido por cuatro genes principales, el gen nanA (codifica para la liasa del ácido N-acetilneurámico), el gen nanT (codifica para una proteína transmenbrana, una Transportadora simporte del ácido N-acetilneurámico, Neu5Ac.), el gen nanE (codifica para la epimerasa de N-acetilglucosamina-6-fosfato) y el gen nanK (codifica a una kinasa de N-acetilmanosamina).

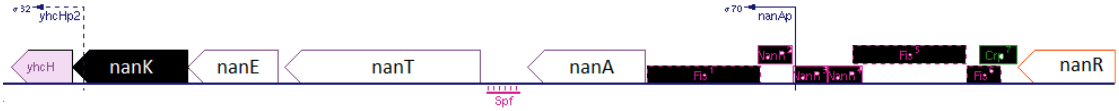


Figura 2- Entorno de localización del operon nan en genoma de *E.coli* MG1665, aparecen los sitios reguladores de unión al ADN, que reprimen la transcripción génica por unión al promotor. Ubicándose de derecha a izquierda (nanA), (nanT), (nanE) y en negro (nanK). La línea discontinua indica que no hay pruebas de alta calidad para confirmar el alcance de esta unidad de transcripción y la existencia o la ubicación de esta entidad.(Extraída de la página web de EcoCyc http://ecocyc.org).

nanA

Es el gen codificante de la enzima liasa del ácido N–acetilneurámico. Esta enzima, cataliza el clivaje del ácido N-acetilneurámico (ácido siálico) para dar dos productos: piruvato y N-acetilmanosamina, vía la formación de una base de Shiff. Es la primera enzima de la vía que conduce a la degradación de ácido siálico (Vimr et al., 1985).

Han sido resueltas tanto las estructuras cristalinas de la liasa del ácido N–acetilneurámico (cuya PDB ID es 1NAL), como las de sus complejos con hydroxipruvato y piruvato, proporcionando información importante acerca de la

ubicación y caracterización del sitio activo. El mismo se encuentra en el entorno de la posición 161 de la cadena polipeptídica, posición ocupada por un residuo de lisina. También brindó información sobre el mecanismo de reacción con el sustrato específico correspondiente. La estructura del sitio activo se exploró usando compuestos de nueve átomos de carbono derivados de NeuNAc modificados (Kiefelt et al., 2000).

La estructura cristalina de la liasa del ácido N-acetilneurámico se presenta bajo la forma de un tetrámero. El trímero aparente encontrado por diversos experimentos de filtración en gel puede ser explicado por la migración anómala debida a la forma asimétrica de dicho tetrámero. Se encontró que la expresión de la enzima es inducida por el crecimiento del ácido siálico en la ausencia de glucosa. La transcripción del conjunto génico integrante del operon (la transcripción desde el nanA hasta nanK) está reprimida por el regulon NanR (Kalivoda et al., 2003).

nanT

El gen nanT codifica para una proteína transmembrana de transporte bi-direccional, la cual facilita el desplazamiento del ácido siálico a través del mecanismo simporte acoplado al transporte de hidrogeniones (H^+) en *E. coli K-* 12. La función transportadora se activa sólo por el gradiente de protones (Mulligan et al., 2012).

Este gen codifica a un miembro principal de la superfamilia de permeasas que tiene catorce segmentos transmembrana, con el extremo C-terminal y N-terminal localizados en el citosol (Martínez et al., 1995). La expresión de nanT es inducida por la presencia de ácido siálico.

nanE

Se supone que nanE codifica para una epimerasa, la N-acetil-manosamina-6-fosfato. Dicha enzima está implicada en la reacción que permite al microorganismo utilizar ácido N-acetilneuramico y ácido N-acetilmanosamina como fuentes de carbono (Plumbridge & Vmir, 1999). Recientemente se purificó de *E. coli* K92, una proteína con actividad enzimática similar a la epimerasa N-acetilmanosamina-6-fosfato, determinándose que su peso molecular es 38,4 kDa (Ferrero et al., 2007).

nanK

El gen nanK codifica para una enzima que pertenece a la familia de quinasas ROK, y se ha demostrado que presenta tanto actividad quinasa, siendo su sustrato el ácido N-acetilmanosamina, como actividad glucoquinasa ordinaria en ambos casos in vitro. Originalmente el sitio de inicio de este gen fue asignado únicamente bajo consideraciones de secuencia. Sin embargo se ha encontrado que el sitio de inicio real

está situado 33 pares de bases corriente abajo de donde se pensaba (Sohanpal et al., 2004).

nanR

Los genes regulados por nanR "regulador del ácido N-acetilneurámico", están involucrados en el transporte y metabolismo del ácido N-acetilneurámico (Chu et al., 2007). La proteína codificada por nanR es inactivada por la presencia en el entorno de ácido N-acetilneurámico y es miembro de la familia FadR / GntR. Los integrantes de esta familia tienen dos dominios: un dominio N-terminal con un motivo de unión al ADN hélice-giro-hélice y un dominio C-terminal con capacidad de formar dímeros. Los modelos tridimensionales de la región N-terminal de FadR y NanR muestran similitud topológica y un 26% de identidad entre sus respectivas secuencias (Kaliboda et al., 2003).

La enzima codificada por el gen nanR regula la transcripción del operon nan al unirse a una región de aproximadamente 30 pares de bases, la cual contiene tres motivos conservados (hexámeros de secuencias GGTATA) dispuestos en forma repetida y consecutiva. Sin embargo, esta enzima puede ser desplazada de dicha región por el ácido N-acetilneurámico. La región de 30 pares de bases a la que se une la proteína está cerca de un sitio de metilación la cual es prevenida por vinculación directa con nanR (Chu et al., 2007).

Aunque se ha demostrado que nanR puede formar homodímeros en solución su estequiometria de unión al ADN es desconocida. Cuando esta proteína está reprimiendo la transcripción la misma se superpone a toda la región promotora.

Gen	Ubicación	Longitud de Los genes	Peso molecular	PH óptimo
nanA	**citosol**	**894 pb / 297 aa**	**105,0 kD**	5.91
nanT	**Memb interna**	**1.491 pb / 496 aa**	**53,551 kD monómero**	-----
nanE	**citosol**	**690 pb / 229 aa**	**24,074 kD**	-----
nanK	**citosol**	**876 bp / 291 aa**	**29,644 kD**	-----
nanR	**citosol**	**792 pb / 263 aa**	**29,524 kD**	**-----**

Tabla 2- Propiedades relevantes de los genes del operon nan, en *Escherichia coli* 07:k1 CE10 (microorganismo de referencia) donde se puede apreciar su ubicación celular, longitud genómica del mismo, y peso molecular de las proteínas por ellos codificadas según la literatura disponible.

Genes vinculados al operon nanA

Al profundizar sobre los alcances del presente trabajo, surgió el interés por ampliar el estudio en lo concerniente a la comparación de la homología de secuencias a tres genes que intervienen en forma directa en la funcionalidad del operon nan, nos referimos a neuA, neuB y neuC.

En la figura 3 puede apreciarse la ubicación de dichos genes en la vía metabólica de amino azúcares y nucleótidos aplicada al modelo de estudio de referencia *E. coli* 07:k1 CE10.

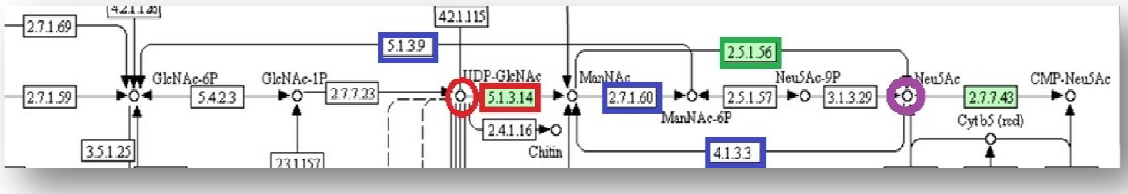


Figura 3 - Pathway eoc00520 por KEGG perteneciente al organismo *E. coli* K1:07 CE10, donde los rectángulos de diversos colores muestran los códigos de las distintas enzimas involucradas en la vía del ácido siálico, los rectángulos en azul corresponden a genes pertenecientes al operon nan (nanA-4.1.3.3/nanK-2.7.1.60/nanE-5.1.3.9), en rojo (neuC-5.1.3.14), en verde (neuB-2.5.1.56) y con fondo verde pero sin color de contorno (neuA-2.7.7.43).

neuA

La estructura de la enzima codificada por este gen aún no se ha podido establecer firmemente. La enzima purificada a partir de extractos de células mostró una banda de 50 kDa en SDS-PAGE (Vann et al., 1997) y su denominación es ''Sintetasa Citidina-5'-monofosfato-ácido-N-acetilneurámico''. Esta enzima que se encuentra tanto en eucariotas como en procariotas, tiene un papel importante en la activación de los ácidos siálicos a sus compuestos derivados-CMP en las reacciones de transferencia del sialato.

La enzima nativa se purificó a partir de la fracción citoplásmica obtenida de *Escherichia coli* 07:k1 CE10. La enzima recombinante se determinó que era activa en una conformación monómerica, pero puede existir como un dímero o agregados mayores bajo algunas condiciones in vitro. La misma se ha purificado y utilizado en la síntesis enzimática de CMP-N-acetil-β-neuraminato mediante mutagénesis, con el objetivo de identificar los residuos de aminoácidos presentes en el sitio activo con importancia catalítica (Stoughton et al., 1999).

En el patógeno *Escherichia coli* 07:k1 CE10, al igual que en las *Leptospiras interrogans* (Lai 56601, Lai IPAV y Copenhageni), el polisacárido capsular es un homopolímero con enlace glucosídico β [2,8] ligado al ácido N-acetilneurámico.

neuB

El gen neuB codifica para la enzima Sintetasa del ácido N-acetilneurámico que participa en la biosíntesis de ácido siálico. La masa molecular aparente de la subunidad enzimática codificada por este gen se determinó por SDS-PAGE (Vann et al., 1997). Mediante la espectrometría de masas (EM) utilizando ionización por electrospray (ESI), se ha podido identificar esta enzima, apreciándose que se trata de un homotetrámero (probablemente compuesta de dos dímeros). La forma activa primaria se cree que es tetramérica.

Esta enzima se distingue de la enzima eucariota (N-acetilneuramico 9-fosfato sintasa) porque utiliza N-acetil-D-manosamina-6-fosfato como sustrato, en lugar de N-acetil-β-D-manosamina-6-fosfato.

neuC

El gen neuC codifica a una enzima que actúa como una epimerasa bacteriana (Vann et al., 2004) la cual es un agente causal de infecciones del tracto urinario. Su cápsula es un factor de virulencia que consiste en un polímero homogéneo lineal de enlace glucosídico α[2,8] ligado al ácido N-acetilneurámico.

La mayoría de las bacterias no biosintetizan ácidos siálicos, pero algunos patógenos y bacterias simbióticas lo hacen como un medio de evadir al sistema inmune del huésped. El ácido siálico se presenta en la superficie de la célula bacteriana (en polisacáridos capsulares) con el fin de imitar las células de mamíferos. Los patógenos que biosintetizan ácidos siálicos incluyen *Neisseria meningitidis, Escherichia coli* K12 *y Campylobacter jejuni*.

En la figura 4 puede apreciarse un esquema referido a la vía metabólica del ácido N-acetilneurámico en bacterias gram-negativas, similar al modelo *E. coli.* 07:K1 CE10 utilizado. El mismo involucra cinco pasos, empezando con la acción de la liasa de ácido N-acetilneurámico (nanA) la cual remueve un grupo piruvato del Neu5Ac convirtiéndolo en N-acetilmanosamina (ManNAc) entonces la kinasa de N-acetilmanosamina (nanK) adiciona un grupo fosfato en la posición del carbono 6, convirtiéndolo en ácido N-acetil-manosamina-6-fosfato (ManNAc-6P). Próximamente sobre la mencionada molécula actuará la epimerasa N-acetilmanosamina (nanE), transformándola en N-acetil-glucosamina-6-P, transcurridos ciertos pasos metabólicos va a ser transformada en Fructuosa-6-P.

Por otro lado, es posible llegar a la molécula de N-acetilmanosamina (ManaNac) desde UDP-N-acetil-glucosamina, vía acción enzimática de UDP-GlcNAc-2-epimerasa (neuC).

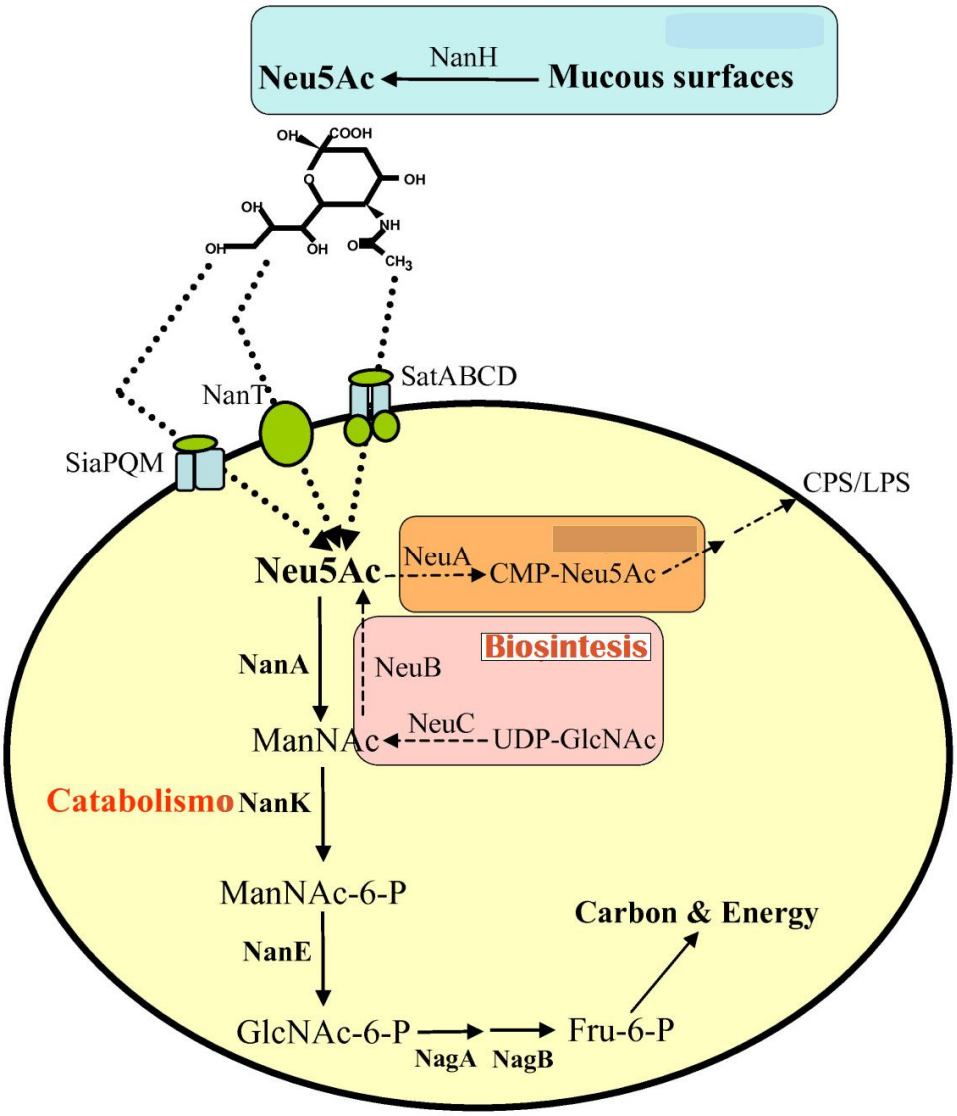


Figura 4 - Representación esquemática del metabolismo del ácido siálico entre el dominio de Bacteria. Resumen de la variedad de vías de utilización del ácido N-acetilneurámico o Neu5Ac (ácido siálico). Las vías catabólicas de Neu5Ac involucran varios pasos que empiezan con la expresión de nanA, prosiguen con la acción de otros integrantes del operon nan , en orden de intervención nanK y nanE. Se destaca también la acción de los genes neuB y neuC, vinculados a la vía del ácido siálico, ligados al proceso de su biosíntesis. Modificado de (Almagro et al.,2009).

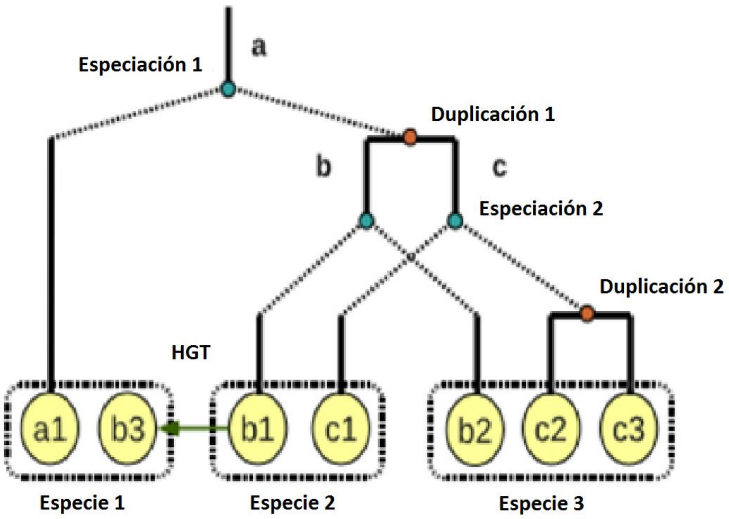
Por último Neu5Ac es convertido en CMP-Neu5Ac por la CMP-Neu5Ac sintetasa (neuA) para ser exportado hacia la membrana externa.

El estudio de homología de secuencia para evaluar la presencia/ausencia de los genes del operon nan en espiroquetas y especialmente en *Leptospiras* se realizó utilizando el software OrthoMCL, que describiremos a continuación.

OrthoMCL

OrthoMCL es una herramienta informática utilizada para identificar y colectar secuencias ortólogas entre distintas especies. Es un algoritmo utilizado para la agrupación de las proteínas en grupos ortólogos en base a la similitud de su secuencia (Li et al., 2003). El software posee además una base de datos pública (OrthoMCL-DB), la cual en su última versión (5) contiene 124.740 grupos ortólogos agrupados de 1.398.546 proteínas. Dichas proteínas fueron obtenidas a partir de 100 genomas eucariotas, 16 genomas de archaea y 34 genomas bacterianos (Chen et al., 2006). Teniendo como objetivo de esta Tesis el establecer grupos ortólogos de las enzimas constituyentes de la vía del ácido siálico en genomas bacterianos del género de las espiroquetas, se procedió a descargar, instalar y ejecutar el software libre OrthoMCL, sobre el conjunto de datos de interés como se describe a continuación.

Un grupo OrthoMCL es un conjunto de proteínas provenientes de una o más especies que representan ortólogos y parálogos putativos. Los genes homólogos, que derivan de una secuencia ancestral común, se dividen en dos grandes grupos dependiendo de la forma en que se produjo su diferenciación. Por un lado, los genes ortólogos han evolucionado por eventos de especiación. Se cree que mantienen funciones similares, conservadas del ancestro común (ancestor). Por otro lado, los genes parálogos marcan el segundo grupo, que a su vez se dividen en out-parálogos y en in-parálogos. Los genes out-parálogos surgieron de una duplicación que precedió a una especiación, mientras que los genes in-parálogos evolucionaron más recientemente por duplicación después de la especiación (Koonin et al., 2005). Todas estas relaciones se ilustran en la figura 5.



Si uno de dos genes parálogos es ortólogo de otra especie, se le llama gen xenólogo. Una distinción clara entre ortólogos y parálogos es crítica para la construcción de una clasificación evolutiva y anotación robusta de los genes.

Figura 5 . Ilustración de las relaciones entre genes homólogos: la imagen representa la situación de tres especies después de eventos de especiación, duplicación y transferencia horizontal de genes (HGT). Los genes c2 y c3 son in-parálogos mientras b1 y c1 son out-parálogos con respecto a la especiación 2. Mientras que c1 es ortólogo tanto a c2 como a c3, los genes b1 y b2 son llamados ortólogos principales. El gen b3 evolucionó por transferencia horizontal de genes (HGT) desde b1 y todos los genes son xenólogos a a1.

El cálculo computacional se realiza sobre la secuencia proteica a pesar que, por definición las relaciones evolutivas que el algoritmo del OrthoMCL intenta descubrir existen a nivel genómico. Una complicación que puede surgir como consecuencia de la utilización de las proteínas, en lugar de secuencias de ADN es que los proteomas pueden contener productos de transcriptos alternativos, consideración irrelevante cuando de bacterias se trata. El motivo de esto, es la utilización por parte del programa de la salida tabular de BLAST, organizando las proteínas en un banco de datos relacional para que posteriormente el algoritmo MCL (se explica más adelante) genere los grupos de ortólogos.

Hay una serie de enfoques utilizables para la identificación de grupos de proteínas con una historia evolutiva común (Chen et al., 2007). Estos incluyen métodos basados en la filogenia, métricas de distancia evolutiva y la similitud de secuencias. OrthoMCL utiliza el método de mejor hit en BLAST recíproco y hace un ajuste de distancias especiales (normalización) para distinguir genes ortólogos de parálogos. Los grupos se generan a partir de las puntuaciones normalizadas utilizando la agrupación de Markov. Un modelo de Markov es un gráfico dirigido que almacena las transiciones entre los diferentes estados. El número de órdenes en una cadena de Markov indica cuántos estados se usan para determinar la probabilidad de los estados actuales. En un modelo de Markov de primer orden, la probabilidad del estado actual depende solo del estado anterior, en uno de segundo orden la probabilidad de un estado depende de los dos estados previos y así sucesivamente. En cada cadena de Markov una matriz de transición almacena las probabilidades de transiciones para cada combinación de estados.

Información general sobre el algoritmo OrthoMCL

Para este programa en particular los proteomas deben encontrarse en formato FASTA. Los archivos con los proteomas se filtran para eliminar las secuencias de baja calidad sobre la base de su longitud (procedimiento detallado más adelante).
El programa requiere el acceso a un sistema de base de datos relacional. En principio es compatible con Oracle® y mySQL®. También requiere de recursos informáticos suficientes para poder realizar un BLASTP (BLAST proteico) sobre el conjunto de proteomas de interés todos-contra-todos. Cabe mencionar que BLAST (Basic Local Alignment Search Tool) es un programa informático de alineamiento heurístico de secuencias de tipo local, que tiene la capacidad de utilizar como fuente ADN o proteínas (como se utilizó en el presente trabajo).
El programa es capaz de comparar una secuencia problema (*query*) contra una gran cantidad de secuencias que se encuentren en una base de datos. Usando un algoritmo heurístico, el programa encuentra las secuencias de la base de datos que tiene mayor identidad local a la secuencia problema. Debido a la naturaleza del algoritmo, no se

puede asegurar que la solución encontrada sea la mejor posible, pero los resultados suelen ser lo suficientemente buenos para fines prácticos.

Crear grupos ortólogos desde los proteomas mediante el software OrthoMCL

Introducción

La entrada al OrthoMCL es un conjunto de proteomas en formato FASTA, mientras que la salida es un conjunto de archivos incluidos en un directorio denominado "pares", el cual contiene las relaciones entre pares de proteínas y sus puntajes. Se clasifican en potenciales ortólogos, xenólogos y parálogos, además del archivo adicional grupos.txt, el cual contiene los grupos creados por la agrupación de los pares con el algoritmo de MCL.

Este algoritmo se basa en el de Markov, es un algoritmo rápido escalable caracterizado por el análisis de conglomerados, el gráfico de la agrupación, la implementación y evaluación comparativa en general. La primera operación utilizada es la expansión, que coincide con la multiplicación de matrices normales. El segundo es la inflación, que es matemáticamente hablando es una matriz de Hadamard (matemático francés quién contribuyo en el desarrollo del análisis infinitesimal y desarrolló el teorema sobre el valor absoluto de un determinante), seguido por un escalado diagonal.

Cómo el software procesa las proteínas?

Hay cuatro fases de procesamiento de la información suministrada por el software: 1) durante la primera fase se preparan los archivos FASTA (uno por proteoma); 2) en la segunda fase se realiza un emparejamiento ("matcheo") de todos-contra-todos vía BLASTP; 3) en la tercera fase se cargan los resultados BLASTP en la base de datos relacional y se ejecuta un programa del OrthoMCL para encontrar pares significativos de proteínas; 4) la cuarta y última fase consiste en utilizar el algoritmo MCL.

Basamos los requisitos de hardware y de las estimaciones de tiempo en este conjunto de datos de referencia. El predictor más importante de las necesidades de recursos y tiempo es el número de similitudes génicas significativas que se alcanzan. La relación entre las mencionadas variables es tal, que conlleva a que cambien de forma no lineal con respecto al procesamiento de la información y el tiempo insumido para procesarla.

orthomclAdjustFasta

La secuencia de comandos orthomclAdjustFasta prepara los archivos para su uso en la base de datos. Para tal fin, se proporcionó un ID de 3 o 4 caracteres para cada organismo. Después de ejecutar el script con cada proteoma, orthomclFilterFasta filtra

y separa la información en dos archivos diferenciados: goodProteins.fasta (donde se agrupan las proteínas potencialmente homólogas de las especies investigadas) y poorProteins.fasta (contiene las proteínas que no presentan identidad aminoacídica por tener e-valor por debajo de lo requerido). El archivo goodProteins será el input para los siguientes pasos.

En la etapa correspondiente al BLASTP (all x all)

Los 57 genomas de las espiroquetas se emparejan todos contra todos, a partir del archivo goodProteins.fasta. Luego se da formato a la entrada con el makeblastdb (comando que prepara la base de datos para el BLASTP), y se ejecuta blastp.

$ makeblastdb -in goodProteins.fasta -dbtype prot

$ blastp -query goodProteins.fasta -db goodProteins.fasta -eValue 0.0000001 -num_threads 4 -best_hit_score_edge 0:05 -out blastp_allXall outfmt 6

*El parámetro ***-****best_hit_score_edge 00:05*- restringe los resultados a los mejores encontrados por cada consulta de acuerdo al e-value seleccionado.
*El parámetro *-outfmt 6* establece los formatos de la salida en formato tabular.
*El parámetro -num_threads 4 permitió el uso de 4 núcleos del servidor de procesamiento.

Parámetros Críticos

Se presentan dos parámetros cruciales en orthomclPairs (pares de orthomcl) que pueden afectar la distancia dentro de un mismo grupo de proteínas vinculadas. El primero es el evalueExponentCutoff cuyo valor recomendado, como se mencionó anteriormente, es de -5. Hay que tener en cuenta que el programa de MCL reconoce la diferencia entre los pares débiles y fuertes. El efecto principal del umbral débil es para evitar la temprana eliminación entre relaciones alejadas. El segundo parámetro es percentMatchCutoff, con un valor recomendado de 50. Para que dos proteínas formen un par, se requiere que el número de aminoácidos de la más corta debe ser de al menos el 50% de la longitud de la de mayor tamaño. Esto aumenta la probabilidad de que las dos proteínas compartan dominios.

Análisis Posterior

Para comprobar que las proteínas asignadas como ortólogas lo fueran realmente, se realizó el alineamiento de cada grupo mediante el programa T-coffee.

La sustitución o la conservación de residuos aminoacídicos en una región específica de la secuencia, cuyas cadenas laterales conservan propiedades químicas similares indica que esa zona para la proteína, tiene importancia estructural o funcional.

Para comprobar las posibles relaciones evolutivas entre los genes homólogos del operon nan en las especies de la familia de las espiroquetas, se utilizaron los alineamientos de secuencias en la construcción e interpretación de árboles filogenéticos para cada gen de interés. El grado de diferenciación entre las secuencias de proteínas homólogas en un conjunto problema determinado, es directamente proporcional a la distancia evolutiva entre ellas es decir, una alta identidad de secuencias sugiere que tienen un ancestro común más reciente, mientras que una baja identidad sugiere que la divergencia es más remota. Si se verifica que algunos de los proteomas tienen buenas asignaciones funcionales, entonces pueden utilizarse las mismas para evaluar la consistencia de los grupos.

Construcción de árboles filogenéticos

Para la construcción de los árboles filogenéticos se utilizó el software libre BEAST v.1.8.0.

El primer paso fue determinar cuál sería el mejor modelo de evolución molecular a utilizar en el software BEAST. Con este fin se empleó el programa libre PROTTEST 3, el cual reconoce como entrada a los archivos que tengan extensión phylum, correspondientes al alineamiento de los proteomas que son codificadas por el operon nan. Por lo tanto hubo de utilizarse un programa libre (sqlconvertor) para convertir al formato requerido los archivos FASTA obtenidos como resultados de la ejecución del programa OrthoMCL.
Asimismo, como los archivos de entrada del software BEAST (v.1.8.0) deben poseer extensión .xml, es preciso utilizar el convertidor anteriormente mencionado para pasar los archivos de formato FASTA a formato nexus y se utilizó el programa "beauti" (dentro del software BEAST) el cual nos permite convertir los archivos nexus en archivos con extensión .xml. Todo esto da como resultado archivos con dos clases de extensiones: por un lado archivos (.log) y por otro lado archivos (.trees). Estos últimos muestran los posibles árboles filogenéticos obtenidos con todas las combinaciones factibles surgidas de 10.000.000 de intentos.

Para la labor de monitorear la convergencia entre los árboles salidos del BEAST, se instala y ejecuta el software de libre distribución "TRACER". Este programa utiliza como input los archivos .log salidos de la ejecución del BEAST. Una vez obtenido la convergencia buscada se ejecuta el programa "treeannotator" presente en el software BEAST1.8.0, utilizando como input los archivos .trees para cada gen del operon nan, compactándolos en el más probable. Como salida se obtiene un archivo .tre el cual representa el árbol filogenético de cada una de las proteínas de estudio.

Para poder visualizar los árboles filogenéticos, como se enseñan a continuación, se empleó el software FIGTREE v.1.5, que representa en forma gráfica los árboles filogenéticos (uno por gen).

RESULTADOS

Operon nan y proteínas vinculadas a la investigación (neuA, neuB y neuC)

Los resultados obtenidos por el OrthoMCL se resumieron en la tabla 3. Allí se muestra para las 57 especies de Espiroquetas estudiadas y una especie de *Escherichia coli* (utilizada como microorganismo de referencia) la presencia/ausencia de los diferentes genes constituyentes del operon nan y genes vinculados al funcionamiento de la vía del ácido siálico. Se dividieron las tablas en ocho columnas ubicándose en las tres primeras los genes neuA, neuB y neuC, luego se destinan cuatro columnas para los genes del operon nan, nanA, nanT, nanE, nanK. Por último se destina una columna al gen dapA. La diferenciación entre la ausencia y presencia del gen en particular se denotó marcando con cero (0) la falta del gen y con (1) la presencia del mismo en el genoma bacteriano.

CODIGO NOMBRE	neuA	neuB	neuC	nanA	nanT	nanE	nanK	dapA
eoc Escherichia coli 07:K1 CE10	0	1	1	1	1	1	1	0
Eco Escherichia coli K-12	0	1	1	1	1	1	1	0
bbz Borrelia burgdorferi B31	0	0	0	0	0	1	0	0
bbz Borrelia burgdorferi ZS7	0	0	0	0	0	1	0	0
Bbn Borrelia burgdorferi N40	0	0	0	0	0	1	0	0
bbj Borrelia burgdorferi JD1	0	0	0	0	0	1	0	0
bga Borrelia garinii PBi	0	0	0	0	0	1	0	0
bgb Borrelia garinii BgVir	0	0	0	0	0	1	0	0
bgn Borrelia garinii NMJW1	0	0	0	0	0	1	0	0
baf Borrelia afzelii (FLI)	0	0	0	0	0	1	0	0
bafz Borrelia afzelii (Maryland)	0	0	0	0	0	1	0	0
bafh Borrelia afzelii HLJ01	0	0	0	0	0	1	0	0
bbs Borrelia bissettii	0	0	0	0	0	1	1	0
btu Borrelia turicatae	0	0	0	0	0	1	1	0
bhr Borrelia hermsii	0	0	0	0	0	1	0	0
bdu Borrelia duttonii	0	0	0	0	0	0	0	0
bre Borrelia recurrentis	0	0	0	0	0	1	0	0
bcw Borrelia crocidurae	0	0	0	0	0	1	0	0
tpa Treponema pallidum subsp. pallidum Nichols	0	1	0	0	0	0	0	0
tpw Treponema pallidum subsp. pallidum Nichols	0	1	0	0	0	0	0	0
tpp Treponema pallidum subsp. pallidum SS14	0	1	0	0	0	0	0	0
tpu Treponema pallidum subsp. pallidum DAL-1	0	1	0	0	0	0	0	0
tph Treponema pallidum subsp. pallidum Chicago	0	1	0	0	0	0	0	0
tpo Treponema pallidum subsp. pallidum Mexico A	0	1	0	0	0	0	0	0
tpc Treponema pallidum subsp. pertenue CDC2	0	1	0	0	0	0	0	0
tpg Treponema pallidum subsp. pertenue Gauthier	0	1	0	0	0	0	0	0
tpm Treponema pallidum subsp. pertenue SamoaD	0	0	0	0	0	0	0	0
tpb Treponema pallidum Fribourg-Blanc	0	0	0	0	0	0	0	0
tde Treponema denticola	0	0	0	0	0	0	0	0
tsu Treponema succinifaciens	0	1	0	0	0	0	0	1
tbe Treponema brennaborense	1	1	1	1	0	0	1	0
taz Treponema azotonutricium	0	1	0	0	0	0	1	0
tpi Treponema primitia	0	1	0	0	0	0	0	1
tpl Treponema paraluiscuniculi	0	0	0	0	0	0	0	0

Tabla 3 - Tabla comparativa donde se muestra la presencia/ausencia de proteínas homólogas codificadas por los genes pertenecientes al operon nan y de tres genes relacionados (neuA, neuB, neuC). Se analizaron las especies de *Escherichia coli* (07:k1 CE10 y K-12), junto con un grupo de espiroquetas, entre las que se encuentran 16 especies del género *Borrelia*, 18 especies del género *Treponema, 8 especies de Spirochaetas, 7 especies de Leptospiras, 7 especies de Brachyspiras* y la especie *Turneriella prava*. El criterio utilizado en esta clasificación es marcando con cero (0) la ausencia de los genes homólogos y con uno (1) la presencia de genes ortólogos y presumiblemente codificantes de proteínas homólogas cumpliendo similares funciones.

CODIGO NOMBRE	ncuA	ncuB	ncuC	nanA	nanT	nanE	nanK	dapA
tcd Treponema caldaria	0	0	0	0	0	0	0	1
Tpe Treponema pedis	1	1	1	0	0	0	0	0
ssm Spirochaeta smaragdinae	1	1	0	0	0	0	1	1
sta Spirochaeta thermophila DSM 6192	0	0	0	0	0	0	0	0
stq Spirochaeta thermophila DSM 6578	0	1	0	0	0	0	1	1
scc Spirochaeta coccoides	0	0	0	0	0	0	1	1
scd Spirochaeta caldaria DSM 7334	1	1	1	0	0	0	1	0
sfc Spirochaeta africana	0	1	0	0	0	0	0	1
lil Leptospira interrogans serovar Lai 56601	1	1	1	0	1	0	1	1
lie Leptospira interrogans serovar Lai IPAV	1	1	1	0	1	0	1	1
lic Leptospira interrogans serovar Copenhageni	1	1	1	0	1	0	1	1
lbj Leptospira borgpetersenii JB197	1	1	1	0	1	0	1	1
lbl Leptospira borgpetersenii L550	1	1	1	0	1	0	1	1
lbi Leptospira biflexa serovar Patoc 1 (Paris)	1	1	0	0	1	0	1	1
lbf Leptospira biflexa serovar Patoc 1 (Ames)	1	1	0	0	1	0	1	1
tpx Turneriella parva	1	1	1	0	0	0	1	1
bhy Brachyspira hyodysenteriae	1	1	0	1	1	0	1	1
brm Brachyspira murdochii	1	1	1	1	1	0	1	1
bpo Brachyspira pilosicoli 95/1000	1	1	0	1	1	1	1	1
bpj Brachyspira pilosicoli B2904	1	1	0	1	0	1	1	1
bpip Brachyspira pilosicoli P43/6/78	1	1	0	1	1	0	1	1
bpw Brachyspira pilosicoli WesB	1	1	0	1	0	0	1	1
bip Brachyspira intermedia	1	1	0	1	1	0	1	1
sbu Spirochaeta sp. Buddy	0	0	1	1	0	0	1	0
sgp Sphaerochaeta pleomorpha Grapes	0	0	1	1	0	0	1	0

Árbol filogenético de las especies del filo *Spirochaetes* incluido el género *Leptospira*

Los resultados obtenidos del OrthoMCL se resumen en la tabla 3. Además de *E. coli,* este trabajo fue realizado sobre otras 57 especies de espiroquetas. Analizando la cuarta columna (correspondiente al gen nanA) se aprecia que, tanto en la mayoría de las cepas evaluadas de las especies del género *Treponema* (con excepción de la especie *Treponema brennaborense*), como en las distintas cepas de las especies de los géneros *Borrelia y Spirochaeta*, en el microorganismo modelo no se presentan homólogos al gen nanA. Esto al menos aparentemente, ya que no fueron detectados por el OrthoMCL, ni a través del "pathway" del KEGG.

En el Anexo 2 se presentan los resultados obtenidos del software T-coffee, mostrando los alineamientos de las proteínas homólogas presuntamente codificadas por nanA, con identidades del orden del 97%. En el organismo modelo (*E.coli* 07:k1 CD10), el sitio activo de la enzima se ubica en la posición 161. La ubicación del mismo varía en algunas especies, pero siempre se asocia con la presencia de un residuo de lisina.

La evidencia más concluyente de la historia evolutiva de genes homólogos es el resultado del análisis filogenético sobre el linaje de ese gen. Los genes que se encuentran dentro del mismo clado (conjunto de especies emparentadas con un antepasado común que forman una única rama en el árbol filogenético) serían homólogos, ya que descienden del mismo ancestro. El patrón de divergencia genética puede utilizarse para inferir las relaciones existentes entre especies.

En la construcción de los árboles filogenéticos a continuación presentados el apoyo estadístico de los nodos está basado en la probabilidad a posteriori. Esta última es una medida de la certidumbre de los grupos inferidos. Usando como umbral el 95% de probabilidad posterior (esto significa que si el valor es inferior a 0.95 los datos no permitirán diferenciar entre topologías o agrupamientos alternativas) valores iguales o mayores a 0.95 se consideran un fuerte soporte del "branching point" o del nodo. Es decir, un fuerte apoyo a la topología y a la separación anteriormente mencionada.

Se utilizó el árbol filogenético de ARNr 16s (Woose et al., 1990) del género *Leptospira* y filo Spirochaetes, usando análisis bayesiano como referencia en la discusión. Debido a la antigüedad del proceso de síntesis de proteínas, el RNA es una excelente molécula para discernir relaciones evolutivas entre organismos vivientes. Tiene distribución universal y cumple una misma función. Cambia muy lentamente y posee regiones altamente conservadas en todos los organismos, permitiendo realizar comparaciones entre organismos muy o muy poco relacionados (fig. 6).

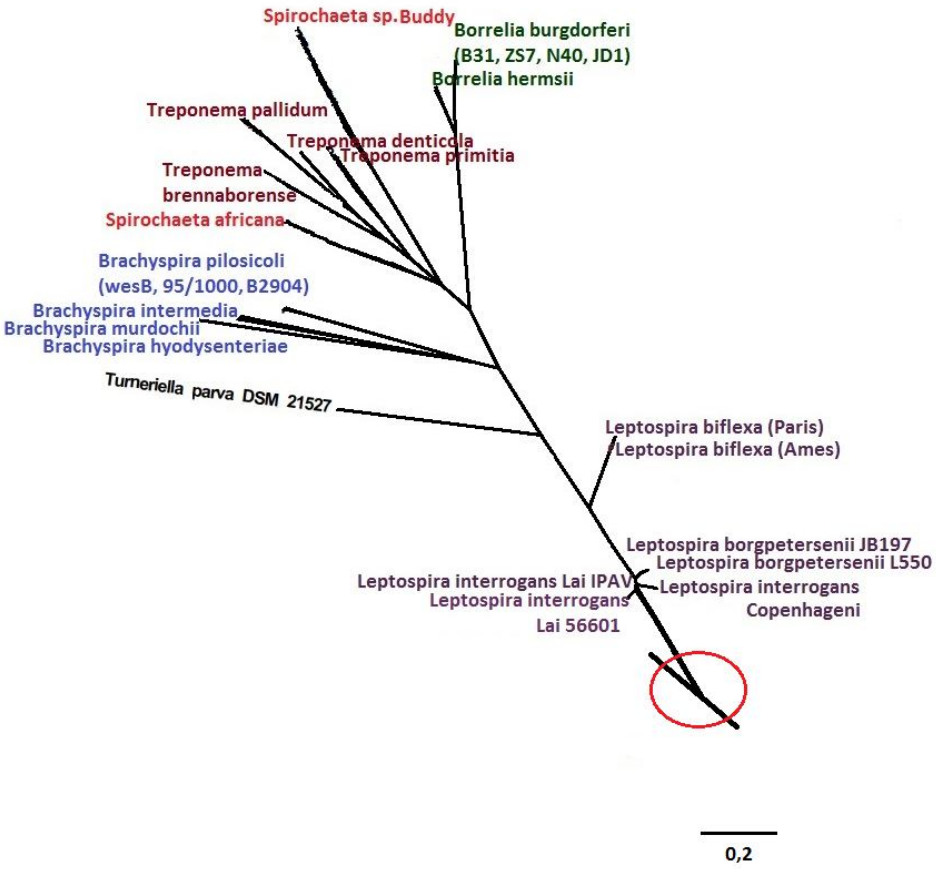


Figura 6 - Árbol filogenético de ARNr 16s del género *Leptospiras* incluido en el filo *Spirochaetes*, obtenido utilizando análisis bayesiano. Se marca con un círculo la localización de la posible raíz del filo.

En el árbol filogenético de ARNr 16S de espiroquetas se utilizó el método de máxima verosimilitud, el cual optimiza la función de verosimilitud con respecto a la topología y el largo de ramas del árbol y a un modelo de evolución nucleotídica. La distancia entre dos secuencias representa la suma del tiempo evolutivo transcurrido desde la separación de su ancestro común más cercano.

Analizando el árbol filogenético de ARNr 16s del filo Spirochaetes y del género *Leptospira* (fig.6), se aprecia que el árbol está enraizado, por lo cual se pueden establecer algunas consideraciones importantes acerca de las relaciones entre los géneros. Para esta labor nos basamos en el sistema de tres dominios, clasificación biológica propuesta por Carl Woese y colaboradores en 1990, quien clasifica el árbol de la vida en tres grupos primarios: Bacteria, Archaea y Eucarya (fig. 7). Considerando que una de las ramas del árbol de la vida fue asignada al filo *Spirochaetes* (constituido en familias, todas incluidas en un único orden *Spirochaetales* formado por varios géneros, entre los que se encuentra el género *Leptospira* observado en profundidad en

la figura 6), podemos concluir que las tres especies del género *Leptospira* evaluadas en la Tesis *(L.Interrogans, L.borgpetersenii y L.biflexa*) son bastante ancestrales dentro de las *Espiroquetas*.

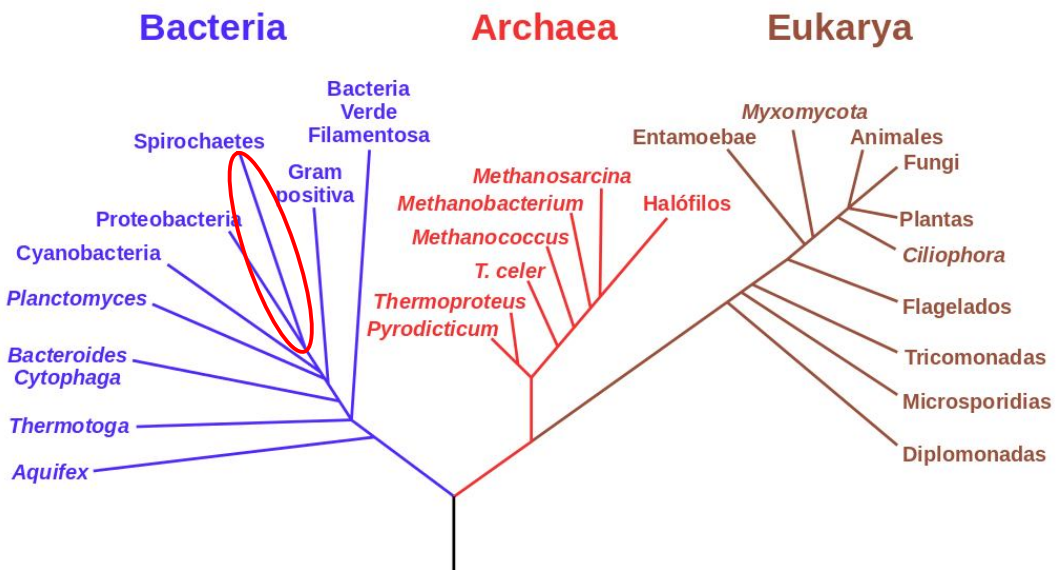


Figura 7 - Árbol filogenético de la vida, clasificación de Carl Woose, quien se basó en las diferencias encontradas en la secuencia del ARN ribosomal de la subunidad menor (ARNr 16s), proponiendo el sistema de tres Dominios. Marcado con una elipse se distingue el filo Spirochaetes.

Es habitual considerar la distancia entre dos especies como una medición del número de años o generaciones transcurridas desde la subdivisión de la especie ancestral común a las especies en cuestión. Se observa que los árboles generados por el software implementado quedan ultramétricos debido al modelo que utiliza BEAST, el cual estima un árbol enraizado (en lugar de otros métodos). En el análisis se utiliza un reloj molecular (técnica para datar la divergencia de dos especies deduciendo el tiempo transcurrido a partir del número de diferencias entre dos secuencias de ADN), cuya flexibilidad permite cuantificar la variación de la tasa evolutiva a lo largo del árbol.

Árbol filogenético de la enzima Liasa del ácido N-acetilneurámico (codificada por nanA).

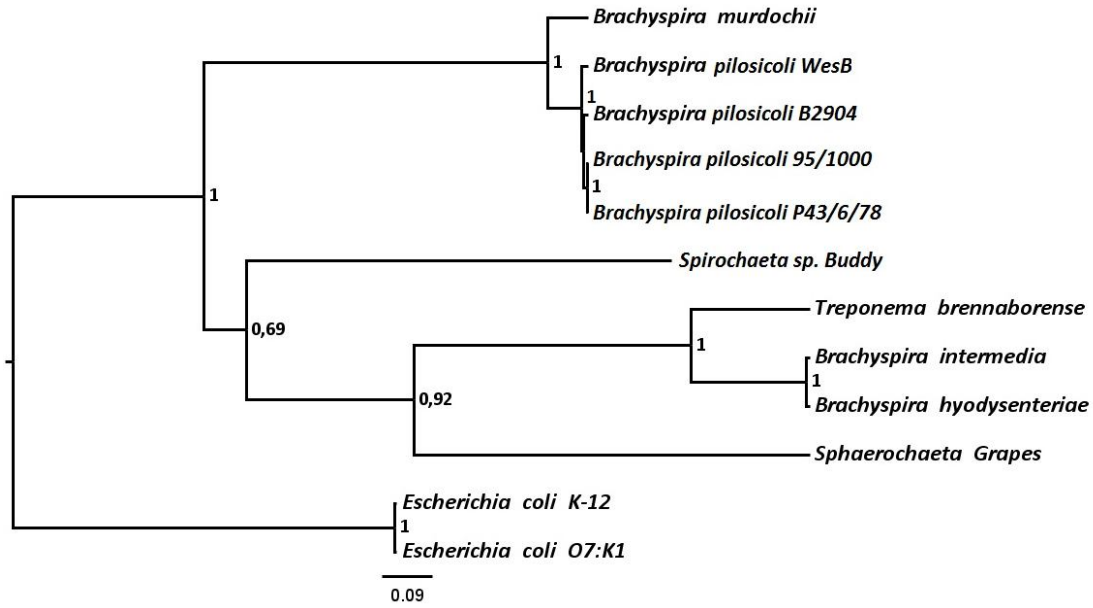


Figura 8- Árbol construido a partir de la enzima liasa del ácido N-acetilneurámico (codificada por nanA). El árbol muestra la probabilidad a posteriori de los clados, mientras que la escala que aparece debajo de la figura representa la distancia de las ramas medidas en sustituciones por sitio.

El árbol filogenético construido a partir del primer gen del operon, que codifica a la enzima liasa del ácido N-acetilneurámico (nanA), muestra la presencia de dos clados (fig. 8). Uno está constituido por dos especies de *Escherichia coli* (el microorganismo modelo y K-12), con probabilidad a posteriori del 100%, mientras que el segundo se subdivide a su vez en dos agrupaciones. Por un lado se ubican cuatro especies de *Brachyspira pilosicoli* (95/1000, wesB, B2904 y P43/6/78), junto a la especie *Brachyspira murdochii,* con un respaldo estadístico del 100%. Esta topología se encuentra en la totalidad de los árboles muestreados. Por el otro lado, con un reducido soporte de 0.69, la especie *Spirochaeta sp.* Buddy se relaciona con la especie *Sphaerochaeta pleomorpha* Grapes. Esta última, con un apoyo estadístico de 0.92 (es decir, dicha topología aparece en 92 de 100 árboles muestreados), se vincula con dos especies del género *Brachyspira* (*B. hyodysenteriae y B. intermedia*) y con la especie *Treponema brennaborense,* teniendo estas una probabilidad a posteriori de 1.

Posiblemente, estas especies expresan genes homólogos a nanA, es decir, derivan de una secuencia ancestral común y evolucionaron por eventos de especiación. De acuerdo a los resultados obtenidos por la implementación de OrthoMCL y la base de datos de KEGG, todos estos genes son xenólogos a nanA de *E.coli* 07:k1 CE10.

Los resultados del múltiple alineamiento producido por T-Coffee (presentados en el Anexo 2) comprueban los resultados obtenidos anteriormente con un porcentaje de identidad del orden del 97%, la cual incluye la zona correspondiente al sitio activo de la enzima.

Los genes presentes en estas especies divergieron a partir de eventos de especiación. De acuerdo a los resultados obtenidos, los mismos podrían ser ortólogos a nanA en el microorganismo modelo *E. coli* 07:k1 CE10. Probablemente hubo un transfer horizontal (HGT) desde la especie *Treponema brennaborense* al ancestor de *Brachyspira intermedia* y *Brachyspira hyodysenteriae* (Bellgard, 2008). Este hecho se corresponde con la filogenia de las especies involucradas (fig. 6). La existencia de otra transferencia horizontal génica entre las especies *Sphaerochaeta pleomorpha* Grapes y *Treponema brennaborense* no puede ser confirmada, debido al reducido apoyo estadístico de 0.92 presentado.

Árbol filogenético de la enzima putativa Epimerasa N-acetilmanosamina-6-Fosfato (codificada por nanE)

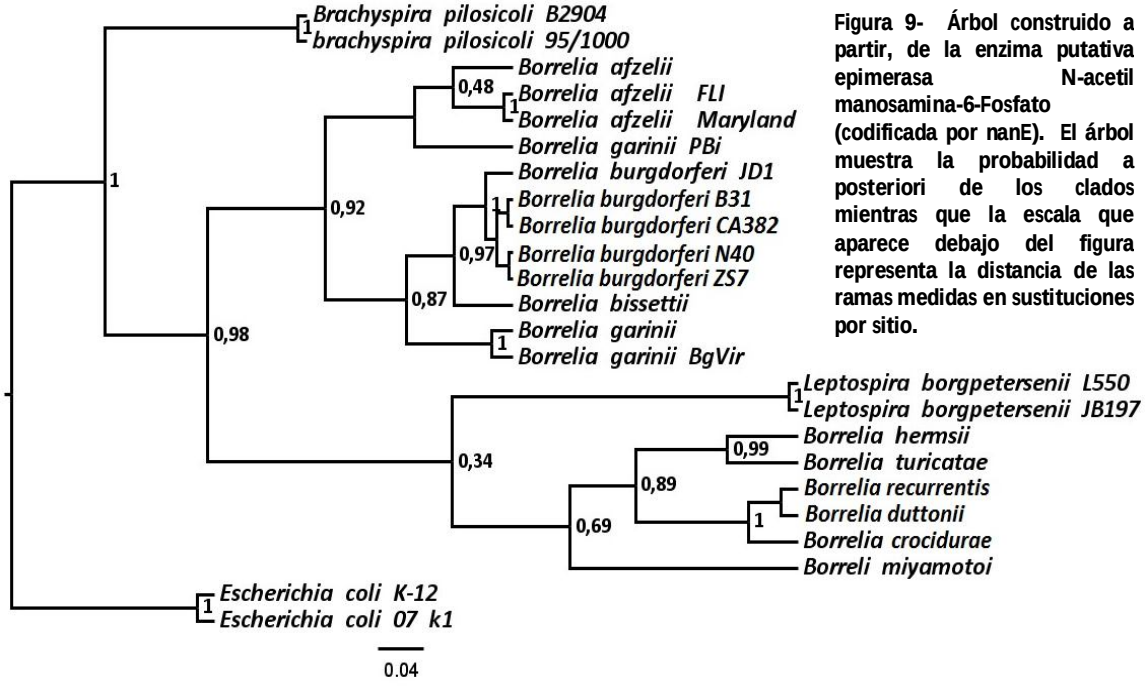


Figura 9- Árbol construido a partir, de la enzima putativa epimerasa N-acetil manosamina-6-Fosfato (codificada por nanE). El árbol muestra la probabilidad a posteriori de los clados mientras que la escala que aparece debajo del figura representa la distancia de las ramas medidas en sustituciones por sitio.

Considerando el gen nanE, el cual probablemente codifique para la epimerasa N-acetilmanosamina-6-fosfato, su situación no está totalmente esclarecida. De acuerdo a los resultados obtenidos en el microorganismo modelo *E. coli* 07:K1 CE10, observamos en la tabla 3 que genes aparentemente ortólogos al gen nanE están presentes en prácticamente la totalidad de las especies del género *Borrelia* (18

especies), dos especies del género *Leptospiras* (*L.borgpetersenii* L550 y *L. borgpetersenii* JB197), dos especies de *Brachyspira pilosicoli* (las cepas B2904 y 95/1000) y en la especie de *E. coli* K-12.

En el árbol filogenético generado a partir de la enzima presumiblemente codificada por este gen, se distinguen dos clados de probabilidad a posteriori del 100%. En uno de ellos se ubican las especies de *Escherichia coli*, el otro se subdivide en dos agrupaciones. Por un lado se ubican (con un fuerte apoyo estadístico del 100%) las dos especies del género *Brachyspira* (*B. pilosicoli* B2904 y *B. pilosicoli* 95/1000), las cuales presumiblemente poseen proteínas homólogas a la enzima codificada por el gen nanE en el microorganismo modelo. Por el otro lado, con un soporte estadístico de 0.98, se observa una subdivisión en dos clados, con un reducido soporte de 0.92 y 0.34. El primero de ellos vincula a 12 especies del género *Borrelia*, mientras que el segundo presenta dos especies de *Leptospiras borgpetersenii* (JB197 y L550) y las restantes seis especies del género *Borrelia* estudiadas en este trabajo.

Los genes presentes en estas especies divergieron a partir de eventos de especiación y de acuerdo a los resultados obtenidos, podrían ser ortólogos a nanE en el microorganismo modelo (*E. coli* 07:k1 CE10). De acuerdo a este árbol, el único posible transfer horizontal del gen nanE habría ocurrido entre las dos cepas de la especie *Leptospira borgpetersenii* (JB197 y L550) hacia el ancestor común a alguna de las seis especies de *Borrelia* (*B. hermsii, B. turicatae, B. duttonii, B. recurrentis, B. miyamotoi* y *B. crocidurae*). Debido al reducido respaldo estadístico que presenta este clado (0.34), no es factible considerar esta topología (recordemos que un fuerte apoyo es considerado para valores mayores o iguales a 0.95, siendo 1 el máximo posible).

Árbol filogenético de enzimas homólogas a la quinasa del ácido N-acetilmanosamina codificada por nanK

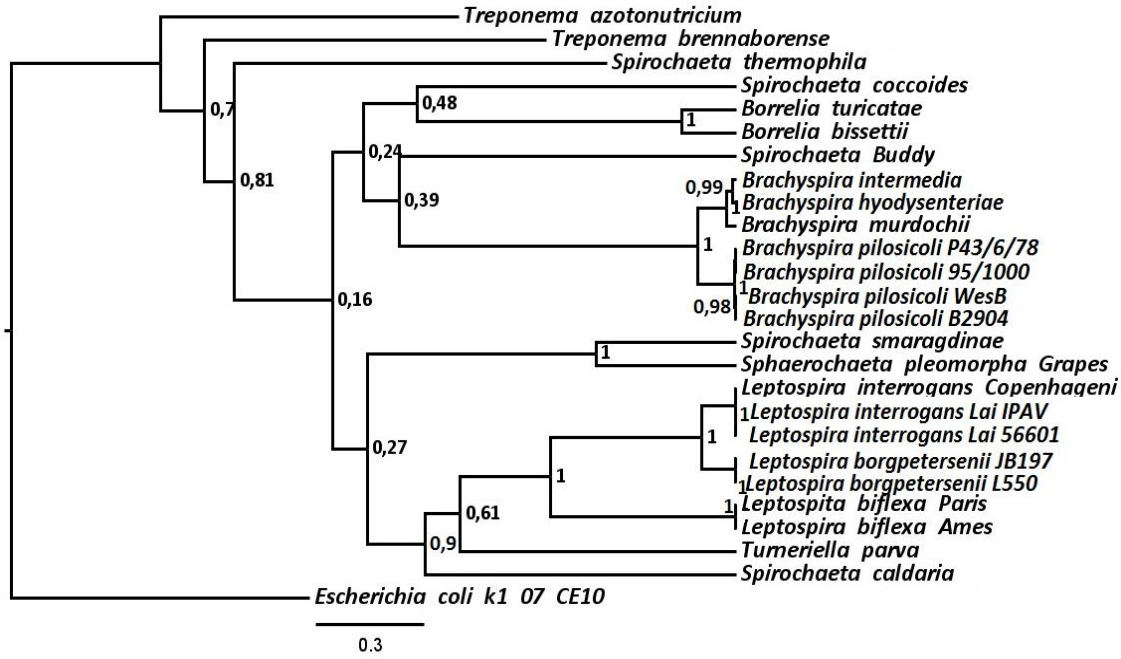


Figura 10- Árbol filogenético construido a partir de la enzima quinasa del ácido N-acetil manosamina (codificada por nanK). El árbol muestra la probabilidad a posteriori de los clados de las ramas mientras que la escala que aparece debajo de la figura representa la distancia de las ramas medidas en sustituciones por sitio.

La kinasa del ácido N-acetilmanosamina pertenece a la familia de ROK. Los resultados obtenidos del OrthoMCL (tabla 3) muestran proteínas homólogas a la del microorganismo modelo para las siete especies del género *Leptospiras* estudiadas y en dos especies del género *Borrelia* (*B. bissettii* y *B. turicatae*). Además se observa su presencia en las siete especies del género *Brachyspira* anteriormente mencionadas, en cinco especies del género *Spirochaeta* (*S. caldaria, S. smaragdinae, S. sp. Buddy, S. coccoides* y *S. thermophila*), en dos especies del género Treponema (*T. azotonutricium* y *T. brennaborense*), en la especie *Turneriella parva*, y en la especie *Sphaerochaeta pleomorpha* Grapes.

El árbol filogenético construido a partir de las proteínas homólogas a esta kinasa presenta un clado que relaciona, con reducida probabilidad a posteriori de 0.70, a dos especies del género Treponema (*T. azotonutricium* y *T. brennaborense*) con la especie *Spirochaeta thermophila.* Por el otro lado, con un bajo apoyo estadístico aparecen dos clados: En uno de ellos, con elevado respaldo en el rango de 0.98 a 1, las siete especies del género de *Brachyspira* estudiadas, las especies *Borrelia turicatae* y *Borrelia bissettii*, y dos especies del género *Spirochaeta* (*S.coccoides* y *S.sp. Buddy*). El otro clado presenta dos subgrupos con alta probabilidad a posteriori, uno de ellos relaciona

(con respaldo estadístico del 100%) a las especies *Spirochaeta smaragdinae* con la especie *Sphaerochaeta pleomorpha Grapes*, mientras que el otro subgrupo vincula (con apoyo del 96%) a la especie *Spirochaeta caldaria* con las siete especies del género *Leptospiras* estudiadas en la investigación y además con la especie *Turneriella parva.*

Los genes presentes en estas especies divergieron a partir de eventos de especiación, y de acuerdo a los resultados obtenidos, se observa que podrían ser ortólogos a nanK en el microorganismo modelo *E.coli* 07:k1 CE10. Puede postularse la ocurrencia de un transfer horizontal (HGT) desde la especie *Spirochaeta caldaria* al ancestor de alguna (o todas) de las especies del género *Leptospira* consideradas, pero este evento no tiene un suficiente apoyo estadístico debido al bajo soporte que presenta: 0.90.

Si consideramos el alineamiento por T-coffee, el resultado extraído indica que las proteínas homólogas presentan una identidad promedio de 80%.

Árbol filogenético de la proteína transmembrana del ácido siálico (codificada por nanT)

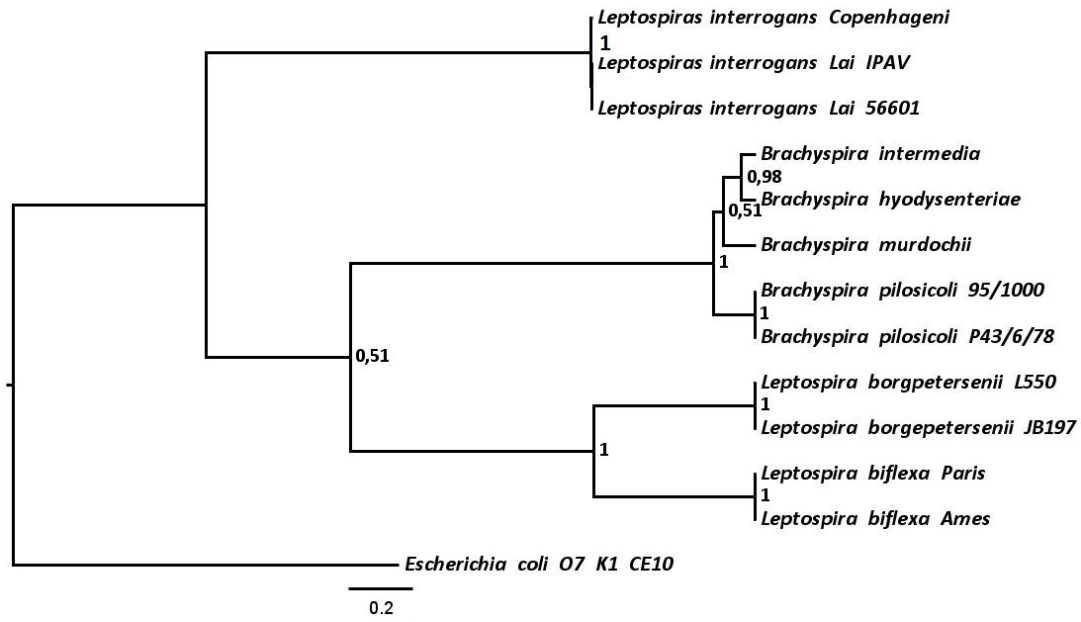


Figura 11- Árbol filogenético construido a partir de la proteína transmembrana del ácido siálico (codificada por nanT). El árbol muestra la probabilidad a posteriori de los clados, mientras que la escala que aparece debajo de la figura representa la distancia de las ramas medidas en sustituciones por sitio.

Los resultados obtenidos por OrthoMCL, los cuales se presentan en la quinta columna de la tabla 3, muestran proteínas probablemente homólogas a la proteína

transmembrana del ácido N-acetilneurámico (ácido siálico) en los siguientes microorganismos: el modelo *E. coli* 07:K1 CE10, en las siete especies del género *Leptospiras*, y cinco especies del género *Brachyspira* (*B. hyodosenteriae, B. murdochii, B. intermedia, B. pilosicoli* 95/1000 *y B. pilosicoli* P43/6/78) dentro de las 57 especies de espiroquetas estudiadas.

Al analizar el árbol filogenético construido a partir de este gen, encontramos dos clados con diferentes valores de respaldo (fig. 11). En el primero de ellos, con apoyo estadístico de 1, figuran las tres especies de *Leptospira interrogans* (Lai 56601, Lai IPAV y Copenhageni). Del segundo clado se derivan, con probabilidad a posteriori de 0.51 (topología presentada en el 51% de los árboles muestreados), cuatro especies del género *Leptospira* (dos de *L. borgpetersenii* y dos de *L. biflexa*) con el alto respaldo estadístico de 100% y las cinco especies del género *Brachyspira* arriba citadas.

Los genes presentes en estas especies divergieron a partir de eventos de especiación. De acuerdo a los resultados obtenidos estos genes podrían ser ortólogos a nanT en el microorganismo modelo *E. coli* 07:k1 CE10. Probablemente debería haber sucedido un transfer horizontal (HGT) desde alguna de las especies de *L. borgpetersenii* o *L. biflexa* al ancestor de alguna (o todas) de las especies del género *Brachyspira* consideradas. Debido al reducido respaldo estadístico presentado, inferimos que no es posible considerar este escenario evolutivo.

Árbol filogenético de la enzima Sintasa 4-hidroxi-tetrahidro picolinato, codificada por el gen dapA

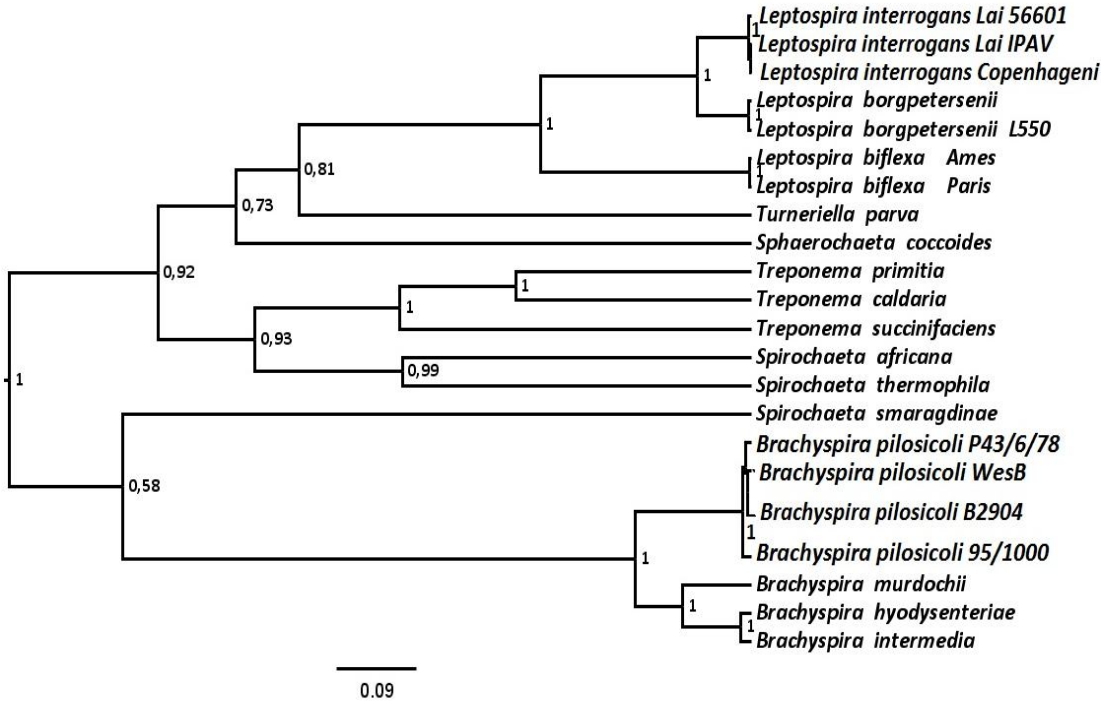


Figura 12 - Árbol filogenético de la enzima Sintasa 4-hidroxi-tetrahidro picolinato, codificada por el gen dapA. El árbol muestra la probabilidad a posteriori de los clados, mientras que la escala que aparece debajo de la figura representa la distancia de las ramas medidas en sustituciones por sitio.

Centrándonos en los resultados obtenidos por el blastp (todos contra todos), tanto del OrthoMCL (caso considerado en la última columna de la tabla comparativa 3), como del realizado en línea utilizando los recursos propios del KEGG, se evidencian dos cosas, a saber: Primero, las especies del género *Leptospira (L. interrogans, L. biflexa y L. borgpetersenii)* no tienen en su genoma el operon nan completo, tal cual se expresa en *E. coli* (07:k1 CE10). Segundo, presentan solo los genes nanE, nanT y nanK del operon nan (Almagro et al., 2009), pero expresan un gen llamado dapA que codifica para la enzima Sintasa 4-hidroxi-tetrahidro picolinato (Sintasa del ácido dihidrodipicolinato). La cual es utilizada por dichas bacterias como precursor del peptidoglicano, principal unidad funcional y estructural de la pared bacteriana.

Según el resultado del OrthoMCL (tabla 3) se observa que el gen dapA es uno de los genes de mayor conservación, al menos para las especies consideradas en este trabajo, expresándose en todas las especies del género *Leptospiras* y en todas las especies del género *Brachyspira* estudiadas. Los resultados se corroboran al analizar el alineamiento de secuencias proteicas, donde se observan porcentajes de identidad del orden de 95% (Anexo 6).

Del análisis del árbol filogenético de proteínas homólogas a la enzima Sintasa 4-hidroxi-tetrahidro picolinato (Sintasa del ácido dihidrodipicolinato) codificada por el

gen dapA, surgen dos clados (fig. 12). El primero de ellos vincula, con poca probabilidad a posteriori (0.81), a las siete especies del género *Leptospira* con la especie *Turneriella parva*, la cual a su vez se vincula con *Spirochaeta coccoides* con un reducido soporte (0.73). Este primer clado vincula también (con la elevada probabilidad a posteriori del 100%) a tres especies del género *Treponema* (*T. primitia, T. caldaria y T. succinifaciens*), con dos especies del género *Spirochaeta* (*S. thermophila y S. africana*). El segundo clado vincula con bajo apoyo a las siete especies del género *Braschyspira* estudiadas, con la especie *Spirochaeta smaragdinae*.

Los genes presentes en estas especies divergieron a partir de eventos de especiación, y de acuerdo a los resultados obtenidos podrían ser ortólogos a dapA. Probablemente habría sucedido un transfer horizontal (HGT) desde la especie *Treponema succinifaciens* al ancestor de alguna de las especies del género *Spirochaeta*, tanto *S. africana* como *S. thermophila*. El reducido soporte estadístico de 0.93 hace poco factible este evento.

Considerando el alineamiento por T-coffee, el resultado obtenido indica que las proteínas homólogas presentan una identidad promedio de 95%.

En este trabajo también se tomaron en cuenta tres genes vinculados directamente con el operon nan, los cuales ya se desarrollaron en instancias anteriores: **neuA, neuB y neuC.**

Árbol filogenético de Sintetasa Citidina-5'-monofosfato-ácido-N-acetilneurámico, codificada por (neuA)

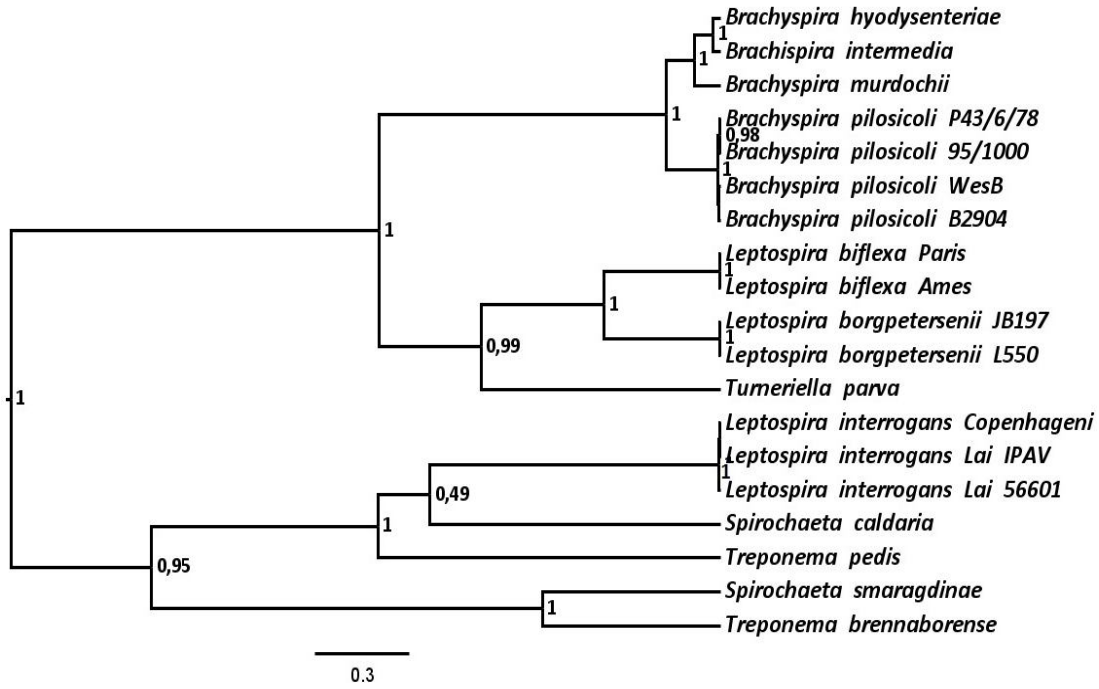


Figura 13- Árbol filogenético construido a partir de la enzima Sintetasa Citidina-5'-monofosfato-ácido-N-acetilneurámico, codificada por neuA. En la generación del presente árbol se consideró la probabilidad a posteriori de las especies involucradas, la escala que aparece debajo de la figura representa la distancia de las ramas medidas en sustituciones por sitio.

Como lo muestra la figura 3, sobre la vía de ácido siálico influyen directamente otras enzimas como la ''Sintetasa Citidina-5'-monofosfato-ácido-N-acetilneurámico'' (se coloca entre comillas ya que es una posible caracterización, su estructura no ha sido todavía establecida). La misma es codificada por neuA. Esta enzima juega un papel importante en la activación de los ácidos siálicos a sus compuestos cíclicos derivados, para su posterior utilización en las reacciones de transferencia del sialato.

Los resultados extraídos del OrthoMCL (Tabla 3) presentan varias especies con posibles proteínas homólogas a esta enzima, como las siete especies del género *Leptospira* estudiadas en este ensayo y las siete especies del género *Brachyspira* consideradas. Asimismo, dos especies del género *Treponema* (*T. pedis* y *T. brennaborense*), dos especies del género *Spirochaeta* (*S. smaragdinae* y *S. caldaria*) y la especie *Turneriella parva.* La ubicación metabólica de dicha enzima es próxima a la vía del ácido N-acetilneurámico. El polímero enzimático codificado por neuA, ha sido clasificado como un homopolímero con enlaces internos glucosídicos β [2,8].

En el estudio analítico del árbol filogenético construido a partir del gen neuA se aprecian dos clados, ambos con fuertes soportes estadísticos (1 y 0.95,

respectivamente). El primero se bifurca a su vez en otros dos clados uno de los cuales relaciona a las siete especies del género *Brachyspira,* mientras que el otro vincula a las dos especies de *L. borgpetersenii* (L550 y JB197) y *L. biflexa* (Ames y Paris) con la especie *Turneriella parva*. Presentándose dicha topología en el 99% de los árboles muestreados. El segundo clado ubica, con probabilidad a posteriori de 0.95, a las especies *Treponema brennaborense* y *Spirochaeta smaragdinae*, vinculándolas con *Treponema pedis, Spirochaeta smaragdinae* y las tres especies de *Leptospira interrogans* (Lai 56601, Lai IPAV y Copenhageni).

Los genes presentes en estas especies divergieron a partir de eventos de especiación, y de acuerdo a los resultados obtenidos podrían ser ortólogos a neuA. Probablemente haya sucedido un transfer horizontal desde la especie *Turneriella parva* al ancestor de alguna de las especies de *Leptospira borgpetersenii* (L550 y JB197) y *Leptospira biflexa* (Paris y Ames). Aunque el soporte estadístico de este evento de transferencia es fuerte (0.99), el mismo se opone a la filogenia del árbol realizado del ARNr 16s de espiroquetas, lo que lo hace poco factible.

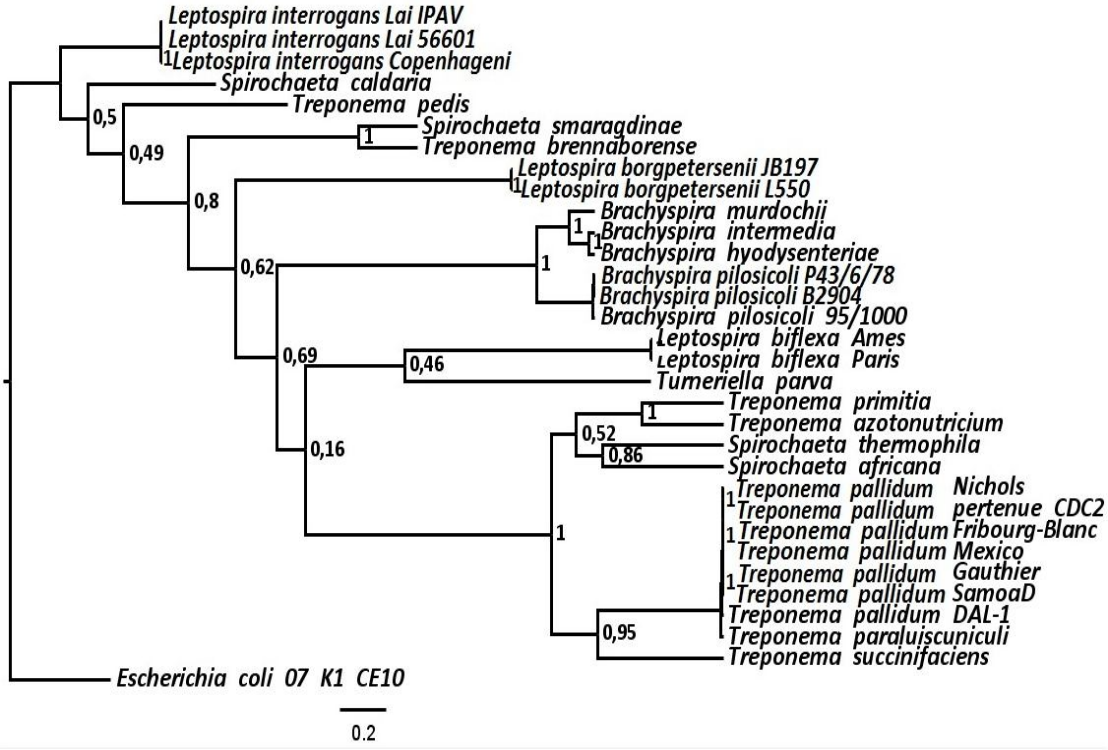
Sumado a esto, los resultados del alineamiento tampoco arrojan demasiada luz sobre el tema, obteniéndose porcentajes de identidad en el rango del 82%.

Árbol filogenético construido a partir de la enzima Sintetasa del ácido N-acetilneurámico codificada por el gen neuB.

Se ha podido identificar la enzima codificada por el gen neuB mediante la espectrometría de masas (EM), utilizando ionización por electrospray (ESI). Se trata de un homotetrámero (probablemente compuesta de dos dímeros), cuya forma activa primaria se cree que es tetramérica. Dicha enzima se identificó como Sintetasa del ácido N-acetilneurámico (sintetasa de ácido siálico) y participa en la biosíntesis del polisacárido del ácido α [2-8] poli-N-acetilneurámico en *E. coli* 07:k1 CE10.

Los resultados del OrthoMCL (Tabla 3) permiten ubicar posibles enzimas homólogas a la codificada por este gen, tanto en el microorganismo modelo utilizado, como en las siete especies del género *Leptospira* analizadas, en las siete especies del género *Brachyspira* estudiadas, en trece especies del género *Treponema,* en cuatro especies del género *Spirochaeta* y en la especie *Turneriella parva*.

Figura 14- Árbol filogenético construido a partir de la enzima Sintetasa del ácido N-acetilneurámico codificada por el gen neuB. Este árbol fue confeccionado según probabilidad a posteriori de las especies involucradas. La escala que aparece debajo de la figura representa la distancia de las ramas medidas en sustituciones por sitio.



Analizando el árbol filogenético construido a partir del gen neuB (fig. 14), se observa un clado del cual surgen otros clados sucesivos de reducido soporte estadístico. Estos vinculan al resto de las especies que aparentemente presentan proteínas homólogas al gen neuB en el microorganismo modelo *E. coli* 07:K1 CE10, con las tres especies del género de *Leptospira interrogans* analizadas en el estudio realizado (Lai 56601, Lai IPAV y Copenhageni). En el primero de los clados, con un bajo respaldo de 0.5 y 0.49, figuran las especies *Spirochaeta caldaria* y *Treponema pedis*, las cuales se relacionan con las especies *Treponema brennaborense* y *Spirochaeta smaragdinae*. En los clados siguientes figuran, con un reducido apoyo de 0.62, las dos especies de *Leptospira borgpetersenii* (L550 y JB197), relacionadas con las siete especies del género *Brachyspira* estudiadas. Posteriormente aparecen, con un soporte casi nulo de 0.16, las especies de *Leptospira biflexa* (Paris y Ames), vinculadas con dos especies del género *Spirochaeta* (*S. Thermophila y S. africana*) y con otras 11 especies del género *Treponema*, estas últimas descritas de la siguiente forma: 7 cepas de la especie

Treponema pallidium, y las especies *T. paraluiscuniculi, T. succinifaciens, T. azotonutricium* y *T. primitia.*

Los genes presentes en estas especies divergieron a partir de eventos de especiación, y de acuerdo a los resultados obtenidos podrían ser ortólogos a neuB. Con el fin de corroborar los resultados se realizó un múltiple alineamiento secuencial, dando como resultado (Anexo 6) un porcentaje de identidad del orden del 88%.

Árbol filogenético construido a partir de la enzima epimerasa codificada por el gen neuC.

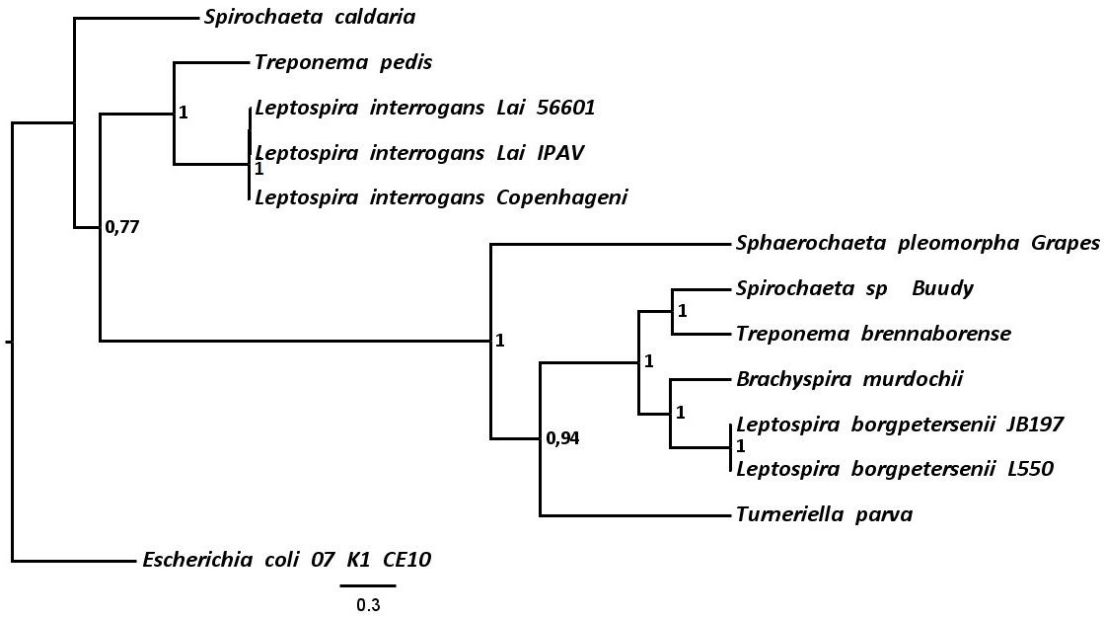


Figura 15- Árbol filogenético construido a partir de la enzima epimerasa codificada por el gen neuC. La confección del presente árbol fue realizada considerando la probabilidad a posteriori de las especies involucradas. La escala que aparece debajo de la figura representa la distancia de las ramas medidas en sustituciones por sitio.

Es sabido que el gen neuC codifica para una epimerasa bacteriana en *Escherichia coli* 07:k1 CE10. También, su cápsula es un importante factor de virulencia, consistiendo en un polímero homogéneo lineal ligado al ácido N-acetilneurámico.

Con referencia a la enzima codificada por el gen neuC, los resultados obtenidos por OrthoMCL (Tabla 3) ubican una homóloga de la misma en dos especies del género *Treponema* (*T. brennaborense* y *T. pedis*) y en las tres especies de *Leptospiras interrogans* (Lai 56601, Lai IPAV y Copenhageni) consideradas en el estudio. También se hallaron proteínas homólogas en las dos especies de *Leptospira borgpetersenii* (JB197, L550), en dos especies del género *Spirochaeta (S. caldaria* y *S. sp. Buddy*), en

Sphaerochaeta pleomorpha Grapes, en una especie del género *Brachyspira (B. murdochii),* y en la especie *Turneriella parva.*

En el análisis realizado sobre el árbol filogenético construido a partir de este gen (fig. 15), se aprecian dos clados con reducido soporte estadístico (0.77). El primero es un clado donde figuran las tres especies de *Leptospira interrogans* (Lai 56601, Lai IPAV y Copenhageni), junto con la especie *Treponema pedis* (con una probabilidad a posteriori del 100%). El segundo clado vincula a *Sphaerochaeta pleomorpha* Grapes (con un fuerte respaldo estadístico de 1), a un clado en donde figuran varias especies con un respaldo estadístico de 0.94. Las mismas se agrupan relacionando a *Treponema brennaborense* y *Spirochaeta sp*. Buddy con probabilidad a posteriori del 100%. Por el otro lado, *Brachyspira murdochii* aparece vinculada con un fuerte apoyo estadístico de 1, a las dos especies de *Leptospira borgpetersenii* (JB197 y L550) estudiadas.

Los genes presentes en estas especies divergieron a partir de eventos de especiación, por lo que, de acuerdo a los resultados obtenidos, podrían ser ortólogos a neuC. Probablemente se sucedieron dos transfer horizontal (HGT), uno de ellos desde *Treponema pedis* al ancestor de las especies de *Leptospira interrogans* consideradas. El otro evento de transferencia horizontal génica es probable que sucediera desde *Brachyspira murdochii* al ancestor de las especies de *Leptospira borgpetersenii* (JB197 y L550). Como los dos eventos se contradicen con la filogenia de espiroquetas (fig. 6), describen entonces escenarios evolutivos poco factibles.

DISCUSION

Como fue indicado en la Introducción de la presente Tesis, se trabajó con secuencias genómicas completas solo en el caso de *E. coli* (microorganismo modelo) y en las 7 especies de *Leptospiras*. Por lo tanto, la ausencia de genes ortólogos a la vía del ácido N-acetilneurámico en las demás especies de espiroquetas no es razón suficiente para afirmar la ausencia en esas especies de genes ortólogos a alguno de los genes del operon nan en particular. Asimismo, es interesante apreciar la existencia dentro del género Brachyspira (o al menos dentro de las especies de este género analizadas en el presente trabajo) de genes ortólogos a la casi totalidad de los genes del operon nan (a excepción del gen nanE en alguna especie). Además, las especies de Brachyspira estudiadas expresan también otros genes como dapA, vinculados a la vía metabólica del ácido siálico, en forma conservada como sucede en la mayoría del filo Spirochaetes.

Una posible explicación a este fenómeno fue presentada en 2009 por parte de S. Almagro y colaboradores, quienes demostraron que el escenario evolutivo que surge al analizar los árboles filogenéticos del operon nan, muestra una evolución en mosaico del operon, debido principalmente a transferencias horizontales de genes (HGT). Dentro del operon nan, el gen nanA, que codifica a la primera enzima del catabolismo del ácido siálico (liasa del ácido N-acetilneurámico), evoluciona en mosaico a diferencia de los genes nanE y nanK que no lo hicieron (Almagro et al., 2009). Este aporte se corresponde con los resultados obtenidos. El género *Leptospira* no posee la vía completa del ácido siálico, expresando solamente los genes nanE, nanK y nanT. Como la propiedad evolutiva en mosaico considera la transferencia horizontal de genes, se deduce que las especies del género *Brachyspira* no pudieron haber adquirido el gen nanA por transferencia vertical de su ancestro común más cercano. De estos resultados obtenidos surge la consideración siguiente: probablemente en el pasado ocurrió un evento de transferencia horizontal génica (HGT) entre algunas especies del género *Escherichia* y la especie *Treponema bernnaborense* (posiblemente también con otras especies del género *Treponema*, aunque no disponemos al momento de la información para confirmar esa aseveración). Asímismo se verifica una transferencia de genes con especies del género *Brachyspira*, al menos en todas las cepas estudiadas, aunque desconocemos el sentido de la misma.

Motro y colaboradores determinaron en 2008 la conservación de una región genómica de 26 Kb entre *Brachyspira* Intestinales, específicamente *B. hyodysenteriae* y varias cepas de *B. pilosicoli*, con especies del género *E. coli.* (Motro et al., 2008). En base a estudios previamente realizados, los investigadores han sugerido tres potenciales explicaciones para este evento de conservación génica (Tamames et al., 2001) (Bellgard, 2008). El primero involucra la región conservada con un proceso de

transferencia vertical, desde un ancestro en común (Lawrence et al., 2005). Si este fuera el caso, se esperaría que la región conservada estuviera presente en otras espiroquetas, y no en dos especies taxonómicamente no relacionadas.

Otra posibilidad es que la integridad de un clúster conservado entre distintas especies es de importancia para la funcionalidad celular mientras exprese unidades transcripcionales (clúster de genes relacionados con la eficiencia en la transcripción) (Wolf et al. 2001) y operones (clúster de genes relacionados con propósitos regulatorios) (Rocha & Danchin et al., 2003) (Pal & Hurst et al., 2004). A pesar de que se predijo que algunos genes presentes en regiones similares de *Brachyspira hyodysenteriae, Brachyspira pilosicoli* en sus distintas cepas y de *E. coli* están asociados con varios operones, no se conocían elementos reguladores ni regiones promotoras, por lo que es improbable que la región completa y todos los genes involucrados sean integrantes de un único operon (Lathe et al., 2000).

La tercera razón implica que la conservación génica presentada entre estas especies se debe a un evento de transferencia horizontal de genes (HGT). Esta posible explicación está basada en la observación de patrones similares a nivel de la composición de bases de la región conservada en las especies involucradas. Este tipo de patrón es característico de ADN adquirido (Buchrieser et al., 2003) y en particular de la presencia de elementos génicos móviles (Schmidt & Hensel et al., 2004). Más evidencias de un potencial evento HGT es la presencia de secuencias repetidas en el extremo 5' de la región conservada de ambas especies de *B. hyodysenteriae* y *B. pilosicoli.* Las secuencias repetidas son conocidas por facilitar la recombinación génica y además permiten la inserción de ADN extracelular (Schmidt & Hensel et al., 2004). Estas secuencias repetidas en genomas de bacterias, han sido identificadas por estar en proximidad a genes funcionales que comúnmente son transferidos, especialmente genes involucrados en procesos metabólicos, tolerancia a antibióticos o resistencia (Nascimento et al., 2004).

CONCLUSIONES

El objetivo de la presente Tesis consistía en la identificación de secuencias homólogas de genes de la vía del ácido siálico (operón nan) en los distintos géneros de espiroquetas, con particular interés en el género de *Leptospiras*. Para ello se buscaron patrones de presencia/ausencia de dichos genes en organismos filogenéticamente cercanos, utilizando herramientas genómicas como el software OrthoMCL y la base de datos "KEGG orthology".

Con respecto al objetivo planteado originalmente, se comprobó que las especies estudiadas del género *Leptospira* poseen ortólogos a tres de los cuatro genes relevantes del operon nan (nanE, nanK y nanT) en relación al microorganismo modelo

Escherichia coli (07:k1 CE10). Es probable que el género *Leptospira* carezca de la enzima responsable del catabolismo del ácido siálico (liasa del ácido N-acetilneurámico), por carecer de ortólogo al gen nanA.

Al analizar la filogenia del filo Spirochaetes de ARNr 16s, se puede concluir que al menos las tres especies del género *Leptospira* evaluadas en la tesis *(L. Interrogans, L. borgpetersenii y L. biflexa*) son bastante ancestrales dentro de las *Espiroquetas*. Por lo anteriormente mencionado, la adquisición del gen nanA por parte de algún género de espiroquetas (por ejemplo, las diversas especies de *Brachyspira* analizadas), no habría sido por transferencia vertical sino por un evento de transferencia horizontal de genes (HGT). El origen y sentido de la transferencia es incierto y escapa al alcance del presente trabajo.

ANEXO 1 - Implementación del software OrthoMCL

El archivo "orthomcl.config.template" le provee al programa, OrthoMCL, la información requerida para su configuración, por lo que se lo edita creando un nuevo directorio (my_orthomcl_dir), que lo incluya y que además sirva para almacenar los resultados de la corrida del Orthomcl.

Se sugiere por parte de los creadores del software, editar el archivo .config.template, utilizando determinadas propiedades de valor:

*dbVendor= indistintamente puede ser oracle o Mysql (se utilizó una base de datos de un servidor Mysql), será utilizada por los comandos, orthomclInstallSchema, orthomclLoadBlast, orthomclPairs del programa.
*dbConnectString= dbi:MySql:orthomcl
dbi:MySql:orthomcl:localhost:5500 (por ejemplo, para un usuario que tiene instalado un servidor MysQL en el puerto 5500)
*se requiere información de logueo del operario, requisito utilizado con comandos del programa, como orthomclInstallSchema, orthomclLoadBlast, orthomclPairs, orthomclDumpPairsFiles
*similarSequencesTable= SimilarSequences, referido al nombre de la tabla donde se cargará los resultados liberados por Blastp atreves del utilitario orthomclLoadBlast, información también levantada por orthomclInstallSchema, orthomclLoadBlast y orthomclPairs.
*orthologTable= Ortholog, constituirá el nombre de la tabla que contendrá los potenciales pares de genes ortólogos.
*inParalogTable= InParalog, se establece como el nombre de la tabla que almacenará los potenciales pares de genes parálogos.
*coOrthologTable= CoOrtholog es el nombre de la tabla que va a contener los potenciales pares coortólogos.
*interTaxonMatchView= InterTaxonMatch, indica el porcentaje mínimo de similitud que deben contener los proteomas,(50%), es decir que los resultados del Blastp, donde se obtenga porcentajes de identidad, menor a este valor, son ignorados.
*se utiliza el comando orthomclPairs , con un exponente de e-value=- 5, (el valor recomendado es de -5 esto significa que todos los pares encontrados por el programa tendrán una puntuación en el BLASTP con valores menores o iguales de 1e-5 (en ambas direcciones).

ANEXO 2 - gen nanA

Resultados de la alineación múltiple secuencial realizada por T-Coffee entre posibles proteínas homólogas a la enzima codificada por nanA (liasa del ácido N-acetilneurámico) de dos especies de *Escherichia coli* (07:k1 CE10 y K-12), de una especie *Treponema brennaborense* y de siete especies del género *Brachyspira* divididas de la siguiente forma, cuatro especies de *B. pilosicoli* (95/1000, WesB, B2904 y P43/6/78), además de *B. hyodysentenariae, B. intermedia y B. murdochii.*

ANEXO 3 - gen nanE

Resultados del alineamiento secuencial realizado entre probables proteínas homólogas a la enzima codificada por nanE (Putativa epimerasa del ácido N-acetilmanosamina 6-fosfato), del microorganismo modelo *Escherichia coli* 07:k1 CE10, de dos especies de *Leptospiras borgpetersenii* (JB197 y L550), de dos especies del género *Brachyspira* (*B. pilosicoli* B2904 y *B. pilosicoli 95/1000)* y de 18 especies del género *Borrelia.*

ANEXO 4 - gen nanT

Resultados del alineamiento múltiple secuencial efectuado entre proteínas homólogas a la enzima codificada por nanT (Proteína transportadora de ácido N-acetilneurámico) de la especie *Escherichia coli* 07:k1 CE10, de siete especies del género *Leptospiras* (distribuidas en tres especies de *L. interrogans*, dos especies de *L. borgpetersenii* y dos especies de *L. biflexa*). Además de cinco especies del género *Brachyspira* (*murdochii, pilosicoli* P43/6/78, *pilosicoli* 95/1000*, intermedia y hyodysentenariae).*

ANEXO 5 - gen nanK

Resultados del alineamiento múltiple secuencial efectuado entre posibles proteínas homólogas a la enzima codificada por nanK (Kinasa del ácido N-acetilmanosamina) de la especie *Escherichia coli* 07:k1 CE10, de dos especies del género *Treponema* (*T. azotonutricium* y *T. brennaborense*), de siete especies del género *Leptospiras* (distribuidas en tres especies de *L. interrogans*, dos especies de *L. borgpetersenii* y dos especies de *L. biflexa*). Además de siete especies del género *Brachyspira* (*B. murdochii, B. pilosicoli* P43/6/78, *B. pilosicoli* 95/1000*, B. pilosicoli* B2904*, B. pilosicoli* WesB*, B. intermedia y B. hyodysentenariae*), de cuatro especies del género *Spirochaeta* (*S. coccoides, S. thermophila, S. caldaria* y *S. smaragdinae*), de la especie *Turneriella parva* y de la especie *Sphaerochaeta pleomorpha* Grapes.

ANEXO 6 - gen dapA

Resultados del alineamiento múltiple secuencial realizado entre proteínas homólogas a la enzima Sintetasa 4-hidroxi-tetrahidro picolinato (Sintetasa del ácido dihidrodipicolinato) codificada por el gen dapA, de las siete especies del género *Leptospira* involucradas en el estudio realizado, del conjunto de todas las especies del género *Brachyspira* hasta el momento estudiado (siete especies), en el presente trabajo. Además el alineamiento secuencial de las proteínas homólogas a la codificada por dapA, de la especie *Turneriella parva,* de tres especies del género *Spirochaeta* (*S. thermophila, S. smaragdinae* y *S. africana*) de tres especies del género *Treponema* (*T. primitia, T. succinifaciens* y *T. caldaria*) y de la especie *Sphaerochaeta coccoides.*

ANEXO 7 - gen neuA

Resultados del alineamiento múltiple secuencial realizado entre proteínas homólogas a la enzima Sintetasa Citidina-5'-monofosfato-ácido-N-acetilneurámico codificada por (neuA), de las siete especies del género *Leptospira* involucradas en la investigación, de las siete especies del género *Brachyspira,* citadas anteriormente, de dos especies del género *Treponema* (*T. brennaborense* y *T. pedis*), de las especies *Turneriella parva* y *Spirochaeta caldaria.*

ANEXO 8 - gen neuB

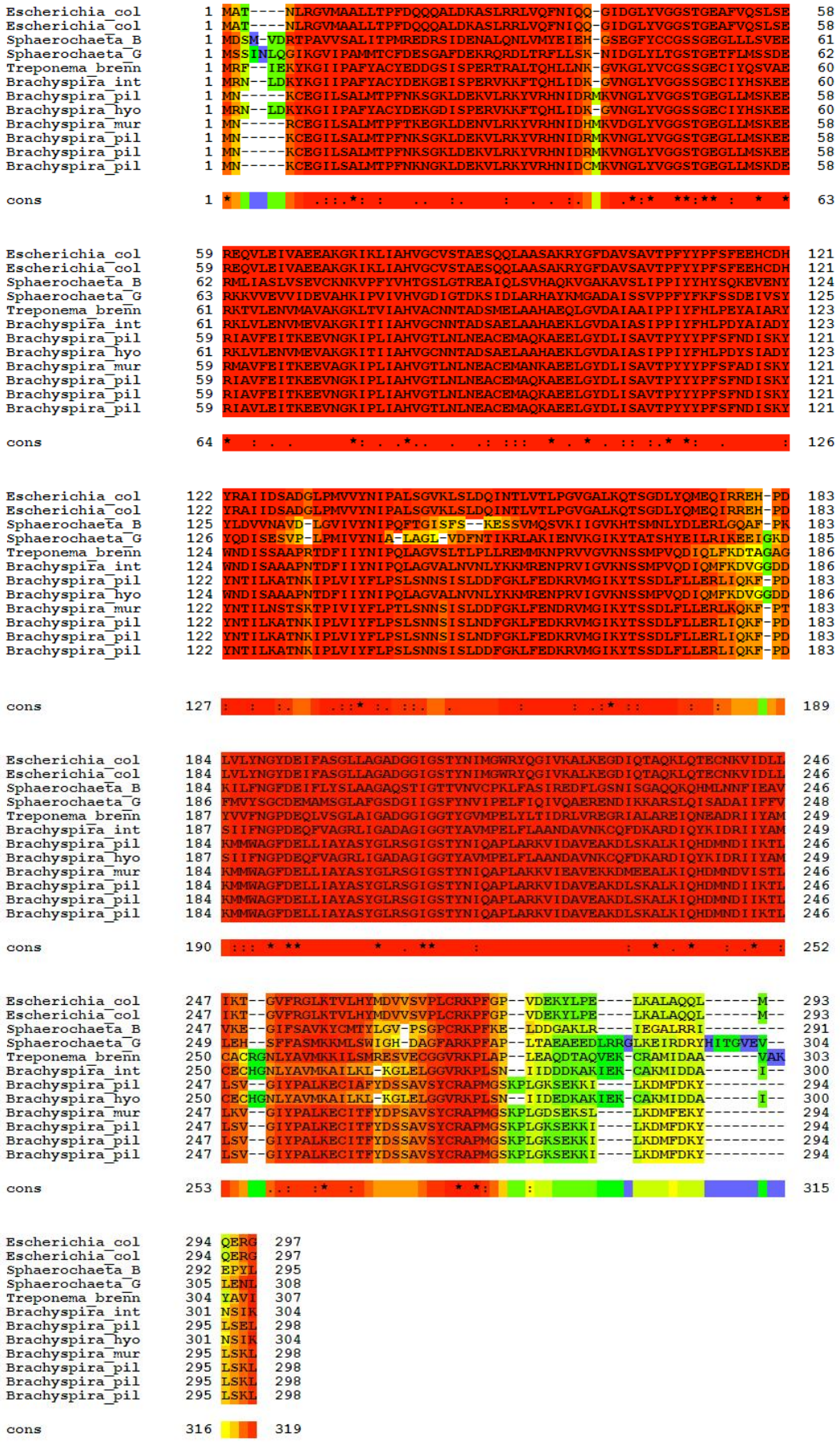
Resultados del alineamiento secuencial realizado entre las proteínas homólogas a la posible enzima codificada por neuB (Sintetasa del ácido N-acetilneurámico) del microorganismo modelo *Escherichia coli* 07:K1 CE10, de siete especies del género *Leptospira* estudiadas, de siete especies del género *Brachyspira* nombradas con anterioridad. Además el alineamiento secuencial contiene posibles proteínas homólogas de 13 especies del género *Treponema* (distribuidas de la siguiente manera, ocho cepas de la especie *T. pallidum,* así también como *T. brennaborense, T. pedis, T. succinifaciens, T. primitia y T. azotonutricium*), de *Turneriella parva* y de cuatro especies del género *Spirochaeta* (*S. thermophila, S. smaragdinae, S. caldaria* y *S. africana*).

ANEXO 9 - gen neuC

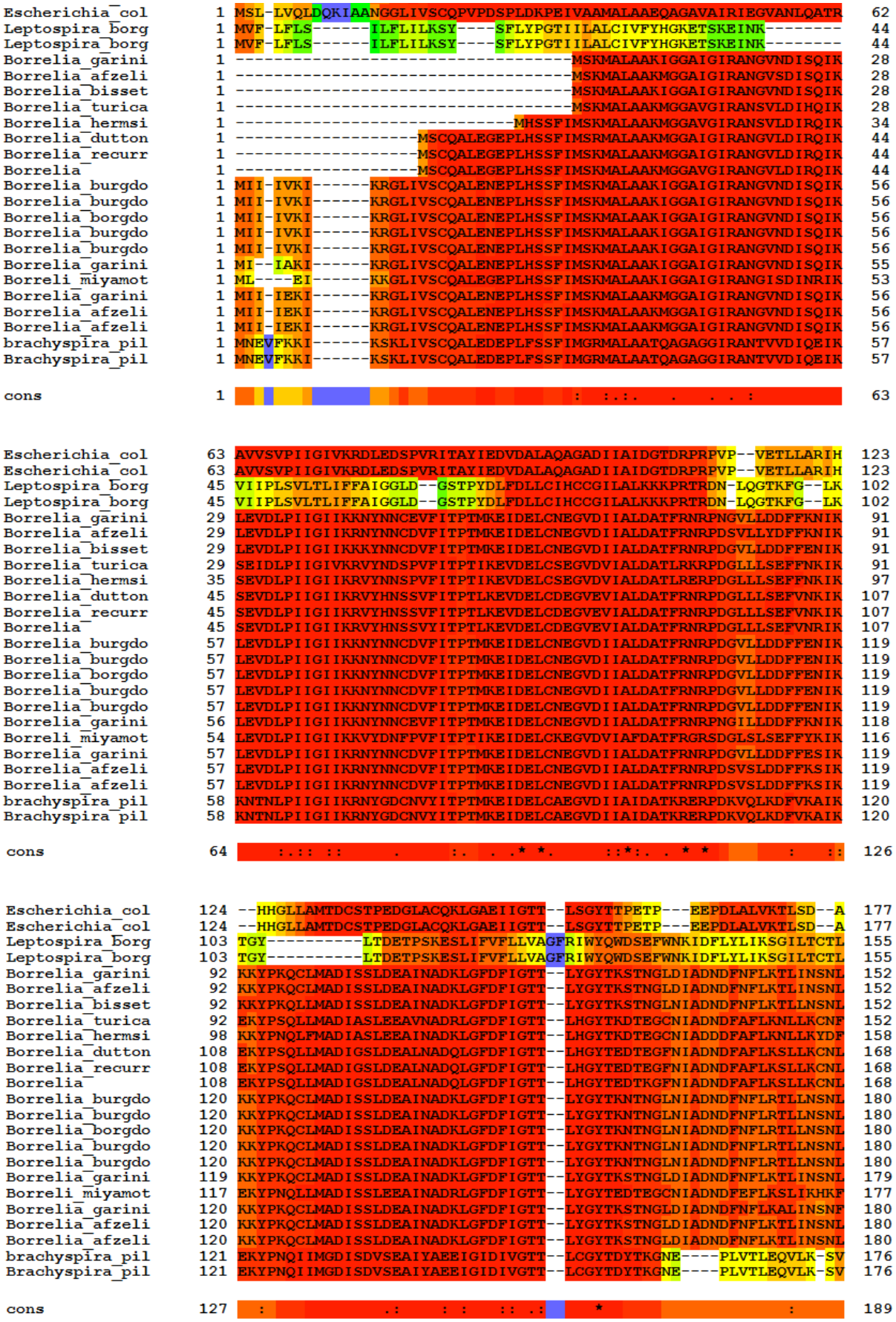
Resultados del alineamiento secuencial realizado en T-Coffee entre las proteínas homólogas a la probable enzima codificada por neuC (Epimerasa UDP-N-acetilglucosamina), de la especie *Escherichia coli* (07:K1 CE10), de cinco especies del género *Leptospira e*studiadas, tres de ellas son *L. interrogans Lai, L. interrogans*

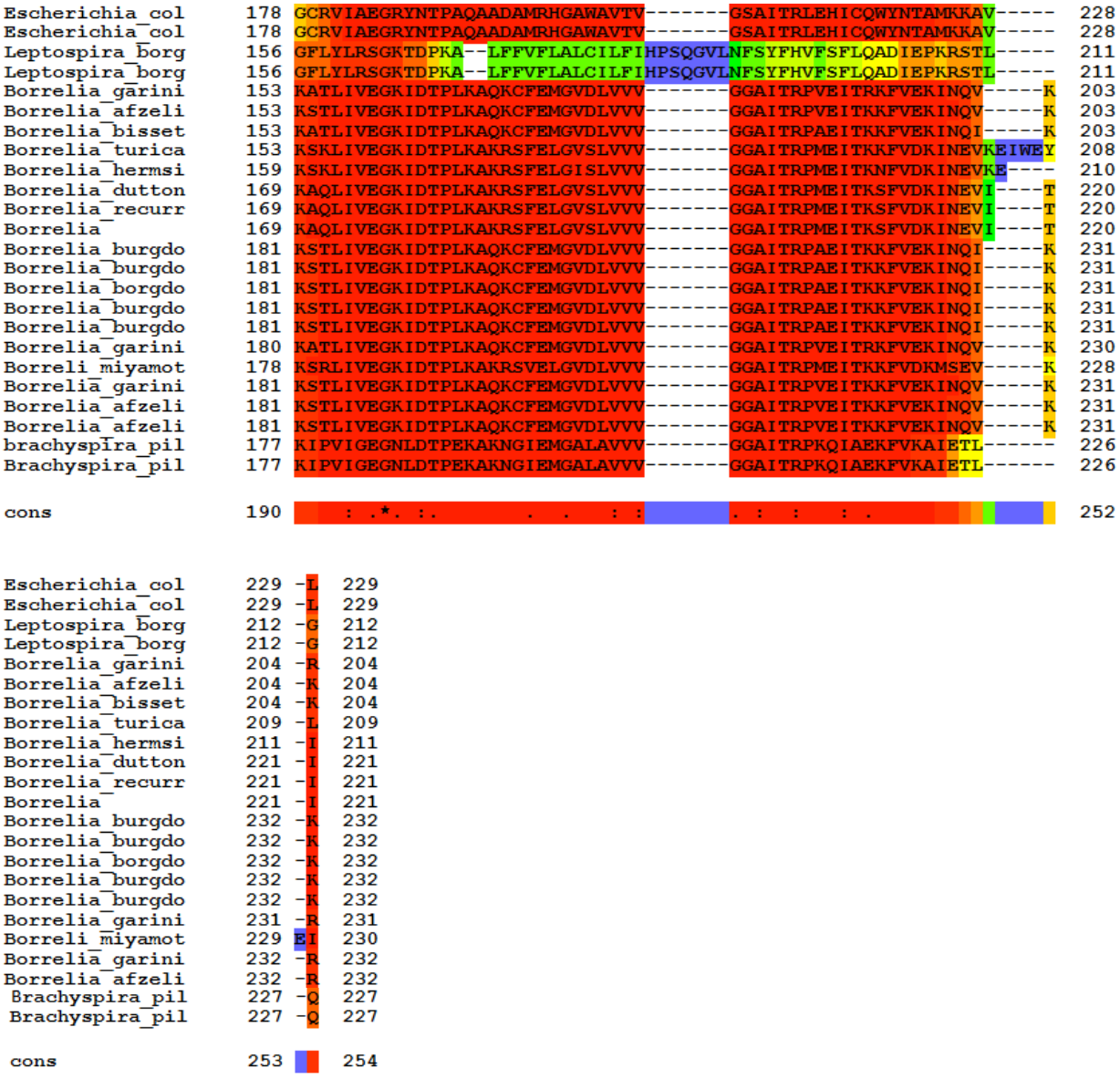
Copenhageni, L. interrogans IPAV las otras dos, son especies de *Leptospira borgpetersenii* (JB197 y L550). También se alinearon proteínas homólogas de dos especies del género *Treponema* (*T. brennaborense* y *T. pedis*), dos especies del género *Spirochaeta (S. sp. Buddy* y *S. caldaria*), de las especies *Brachyspira murdochii, Sphaerochaeta pleomorpha* Grapes y *Turneriella parva.*

ANEXO 2 (nanA) - score : 93

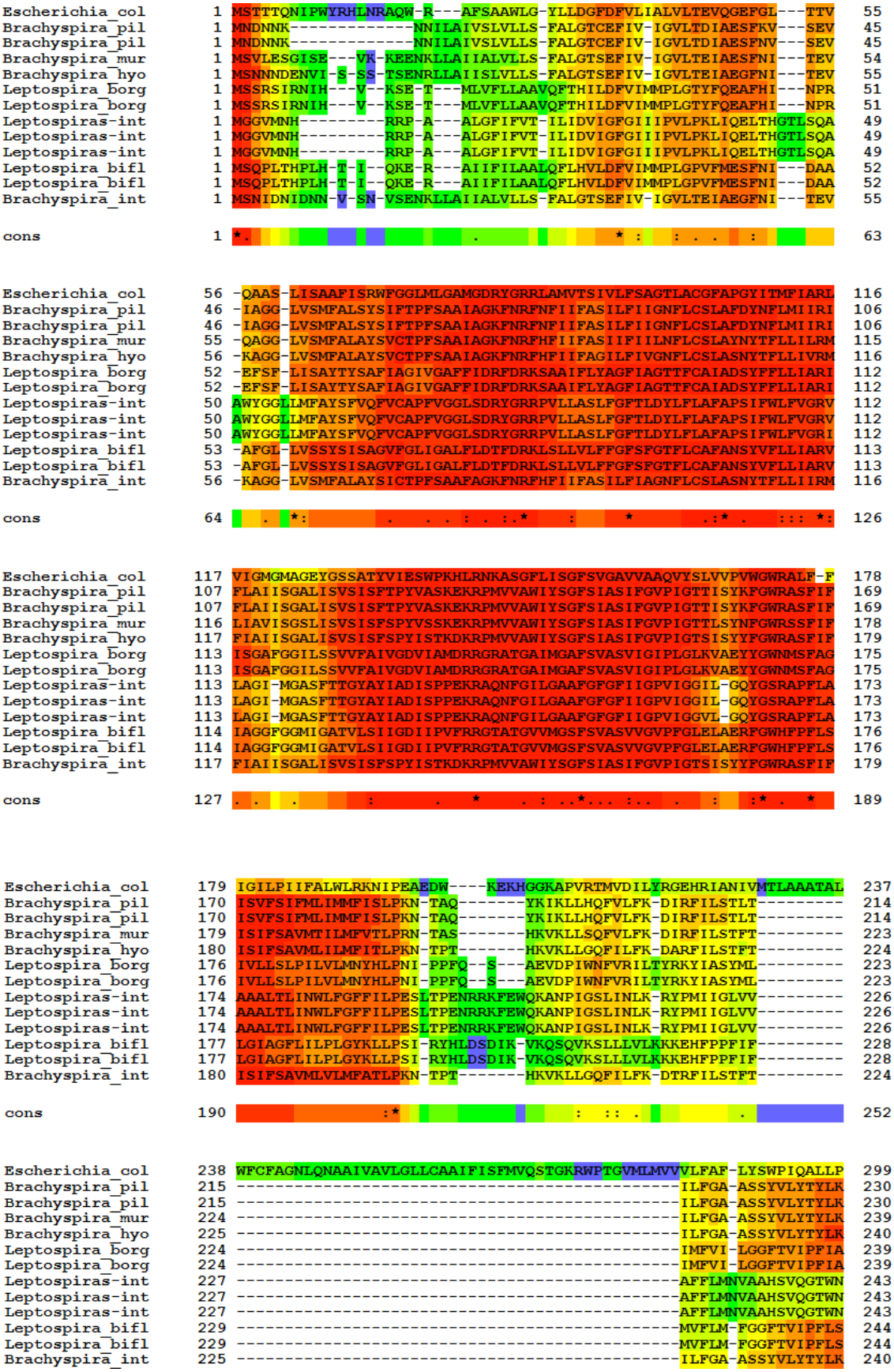


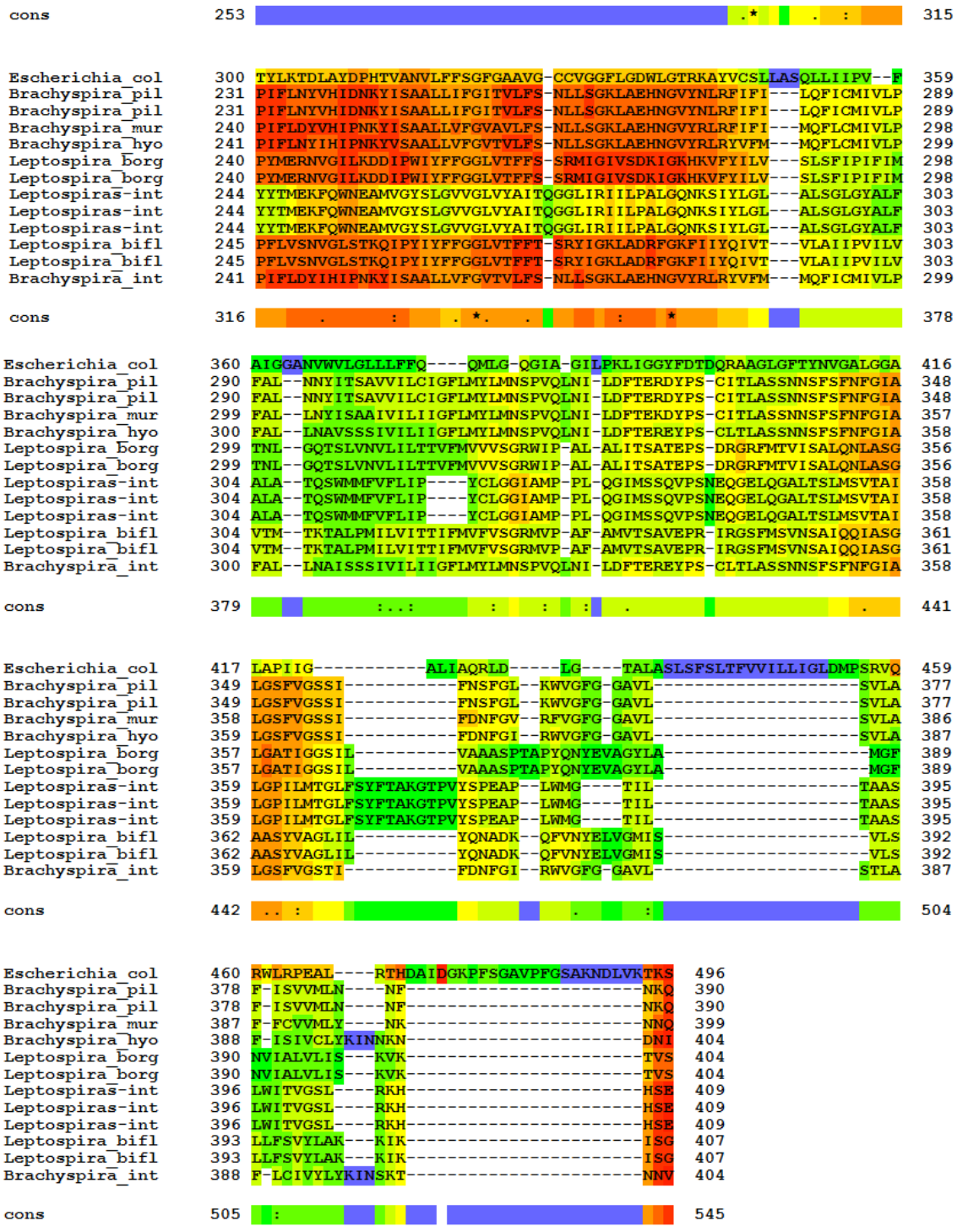
ANEXO 3 (nanE) - score : 90



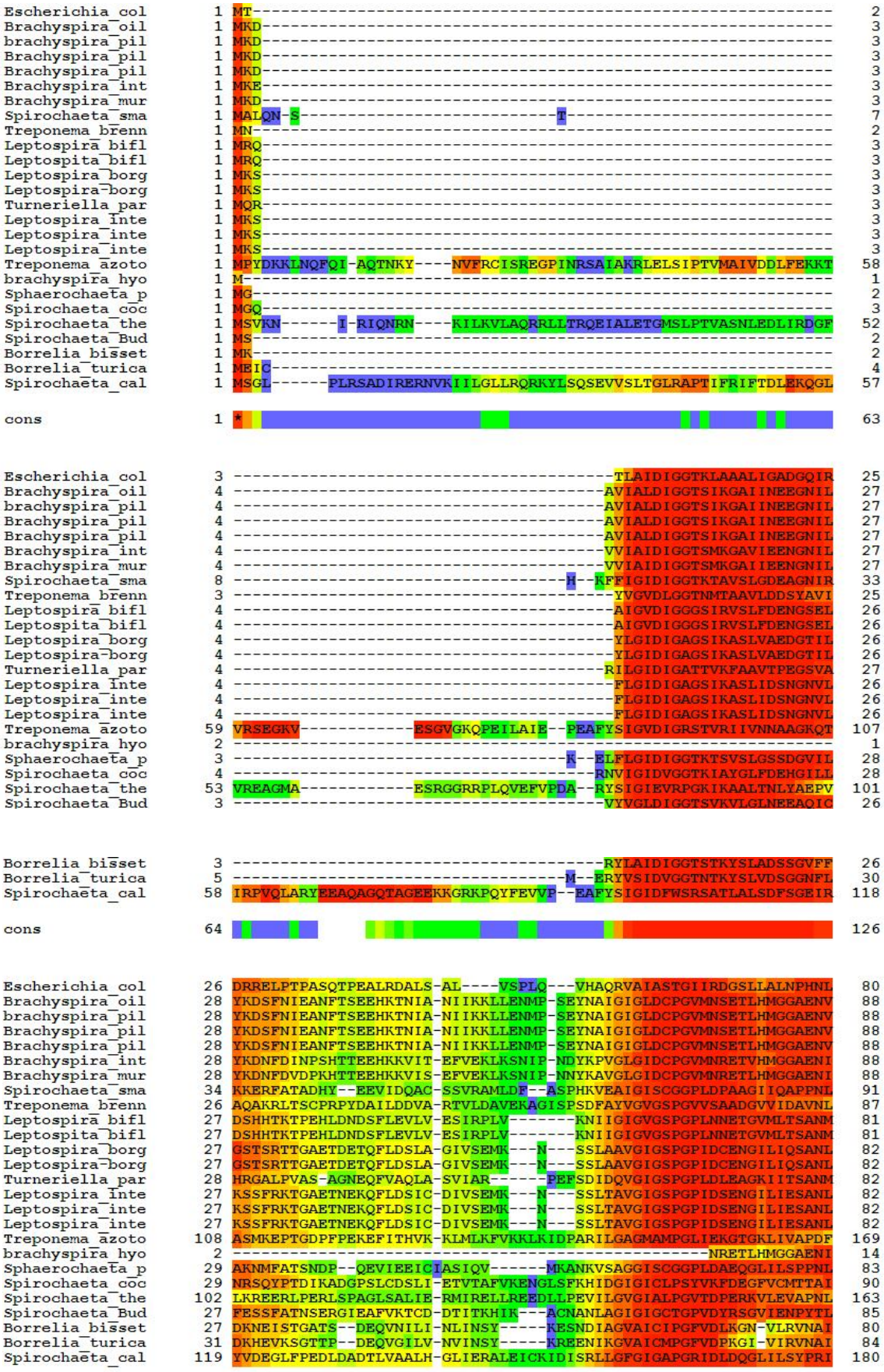

Escherichia_col 178 GCRVIAEGRYNTPAQAADAMRHGAWAVTV-------GSAITRLEHICQWYNTAMKKAV----- 228
Escherichia_col 178 GCRVIAEGRYNTPAQAADAMRHGAWAVTV-------GSAITRLEHICQWYNTAMKKAV----- 228
Leptospira_borg 156 GFLYLRSGKTDPKA--LFFVFLALCILFIHPSQGVLNFSYFHVFSFLQADIEPKRSTL----- 211
Leptospira_borg 156 GFLYLRSGKTDPKA--LFFVFLALCILFIHPSQGVLNFSYFHVFSFLQADIEPKRSTL----- 211
Borrelia_garini 153 KATLIVEGKIDTPLKAQKCFEMGVDLVVV-------GGAITRPVEITRKFVEKINQV-----K 203
Borrelia_afzeli 153 KSTLIVEGKIDTPLKAQKCFEMGVDLVVV-------GGAITRPVEITKKFVEKINQV-----K 203
Borrelia_bisset 153 KATLIVEGKIDTPLKAQKCFEMGVDLVVV-------GGAITRPAEITKKFVEKINQI-----K 203
Borrelia_turica 153 KSKLIVEGKIDTPLKAKRSFELGVSLVVV-------GGAITRPMEITKKFVDKINEVKEIWEY 208
Borrelia_hermsi 159 KSKLIVEGKIDTPLKAKRSFELGISLVVV-------GGAITRPMEITKNFVDKINEVKE---- 210
Borrelia_dutton 169 KAQLIVEGKIDTPLKAKRSFELGVSLVVV-------GGAITRPMEITKSFVDKINEVI----T 220
Borrelia_recurr 169 KAQLIVEGKIDTPLKAKRSFELGVSLVVV-------GGAITRPMEITKSFVDKINEVI----T 220
Borrelia 169 KAQLIVEGKIDTPLKAKRSFELGVSLVVV-------GGAITRPMEITKSFVDKINEVI----T 220
Borrelia_burgdo 181 KSTLIVEGKIDTPLKAQKCFEMGVDLVVV-------GGAITRPAEITKKFVEKINQI-----K 231
Borrelia_burgdo 181 KSTLIVEGKIDTPLKAQKCFEMGVDLVVV-------GGAITRPAEITKKFVEKINQI-----K 231
Borrelia_borgdo 181 KSTLIVEGKIDTPLKAQKCFEMGVDLVVV-------GGAITRPAEITKKFVEKINQI-----K 231
Borrelia_burgdo 181 KSTLIVEGKIDTPLKAQKCFEMGVDLVVV-------GGAITRPAEITKKFVEKINQI-----K 231
Borrelia_burgdo 181 KSTLIVEGKIDTPLKAQKCFEMGVDLVVV-------GGAITRPAEITKKFVEKINQI-----K 231
Borrelia_garini 180 KATLIVEGKIDTPLKAQKCFEMGVDLVVV-------GGAITRPVEITRKFVEKINQV-----K 230
Borreli_miyamot 178 KSRLIVEGKIDTPLKAKRSVELGVDLVVV-------GGAITRPMEITKKFVDKMSEV-----K 228
Borrelia_garini 181 KSTLIVEGKIDTPLKAQKCFEMGVDLVVV-------GGAITRPVEITKKFVEKINQV-----K 231
Borrelia_afzeli 181 KSTLIVEGKIDTPLKAQKCFEMGVDLVVV-------GGAITRPVEITKKFVEKINQV-----K 231
Borrelia_afzeli 181 KSTLIVEGKIDTPLKAQKCFEMGVDLVVV-------GGAITRPVEITKKFVEKINQV-----K 231
brachyspira_pil 177 KIPVIGEGNLDTPEKAKNGIEMGALAVVV-------GGAITRPKQIAEKFVKAIETL------ 226
Brachyspira_pil 177 KIPVIGEGNLDTPEKAKNGIEMGALAVVV-------GGAITRPKQIAEKFVKAIETL------ 226
cons 190 252
Escherichia_col 229 -L 229
Escherichia_col 229 -L 229
Leptospira_borg 212 -G 212
Leptospira_borg 212 -G 212
Borrelia_garini 204 -R 204
Borrelia_afzeli 204 -K 204
Borrelia_bisset 204 -K 204
Borrelia_turica 209 -L 209
Borrelia_hermsi 211 -I 211
Borrelia_dutton 221 -I 221
Borrelia_recurr 221 -I 221
Borrelia 221 -I 221
Borrelia_burgdo 232 -K 232
Borrelia_burgdo 232 -K 232
Borrelia_borgdo 232 -K 232
Borrelia_burgdo 232 -K 232
Borrelia_burgdo 232 -K 232
Borrelia_garini 231 -R 231
Borreli_miyamot 229 EI 230
Borrelia_garini 232 -R 232
Borrelia_afzeli 232 -R 232
Brachyspira_pil 227 -Q 227
Brachyspira_pil 227 -Q 227
cons 253 254

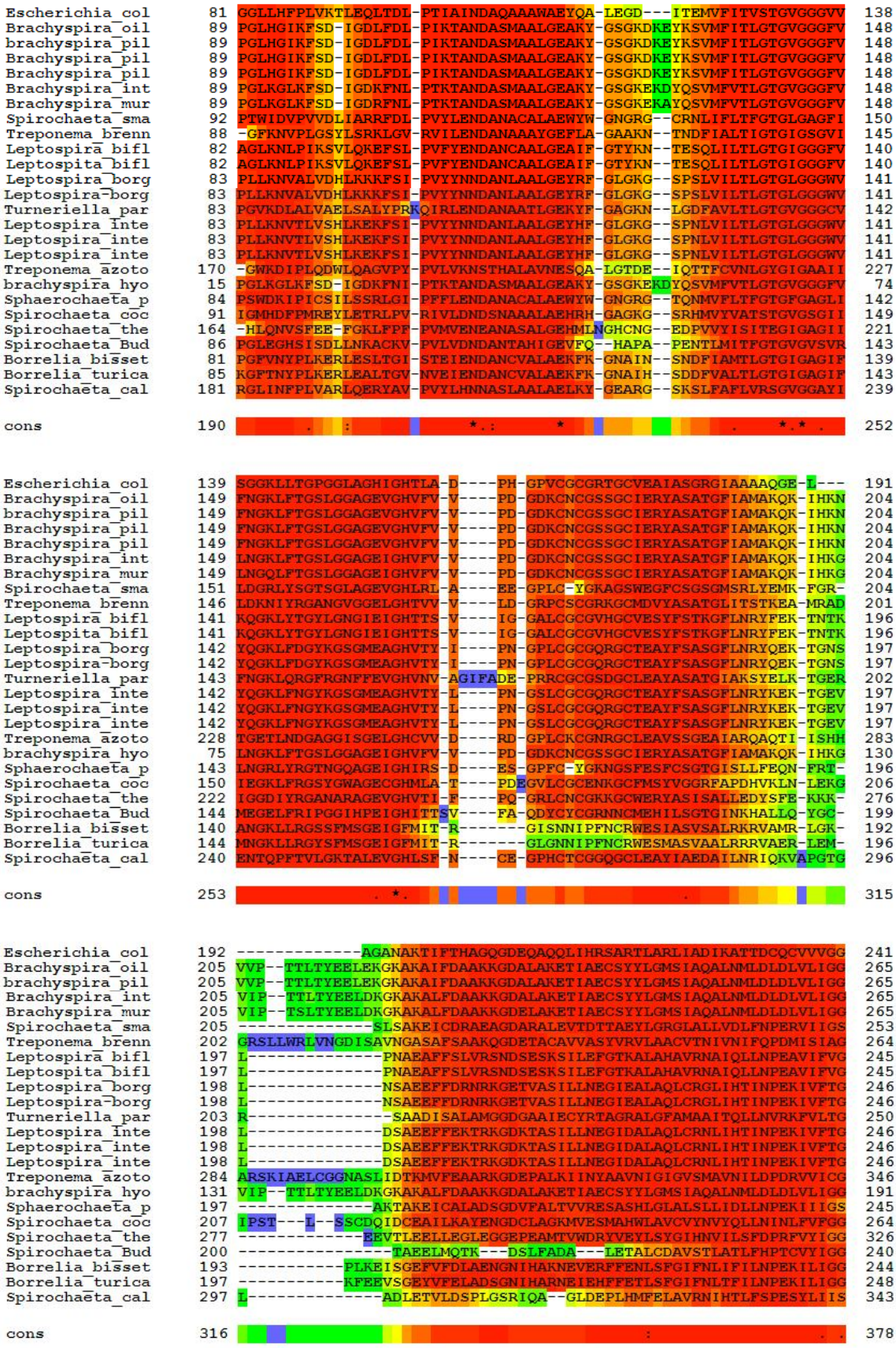
ANEXO 4 (nanT) - score : 69

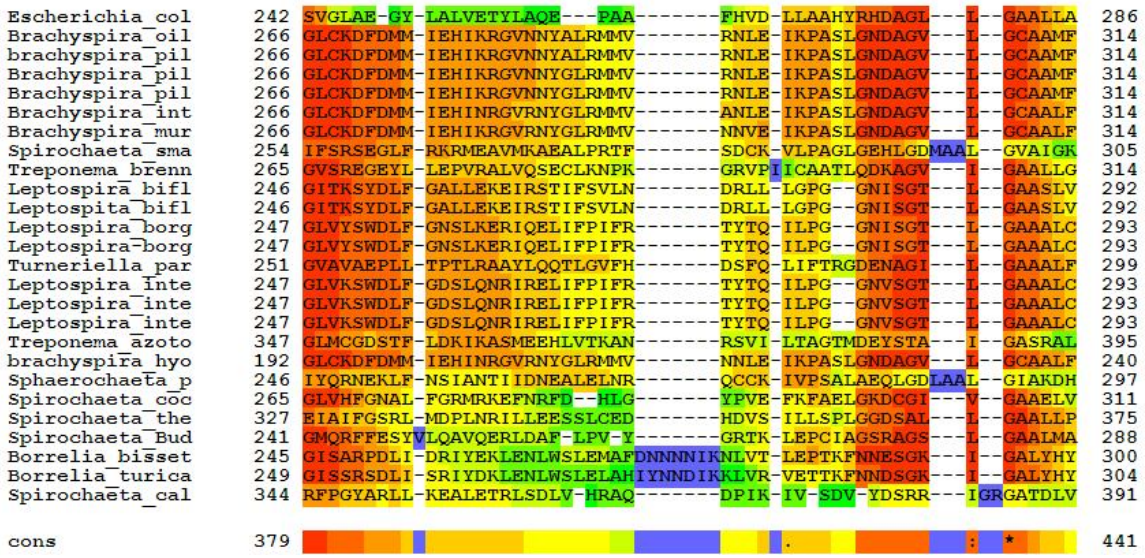



cons 253 315
Escherichia_col 300 TYLKTDLAYDPHTVANVLFFSGFGAAVG-CCVGGFLGDWLGTRKAYVCSLLASQLLIIPV--F 359
Brachyspira_pil 231 PIFLNYVHIDNKYISAALLIFGITVLFS-NLLSGKLAEHNGVYNLRFIFI---LQFICMIVLP 289
Brachyspira_pil 231 PIFLNYVHIDNKYISAALLIFGITVLFS-NLLSGKLAEHNGVYNLRFIFI---LQFICMIVLP 289
Brachyspira_mur 240 PIFLDYVHIPNKYISAALLVFGVAVLFS-NLLSGKLAEHNGVYRLRFIFI---MQFLCMIVLP 298
Brachyspira_hyo 241 PIFLNYIHIPNKYVSAALLVFGVTVLFS-NLLSGKLAEHNGVYRLRYVFM---MQFLCMIVLP 299
Leptospira_borg 240 PYMERNVGILKDDIPWIYFFGGLVTFFS-SRMIGIVSDKIGKHKVFYILV---SLSFIPIFIM 298
Leptospira_borg 240 PYMERNVGILKDDIPWIYFFGGLVTFFS-SRMIGIVSDKIGKHKVFYILV---SLSFIPIFIM 298
Leptospiras-int 244 YYTMEKFQWNEAMVGYSLGVVGLVYAITQGGLIRIILPALGQNKSIYLGL---ALSGLGYALF 303
Leptospiras-int 244 YYTMEKFQWNEAMVGYSLGVVGLVYAITQGGLIRIILPALGQNKSIYLGL---ALSGLGYALF 303
Leptospiras-int 244 YYTMEKFQWNEAMVGYSLGVVGLVYAITQGGLIRIILPALGQNKSIYLGL---ALSGLGYALF 303
Leptospira_bifl 245 PFLVSNVGLSTKQIPYIYFFGGLVTFFT-SRYIGKLADRFGKFIIYQIVT---VLAIIPVILV 303
Leptospira_bifl 245 PFLVSNVGLSTKQIPYIYFFGGLVTFFT-SRYIGKLADRFGKFIIYQIVT---VLAIIPVILV 303
Brachyspira_int 241 PIFLDYIHIPNKYISAALLVFGVTVLFS-NLLSGKLAEHNGVYRLRYVFM---MQFICMIVLP 299
cons 316 378
Escherichia_col 360 AIGGANVWVLGLLLFFQ----QMLG-QGIA-GILPKLIGGYFDTDQRAAGLGFTYNVGALGGA 416
Brachyspira_pil 290 FAL--NNYITSAVVILCIGFLMYLMNSPVQLNI-LDFTERDYPS-CITLASSNNSFSFNFGIA 348
Brachyspira_pil 290 FAL--NNYITSAVVILCIGFLMYLMNSPVQLNI-LDFTERDYPS-CITLASSNNSFSFNFGIA 348
Brachyspira_mur 299 FAL--LNYISAAIVILIIGFLMYLMNSPVQLNI-LDFTEKDYPS-CITLASSNNSFSFNFGIA 357
Brachyspira_hyo 300 FAL--LNAVSSSIVILIIGFLMYLMNSPVQLNI-LDFTEREYPS-CLTLASSNNSFSFNFGIA 358
Leptospira_borg 299 TNL--GQTSLVNVLILTTVFMVVVSGRWIP-AL-ALITSATEPS-DRGRFMTVISALQNLASG 356
Leptospira_borg 299 TNL--GQTSLVNVLILTTVFMVVVSGRWIP-AL-ALITSATEPS-DRGRFMTVISALQNLASG 356
Leptospiras-int 304 ALA--TQSWMMFVFLIP----YCLGGIAMP-PL-QGIMSSQVPSNEQGELQGALTSLMSVTAI 358
Leptospiras-int 304 ALA--TQSWMMFVFLIP----YCLGGIAMP-PL-QGIMSSQVPSNEQGELQGALTSLMSVTAI 358
Leptospiras-int 304 ALA--TQSWMMFVFLIP----YCLGGIAMP-PL-QGIMSSQVPSNEQGELQGALTSLMSVTAI 358
Leptospira_bifl 304 VTM--TKTALPMILVITTIFMVFVSGRMVP-AF-AMVTSAVEPR-IRGSFMSVNSAIQQIASG 361
Leptospira_bifl 304 VTM--TKTALPMILVITTIFMVFVSGRMVP-AF-AMVTSAVEPR-IRGSFMSVNSAIQQIASG 361
Brachyspira_int 300 FAL--LNAISSSIVILIIGFLMYLMNSPVQLNI-LDFTEREYPS-CLTLASSNNSFSFNFGIA 358
cons 379 441
Escherichia_col 417 LAPIIG----------ALIAQRLD-----LG----TALASLSFSLTFVVILLIGLDMPSRVQ 459
Brachyspira_pil 349 LGSFVGSSI----------FNSFGL--KWVGFG-GAVL----------SVLA 377
Brachyspira_pil 349 LGSFVGSSI----------FNSFGL--KWVGFG-GAVL----------SVLA 377
Brachyspira_mur 358 LGSFVGSSI----------FDNFGV--RFVGFG-GAVL----------SVLA 386
Brachyspira_hyo 359 LGSFVGSSI----------FDNFGI--RWVGFG-GAVL----------SVLA 387
Leptospira_borg 357 LGATIGGSIL----------VAAASPTAPYQNYEVAGYLA----------MGF 389
Leptospira_borg 357 LGATIGGSIL----------VAAASPTAPYQNYEVAGYLA----------MGF 389
Leptospiras-int 359 LGPILMTGLFSYFTAKGTPVYSPEAP--LWMG----TIL----------TAAS 395
Leptospiras-int 359 LGPILMTGLFSYFTAKGTPVYSPEAP--LWMG----TIL----------TAAS 395
Leptospiras-int 359 LGPILMTGLFSYFTAKGTPVYSPEAP--LWMG----TIL----------TAAS 395
Leptospira_bifl 362 AASYVAGLIL----------YQNADK--QFVNYELVGMIS----------VLS 392
Brachyspira_int 362 AASYVAGLIL----------YQNADK--QFVNYELVGMIS----------VLS 392
Brachyspira_int 359 LGSFVGSTI----------FDNFGI--RWVGFG-GAVL----------STLA 387
cons 442 504
Escherichia_col 460 RWLRPEAL----RTHDAIDGKPFSGAVPFGSAKNDLVKTKS 496
Brachyspira_pil 378 F-ISVVMLN----NF----------NKQ 390
Brachyspira_pil 378 F-ISVVMLN----NF----------NKQ 390
Brachyspira_mur 387 F-FCVVMLY----NK----------NNQ 399
Brachyspira_hyo 388 F-ISIVCLYKINNKN----------DNI 404
Leptospira_borg 390 NVIALVLIS---KVK----------TVS 404
Leptospira_borg 390 NVIALVLIS---KVK----------TVS 404
Leptospiras-int 396 LWITVGSL----RKH----------HSE 409
Leptospiras-int 396 LWITVGSL----RKH----------HSE 409
Leptospiras-int 396 LWITVGSL----RKH----------HSE 409
Leptospira_bifl 393 LLFSVYLAK---KIK----------ISG 407
Leptospira_bifl 393 LLFSVYLAK---KIK----------ISG 407
Brachyspira_int 388 F-LCIVYLYKINSKT----------NNV 404
cons 505 545

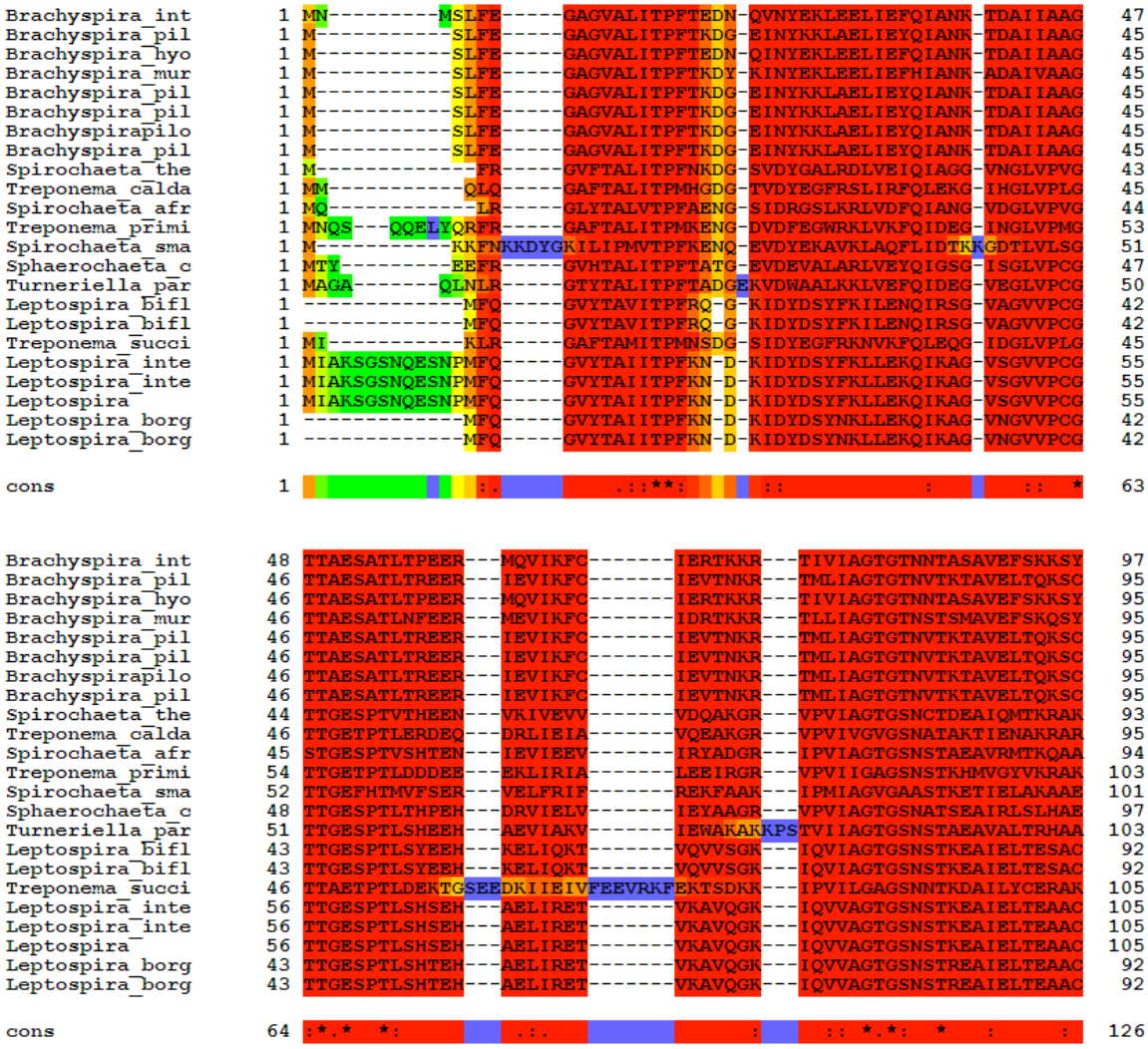
ANEXO 5 (nanK) - score : 80

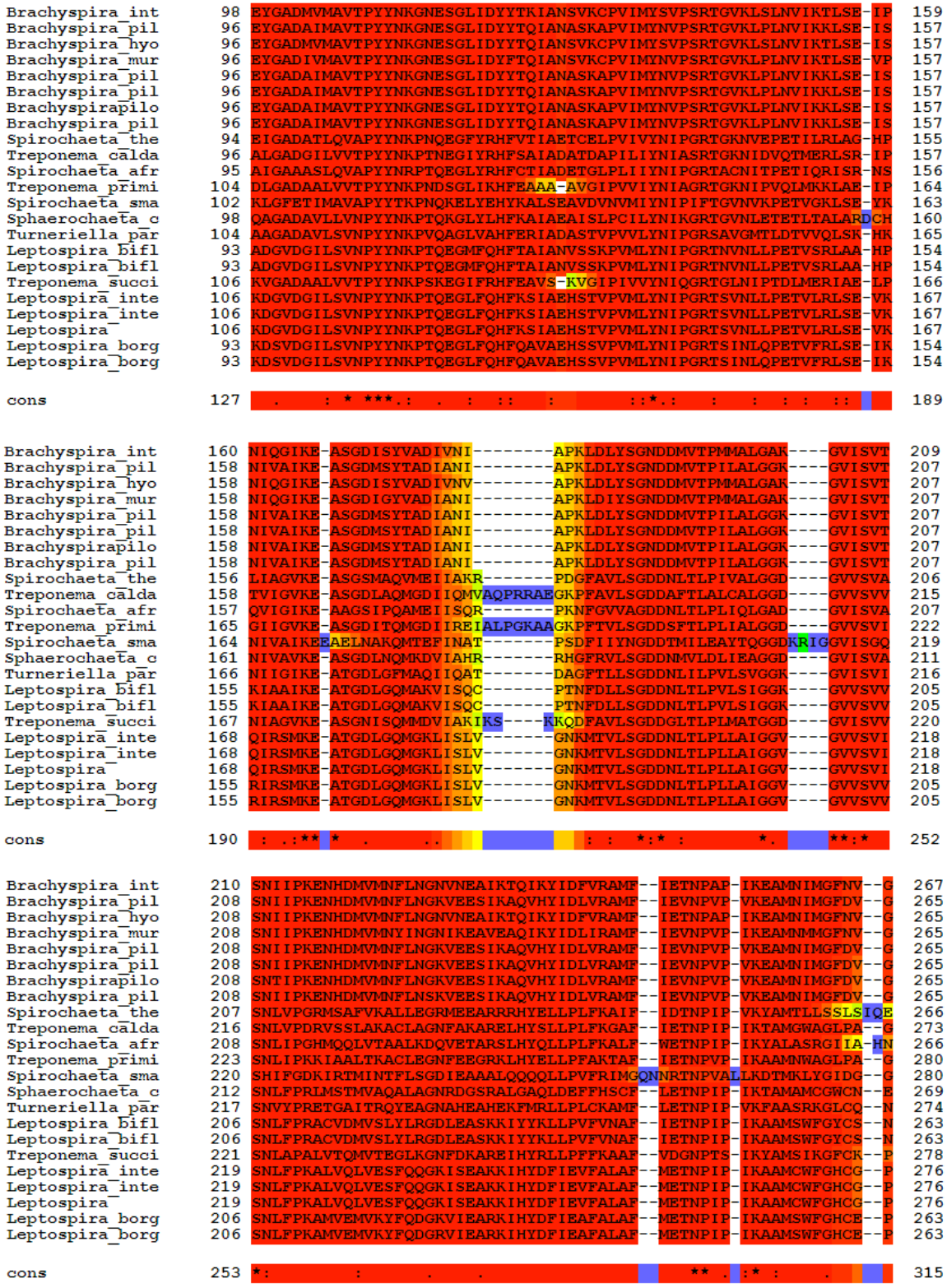


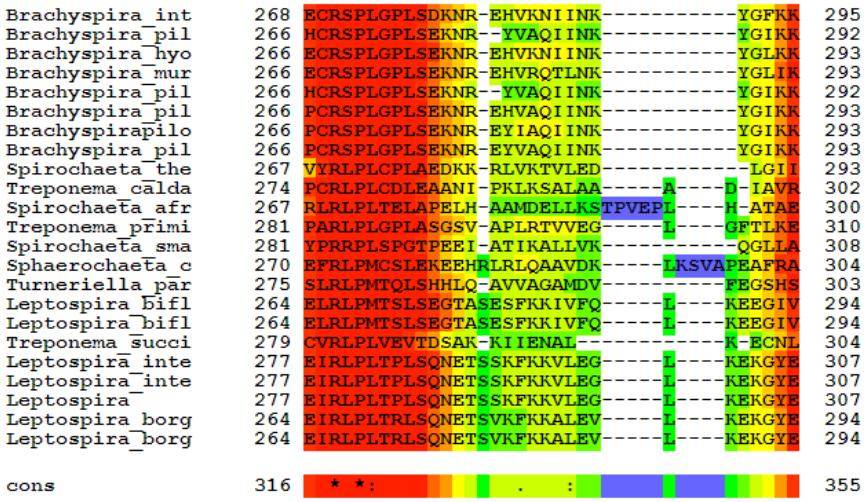

Escherichia_col 81 GGLLHFPLVKTLEQLTDL-PTIAINDAQAAAWAEYQA-LEGD---ITEMVFITVSTGVGGGVV 138
Brachyspira_oil 89 PGLHGIKFSD-IGDLFDL-PIKTANDASMAALGEAKY-GSGKDKEYKSVMFITLGTGVGGGFV 148
brachyspira_pil 89 PGLHGIKFSD-IGDLFDL-PIKTANDASMAALGEAKY-GSGKDKEYKSVMFITLGTGVGGGFV 148
Brachyspira_pil 89 PGLHGIKFSD-IGDLFDL-PIKTANDASMAALGEAKY-GSGKDKEYKSVMFITLGTGVGGGFV 148
Brachyspira_pil 89 PGLHGIKFSD-IGDLFDL-PIKTANDASMAALGEAKY-GSGKDKEYKSVMFITLGTGVGGGFV 148
Brachyspira_int 89 PGLKGLKFSD-IGDKFNL-PTKTANDASMAALGEAKY-GSGKEKDYQSVMFVTLGTGVGGGFV 148
Brachyspira_mur 89 PGLKGLKFSD-IGDRFNL-PTKTANDASMAALGEAKY-GSGKEKAYQSVMFVTLGTGVGGGFV 148
Spirochaeta_sma 92 PTWIDVPVVDLIARRFDL-PVYLENDANACALAEWYW-GNGRG---CRNLIFLTFGTGLGAGFI 150
Treponema_brenn 88 -GFKNVPLGSYLSRKLGV-RVILENDANAAAYGEFLA-GAAKN---TNDFIALTIGTGIGSGVI 145
Leptospira_bifl 82 AGLKNLPIKSVLQKEFSL-PVFYENDANCAALGEAIF-GTYKN---TESQLILTLGTGIGGGFV 140
Leptospita_bifl 82 AGLKNLPIKSVLQKEFSL-PVFYENDANCAALGEAIF-GTYKN---TESQLILTLGTGIGGGFV 140
Leptospira_borg 83 PLLKNVALVDHLKKKFSI-PVYYNNDANLAALGEYRF-GLGKG---SPSLVILTLGTGLGGGWV 141
Leptospira_borg 83 PLLKNVALVDHLKKKFSI-PVYYNNDANLAALGEYRF-GLGKG---SPSLVILTLGTGLGGGWV 141
Turneriella_par 83 PGVKDLALVAELSALYPRKQIRLENDANAATLGEKYF-GAGKN---LGDFAVLTLGTGVGGGCV 142
Leptospira_inte 83 PLLKNVTLVSHLKEKFSI-PVYYNNDANLAALGEYHF-GLGKG---SPNLVILTLGTGLGGGWV 141
Leptospira_inte 83 PLLKNVTLVSHLKEKFSI-PVYYNNDANLAALGEYHF-GLGKG---SPNLVILTLGTGLGGGWV 141
Leptospira_inte 83 PLLKNVTLVSHLKEKFSI-PVYYNNDANLAALGEYHF-GLGKG---SPNLVILTLGTGLGGGWV 141
Treponema_azoto 170 -GWKDIPLQDWLQAGVPY-PVLVKNSTHALAVNESQA-LGTDE---IQTTFCVNLGYGIGAAII 227
brachyspira_hyo 15 PGLKGLKFSD-IGDKFNI-PTKTANDASMAALGEAKY-GSGKEKDYQSVMFVTLGTGVGGGFV 74
Sphaerochaeta_p 84 PSWDKIPICSILSSRLGI-PFFLENDANACALAEWYW-GNGRG---TQNMVFLTFGTGFGAGLI 142
Spirochaeta_coc 91 IGMHDFPMREYLETRLPV-RIVLDNDSNAAALAEHRH-GAGKG---SRHMVYVATSTGVGSGII 149
Spirochaeta_the 164 -HLQNVSFEE-FGKLFPF-PVMVENEANASALGEHMLNGHCNG---EDPVVYISITEGIGAGII 221
Spirochaeta_Bud 86 PGLEGHSISDLLNKACKV-PVLVDNDANTAHIGEVFQ---HAPA---PENTLMITFGTGVGVSVR 143
Borrelia_bisset 81 PGFVNYPLKERLESLTGI-STEIENDANCVALAEKFK-GNAIN---SNDFIAMTLGTGIGAGIF 139
Borrelia_turica 85 KGFTNYPLKERLEALTGV-NVEIENDANCVALAEKFK-GNAIH---SDDFVALTLGTGIGAGIF 143
Spirochaeta_cal 181 RGLINFPLVARLQERYAV-PVYLHNNASLAALAELKY-GEARG---SKSLFAFLVRSGVGGAYI 239
cons 190 252
Escherichia_col 139 SGGKLLTGPGGLAGHIGHTLA-D----PH-GPVCGCGRTGCVEAIASGRGIAAAAQGE-L--- 191
Brachyspira_oil 149 FNGKLFTGSLGGAGEVGHVFV-V----PD-GDKCNCGSSGCIERYASATGFIAMAKQK-IHKN 204
brachyspira_pil 149 FNGKLFTGSLGGAGEVGHVFV-V----PD-GDKCNCGSSGCIERYASATGFIAMAKQK-IHKN 204
Brachyspira_pil 149 FNGKLFTGSLGGAGEVGHVFV-V----PD-GDKCNCGSSGCIERYASATGFIAMAKQK-IHKN 204
Brachyspira_pil 149 FNGKLFTGSLGGAGEVGHVFV-V----PD-GDKCNCGSSGCIERYASATGFIAMAKQK-IHKN 204
Brachyspira_int 149 LNGKLFTGSLGGAGEIGHVFV-V----PD-GDKCNCGSSGCIERYASATGFIAMAKQK-IHKG 204
Brachyspira_mur 149 LNGQLFTGSLGGAGEIGHVFV-V----PD-GDKCNCGSSGCIERYASATGFIAMAKQK-IHKG 204
Spirochaeta_sma 151 LDGRLYSGTSGLAGEVGHLRI-A----EE-GPLC-YGKAGSWEGFCSGSGMSRLYEMK-FGR- 204
Treponema_brenn 146 LDKNIYRGANGVGGELGHTVV-V----LD-GRPCSCGRKGCMDVYASATGLITSTKEA-MRAD 201
Leptospira_bifl 141 KQGKLYTGYLGNGIEIGHTTS-V----IG-GALCGCGVHGCVESYFSTKGFLNRYFEK-TNTK 196
Leptospita_bifl 141 KQGKLYTGYLGNGIEIGHTTS-V----IG-GALCGCGVHGCVESYFSTKGFLNRYFEK-TNTK 196
Leptospira_borg 142 YQGKLFDGYKGSGMEAGHVTY-I----PN-GPLCGCGQRGCTEAYFSASGFLNRYQEK-TGNS 197
Leptospira_borg 142 YQGKLFDGYKGSGMEAGHVTY-I----PN-GPLCGCGQRGCTEAYFSASGFLNRYQEK-TGNS 197
Turneriella_par 143 FNGKLQRGFRGNFFEVGHVNV-AGIFADE-PRRCGCGSDGCLEAYASATGIAKSYELK-TGER 202
Leptospira_inte 142 YQGKLFNGYKGSGMEAGHVTY-L----PN-GSLCGCGQRGCTEAYFSASGFLNRYKEK-TGEV 197
Leptospira_inte 142 YQGKLFNGYKGSGMEAGHVTY-L----PN-GSLCGCGQRGCTEAYFSASGFLNRYKEK-TGEV 197
Leptospira_inte 142 YQGKLFNGYKGSGMEAGHVTY-L----PN-GSLCGCGQRGCTEAYFSASGFLNRYKEK-TGEV 197
Treponema_azoto 228 TGETLNDGAGGISGELGHCVV-D----RD-GPLCKCGNRGCLEAVSSGEAIARQAQTI-ISHH 283
brachyspira_hyo 75 LNGKLFTGSLGGAGEIGHVFV-V----PD-GDKCNCGSSGCIERYASATGFIAMAKQK-IHKG 130
Sphaerochaeta_p 143 LNGRLYRGTNGQAGEIGHIRS-D----ES-GPFC-YGKNGSFESFCSGTGISLLFEQN-FRT- 196
Spirochaeta_coc 150 IEGKLFRGSYGWAGECGHMLA-T----PDEGVLCGCENKGCFMSYVGGRFAPDHVKLN-LEKG 206
Spirochaeta_the 222 IGGDIYRGANARAGEVGHVTI-F----PQ-GRLCNCGKKGCWERYASISALLEDYSFE-KKK- 276
Spirochaeta_Bud 144 MEGELFRIPGGIHPEIGHITTSV----FA-QDYCYCGRNNCMEHILSGTGINKHALLQ-YGC- 199
Borrelia_bisset 140 ANGKLLRGSSFMSGEIGFMIT-R-------GISNNIPFNCRWESIASVSALRKRVAMR-LGK- 192
Borrelia_turica 144 MNGKLLRGYSFMSGEIGFMIT-R-------GLGNNIPFNCRWESMASVAALRRRVAER-LEM- 196
Spirochaeta_cal 240 ENTQPFTVLGKTALEVGHLSF-N----CE-GPHCTCGGQGCLEAYIAEDAILNRIQKVAPGTG 296
cons 253 315
Escherichia_col 192 ----------AGANAKTIFTHAGQGDEQAQQLIHRSARTLARLIADIKATTDCQCVVVGG 241
Brachyspira_oil 205 VVP--TTLTYEELEKGKAKAIFDAAKKGDALAKETIAECSYYLGMSIAQALNMLDLDLVLIGG 265
brachyspira_pil 205 VVP--TTLTYEELEKGKAKAIFDAAKKGDALAKETIAECSYYLGMSIAQALNMLDLDLVLIGG 265
Brachyspira_int 205 VIP--TTLTYEELDKGKAKALFDAAKKGDALAKETIAECSYYLGMSIAQALNMLDLDLVLIGG 265
Brachyspira_mur 205 VIP--TSLTYEELDKGKAKALFDAAKKGDELAKETIAECSYYLGMSIAQALNMLDLDLVLIGG 265
Spirochaeta_sma 205 ----------SLSAKEICDRAEAGDARALEVTDTTAEYLGRGLALLVDLFNPERVIIGS 253
Treponema_brenn 202 GRSLLWRLVNGDISAVNGASAFSAAKQGDETACAVVASYVRVLAACVTNIVNIFQPDMISIAG 264
Leptospira_bifl 197 L----------PNAEAFFSLVRSNDSESKSILEFGTKALAHAVRNAIQLLNPEAVIFVG 245
Leptospita_bifl 197 L----------PNAEAFFSLVRSNDSESKSILEFGTKALAHAVRNAIQLLNPEAVIFVG 245
Leptospira_borg 198 L----------NSAEEFFDRNRKGETVASILLNEGIEALAQLCRGLIHTINPEKIVFTG 246
Leptospira_borg 198 L----------NSAEEFFDRNRKGETVASILLNEGIEALAQLCRGLIHTINPEKIVFTG 246
Turneriella_par 203 R----------SAADISALAMGGDGAAIECYRTAGRALGFAMAAITQLLNVRKFVLTG 250
Leptospira_inte 198 L----------DSAEEFFEKTRKGDKTASILLNEGIDALAQLCRNLIHTINPEKIVFTG 246
Leptospira_inte 198 L----------DSAEEFFEKTRKGDKTASILLNEGIDALAQLCRNLIHTINPEKIVFTG 246
Leptospira_inte 198 L----------DSAEEFFEKTRKGDKTASILLNEGIDALAQLCRNLIHTINPEKIVFTG 246
Treponema_azoto 284 ARSKIAELCGGNASLIDTKMVFEAARKGDEPALKIINYAAVNIGIGVSMAVNILDPDRVVICG 346
brachyspira_hyo 131 VIP--TTLTYEELDKGKAKALFDAAKKGDALAKETIAECSYYLGMSIAQALNMLDLDLVLIGG 191
Sphaerochaeta_p 197 ----------AKTAKEICALADSGDVFALTVVRESASHLGLALSLLIDLLNPEKIIIGS 245
Spirochaeta_coc 207 IPST---L--SSCDQIDCEAILKAYENGDCLAGKMVESMAHWLAVCVYNVYQLLNINLFVFGG 264
Spirochaeta_the 277 ----------EEVTLEELLEGLEGGEPEAMTVWDRYVEYLSYGIHNVILSFDPRFVYIGG 326
Spirochaeta_Bud 200 ----------TAEELMQTK---DSLFADA---LETALCDAVSTLATLFHPTCVYIGG 240
Borrelia_bisset 193 ----------PLKEISGEFVFDLAENGNIHAKNEVERFFENLSFGIFNLIFILNPEKILIGG 244
Borrelia_turica 197 ----------KFEEVSGEYVFELADSGNIHARNEIEHFFETLSFGIFNLTFILNPEKILIGG 248
Spirochaeta_cal 297 L----------ADLETVLDSPLGSRIQA--GLDEPLHMFELAVRNIHTLFSPESYLIIS 343
cons 316 378



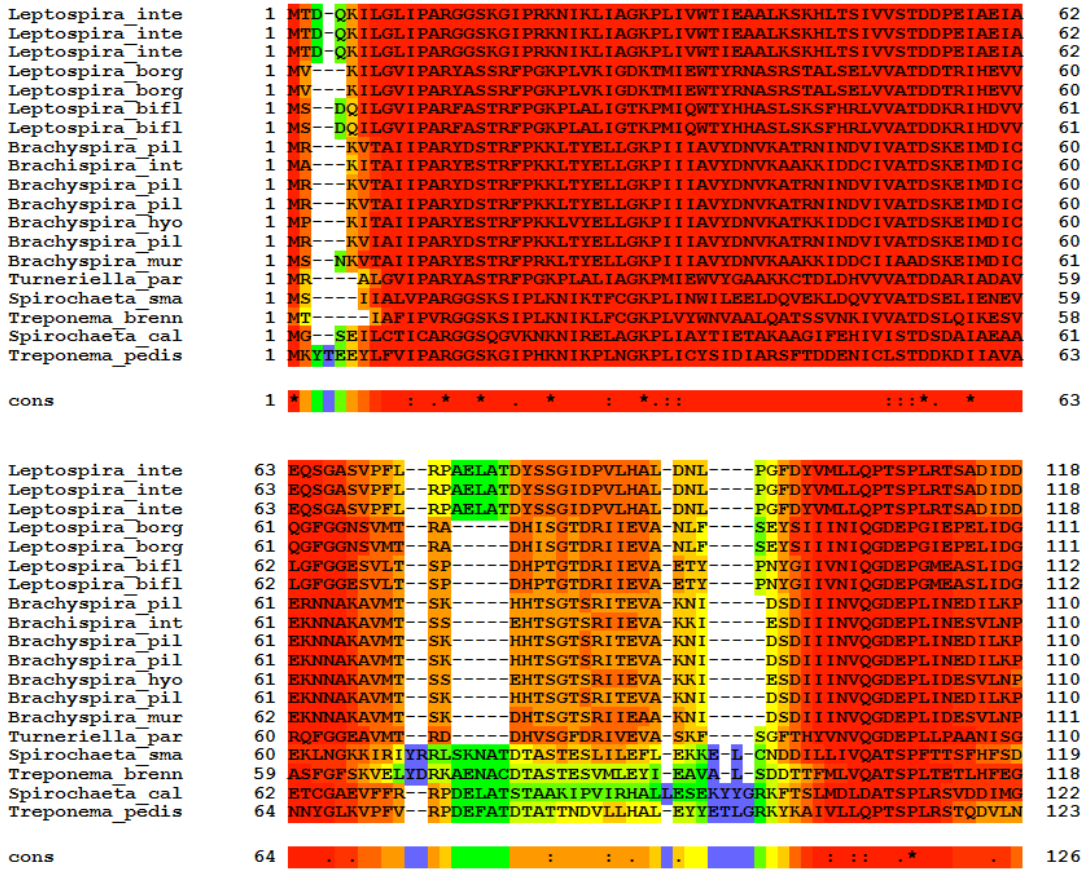
ANEXO 6 (dapA) - score: 95

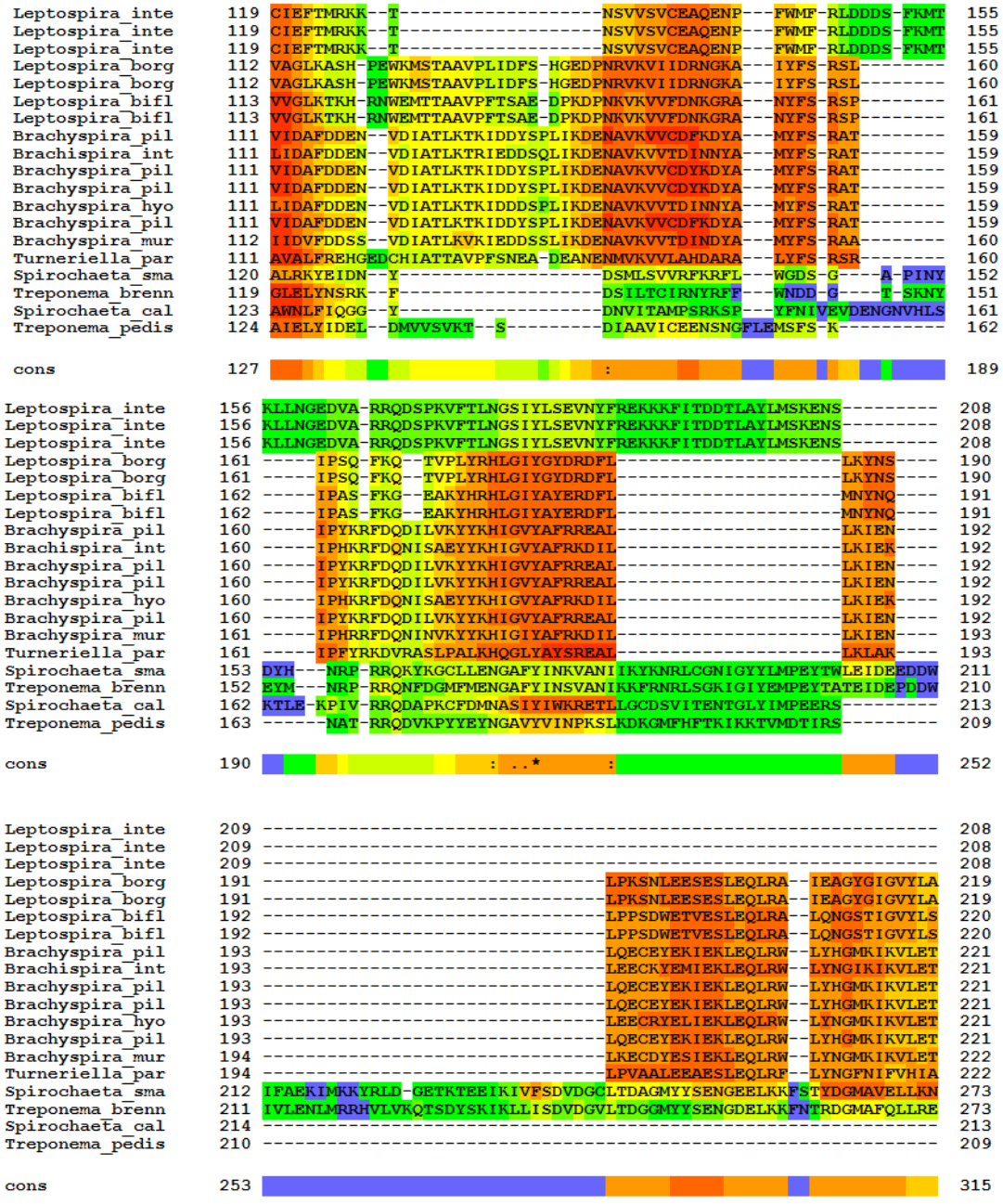


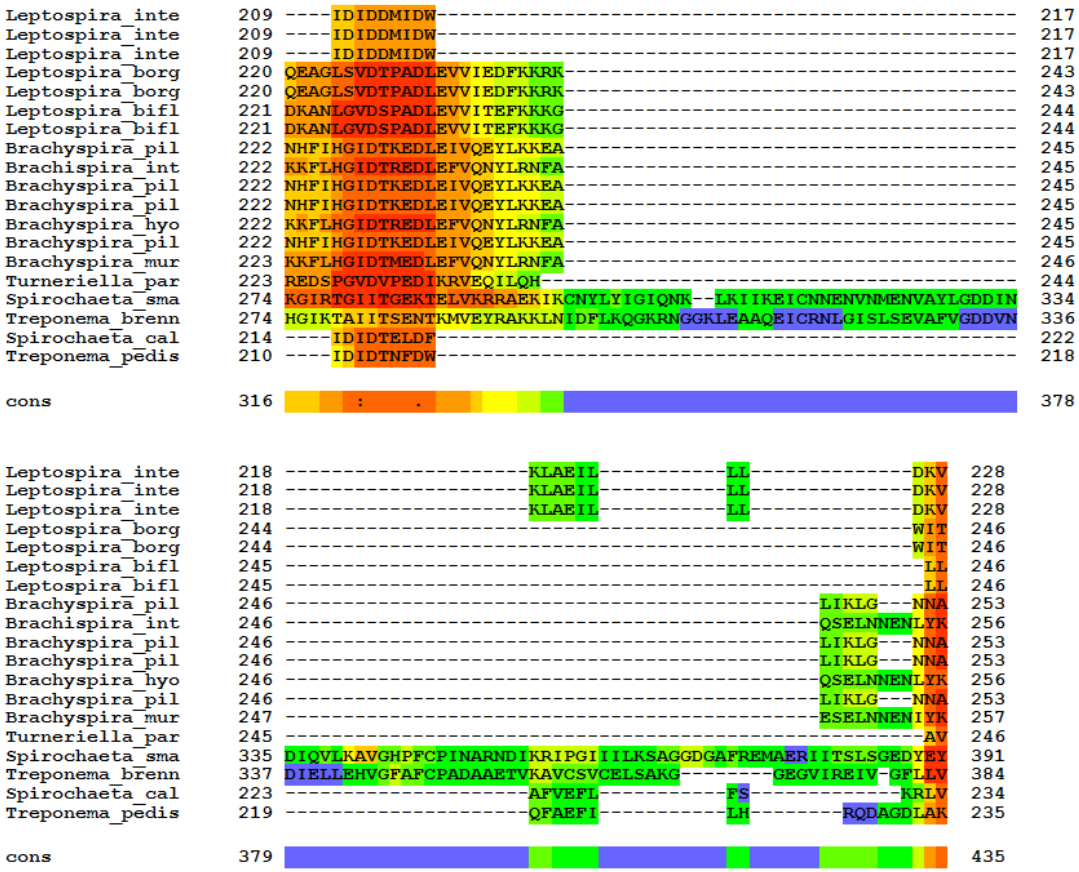

Brachyspira_int 98 EYGADMVMAVTPYYNKGNESGLIDYYTKIANSVKCPVIMYSVPSRTGVKLSLNVIKTLSE-IP 159
Brachyspira_pil 96 EYGADAIMAVTPYYNKGNESGLIDYYTQIANASKAPVIMYNVPSRTGVKLPLNVIKKLSE-IS 157
Brachyspira_hyo 96 EYGADMVMAVTPYYNKGNESGLIDYYTQIANSVKCPVIMYSVPSRTGVKLSLNVIKTLSE-IS 157
Brachyspira_mur 96 EYGADIVMAVTPYYNKGNESGLIDYFTQIANSVKCPVIMYNVPSRTGVKLPLNVIKTLSE-VP 157
Brachyspira_pil 96 EYGADAIMAVTPYYNKGNESGLIDYYTQIANASKAPVIMYNVPSRTGVKLPLNVIKKLSE-IS 157
Brachyspira_pil 96 EYGADAIMAVTPYYNKGNESGLIDYYTQIANASKAPVIMYNVPSRTGVKLPLNVIKKLSE-IS 157
Brachyspirapilo 96 EYGADAIMAVTPYYNKGNESGLIDYYTQIANASKAPVIMYNVPSRTGVKLPLNVIKKLSE-IS 157
Brachyspira_pil 96 EYGADAIMAVTPYYNKGNESGLIDYYTQIANASKAPVIMYNVPSRTGVKLPLNVIKKLSE-IS 157
Spirochaeta_the 94 EIGADATLQVAPYYNKPNQEGFYRHFVTIAETCELPVIVYNIPGRTGKNVEPETILRLAG-HP 155
Treponema_calda 96 ALGADGILVVTPYYNKPTNEGIYRHFSAIADATDAPILIYNIASRTGKNIDVQTMERLSR-IP 157
Spirochaeta_afr 95 AIGAAASLQVAPYYNRPTQEGLYRHFCTIADETGLPLIIYNIPGRTACNITPETIQRISR-NS 156
Treponema_primi 104 DLGADAALVVTPYYNKPNDSGLIKHFEAAA-AVGIPVVIYNIAGRTGKNIPVQLMKKLAE-IP 164
Spirochaeta_sma 102 KLGFETIMAVAPYYTKPNQKELYEHYKALSEAVDVNVMIYNIPIFTGVNVKPETVGKLSE-YK 163
Sphaerochaeta_c 98 QAGADAVLLVNPYYNKPTQKGLYLHFKAIAEAISLPCILYNIKGRTGVNLETETLTALARDCH 160
Turneriella_par 104 AAGADAVLSVNPYYNKPVQAGLVAHFERIADASTVPVVLYNIPGRSAVGMTLDTVVQLSK-HK 165
Leptospira_bifl 93 ADGVDGILSVNPYYNKPTQEGMFQHFTAIANVSSKPVMLYNIPGRTNVNLLPETVSRLAA-HP 154
Leptospira_bifl 93 ADGVDGILSVNPYYNKPTQEGMFQHFTAIANVSSKPVMLYNIPGRTNVNLLPETVSRLAA-HP 154
Treponema_succi 106 KVGADAALVVTPYYNKPSKEGIFRHFEAVS-KVGIPIVVYNIQGRTGLNIPTDLMERIAE-LP 166
Leptospira_inte 106 KDGVDGILSVNPYYNKPTQEGLFQHFKSIAEHSTVPVMLYNIPGRTSVNLLPETVLRLSE-VK 167
Leptospira_inte 106 KDGVDGILSVNPYYNKPTQEGLFQHFKSIAEHSTVPVMLYNIPGRTSVNLLPETVLRLSE-VK 167
Leptospira 106 KDGVDGILSVNPYYNKPTQEGLFQHFKSIAEHSTVPVMLYNIPGRTSVNLLPETVLRLSE-VK 167
Leptospira_borg 93 KDSVDGILSVNPYYNKPTQEGLFQHFQAVAEHSSVPVMLYNIPGRTSINLQPETVFRLSE-IK 154
Leptospira_borg 93 KDSVDGILSVNPYYNKPTQEGLFQHFQAVAEHSSVPVMLYNIPGRTSINLQPETVFRLSE-IK 154
cons 127 189
Brachyspira_int 160 NIQGIKE-ASGDISYVADIVNI--------APKLDLYSGNDDMVTPMMALGAK----GVISVT 209
Brachyspira_pil 158 NIVAIKE-ASGDMSYTADIANI--------APKLDLYSGNDDMVTPILALGGK----GVISVT 207
Brachyspira_hyo 158 NIQGIKE-ASGDISYVADIVNV--------APKLDLYSGNDDMVTPMMALGAK----GVISVT 207
Brachyspira_mur 158 NIQGIKE-ASGDIGYVADIANI--------APKLDIYSGNDDMVTPMMALGAK----GVISVT 207
Brachyspira_pil 158 NIVAIKE-ASGDMSYTADIANI--------APKLDLYSGNDDMVTPILALGGK----GVISVT 207
Brachyspira_pil 158 NIVAIKE-ASGDMSYTADIANI--------APKLDLYSGNDDMVTPILALGGK----GVISVT 207
Brachyspirapilo 158 NIVAIKE-ASGDMSYTADIANI--------APKLDLYSGNDDMVTPILALGGK----GVISVT 207
Brachyspira_pil 158 NIVAIKE-ASGDMSYTADIANI--------APKLDLYSGNDDMVTPILALGGK----GVISVT 207
Spirochaeta_the 156 LIAGVKE-ASGSMAQVMEIIAKR-------PDGFAVLSGDDNLTLPIVALGGD----GVVSVA 206
Treponema_calda 158 TVIGVKE-ASGDLAQMGDIIQMVAQPRRAEGKPFAVLSGDDAFTLALCALGGD----GVVSVV 215
Spirochaeta_afr 157 QVIGIKE-AAGSIPQAMEIISQR-------PKNFGVVAGDDNLTLPLIQLGAD----GVISVA 207
Treponema_primi 165 GIIGVKE-ASGDITQMGDIIREIALPGKAAGKPFTVLSGDDSFTLPLIALGGD----GVVSVI 222
Spirochaeta_sma 164 NIVAIKEEAELNAKQMTEFINAT-------PSDFIIYNGDDTMILEAYTQGGDKRIGGVISGQ 219
Sphaerochaeta_c 161 NIVAVKE-ASGDLNQMKDVIAHR-------RHGFRVLSGDDNMVLDLIEAGGD----GVISVA 211
Turneriella_par 166 NIIGIKE-ATGDLGFMAQIIQAT-------DAGFTLLSGDDNLILPVLSVGGK----GVISVI 216
Leptospira_bifl 155 KIAAIKE-ATGDLGQMAKVISQC-------PTNFDLLSGDDNLTLPVLSIGGK----GVVSVV 205
Leptospira_bifl 155 KIAAIKE-ATGDLGQMAKVISQC-------PTNFDLLSGDDNLTLPVLSIGGK----GVVSVV 205
Treponema_succi 167 NIAGVKE-ASGNISQMMDVIAKIKS----KKQDFAVLSGDDGLTLPLMATGGD----GVISVV 220
Leptospira_inte 168 QIRSMKE-ATGDLGQMGKLISLV-------GNKMTVLSGDDNLTLPLLAIGGV----GVVSVI 218
Leptospira_inte 168 QIRSMKE-ATGDLGQMGKLISLV-------GNKMTVLSGDDNLTLPLLAIGGV----GVVSVI 218
Leptospira 168 QIRSMKE-ATGDLGQMGKLISLV-------GNKMTVLSGDDNLTLPLLAIGGV----GVVSVI 218
Leptospira_borg 155 RIRSMKE-ATGDLGQMGKLISLV-------GNKMTVLSGDDNLTLPLLAIGGV----GVVSVV 205
Leptospira_borg 155 RIRSMKE-ATGDLGQMGKLISLV-------GNKMTVLSGDDNLTLPLLAIGGV----GVVSVV 205
cons 190 252
Brachyspira_int 210 SNIIPKENHDMVMNFLNGNVNEAIKTQIKYIDFVRAMF--IETNPAP-IKEAMNIMGFNV--G 267
Brachyspira_pil 208 SNIIPKENHDMVMNFLNGKVEESIKAQVHYIDLVRAMF--IEVNPVP-VKEAMNIMGFDV--G 265
Brachyspira_hyo 208 SNIIPKENHDMVMNFLNGNVNEAIKTQIKYIDFVRAMF--IETNPAP-IKEAMNIMGFNV--G 265
Brachyspira_mur 208 SNIIPKENHDMVMNYINGNIKEAVEAQIKYIDLIRAMF--IEVNPVP-IKEAMNMMGFNV--G 265
Brachyspira_pil 208 SNIIPKENHDMVMNFLNGKVEESIKAQVHYIDLVRAMF--IEVNPVP-VKEAMNIMGFDV--G 265
Brachyspira_pil 208 SNIIPKENHDMVMNFLNGKVEESIKAQVHYIDLVRAMF--IEVNPVP-VKEAMNIMGFDV--G 265
Brachyspirapilo 208 SNTIPKENHDMVMNFLNGKVEESIKAQVHYIDLVRAMF--IEVNPVP-VKEAMNIMGFDV--G 265
Brachyspira_pil 208 SNIIPKENHDMVMNFLNSKVEESIKAQVHYIDLVRAMF--IEVNPVP-VKEAMNIMGFDV--G 265
Spirochaeta_the 207 SNLVPGRMSAFVKALLEGRMEEARRRHYELLPLFKAIF--IDTNPIP-VKYAMTLLSSLSIQE 266
Treponema_calda 216 SNLVPDRVSSLAKACLAGNFAKARELHYSLLPLFKGAF--IETNPIP-IKTAMGWAGLPA--G 273
Spirochaeta_afr 208 SNLIPGHMQQLVTAALKDQVETARSLHYQLLPLFKALF--WETNPIP-IKYALASRGIIA-HN 266
Treponema_primi 223 SNLIPKKIAALTKACLEGNFEEGRKLHYELLPFAKTAF--IETNPVP-IKAAMNWAGLPA--G 280
Spirochaeta_sma 220 SHIFGDKIRTMINTFLSGDIEAAALQQQQLLPVFRIMGQNNRTNPVALLKDTMKLYGIDG--G 280
Sphaerochaeta_c 212 SNLFPRLMSTMVAQALAGNRDGSRALGAQLDEFFHSCF--LETNPIP-IKTAMAMCGWCN--E 269
Turneriella_par 217 SNVYPRETGAITRQYEAGNAHEAHEKFMRLLPLCKAMF--LETNPIP-VKFAASRKGLCQ--N 274
Leptospira_bifl 206 SNLFPRACVDMVSLYLRGDLEASKKIYYKLLPVFVNAF--IETNPIP-IKAAMSWFGYCS--N 263
Leptospira_bifl 206 SNLFPRACVDMVSLYLRGDLEASKKIYYKLLPVFVNAF--IETNPIP-IKAAMSWFGYCS--N 263
Treponema_succi 221 SNLAPALVTQMVTEGLKGNFDKAREIHYRLLPFFKAAF--VDGNPTS-IKYAMSIKGFCK--P 278
Leptospira_inte 219 SNLFPKALVQLVESFQQGKISEAKKIHYDFIEVFALAF--METNPIP-IKAAMCWFGHCG--P 276
Leptospira_inte 219 SNLFPKALVQLVESFQQGKISEAKKIHYDFIEVFALAF--METNPIP-IKAAMCWFGHCG--P 276
Leptospira 219 SNLFPKALVQLVESFQQGKISEAKKIHYDFIEVFALAF--METNPIP-IKAAMCWFGHCG--P 276
Leptospira_borg 206 SNLFPKAMVEMVKYFQDGKVIEARKIHYDFIEAFALAF--METNPIP-IKAAMSWFGHCE--P 263
Leptospira_borg 206 SNLFPKAMVEMVKYFQDGRVIEARKIHYDFIEAFALAF--METNPIP-IKAAMSWFGHCE--P 263
cons 253 315



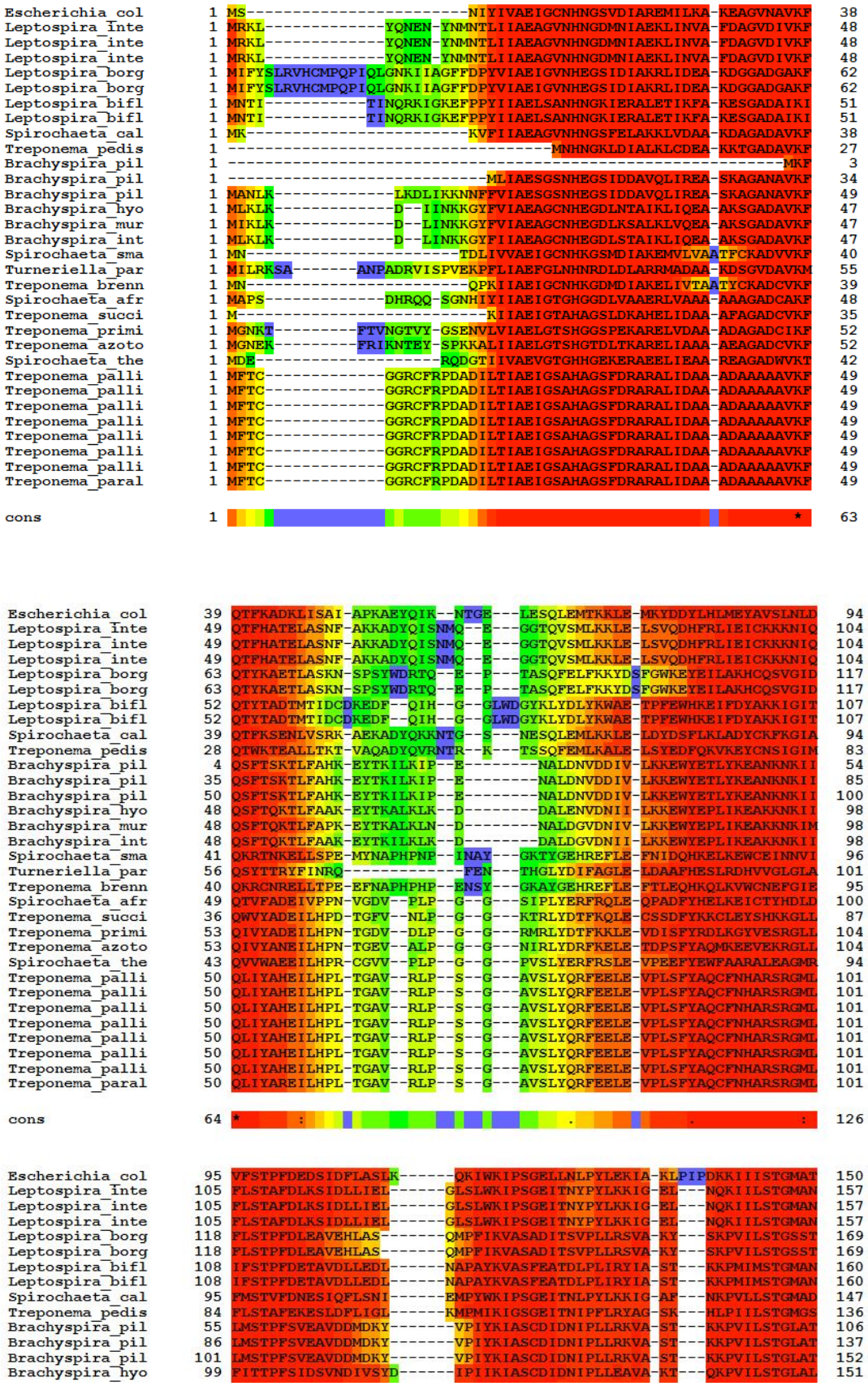
ANEXO 7 (neuA) - score : 82

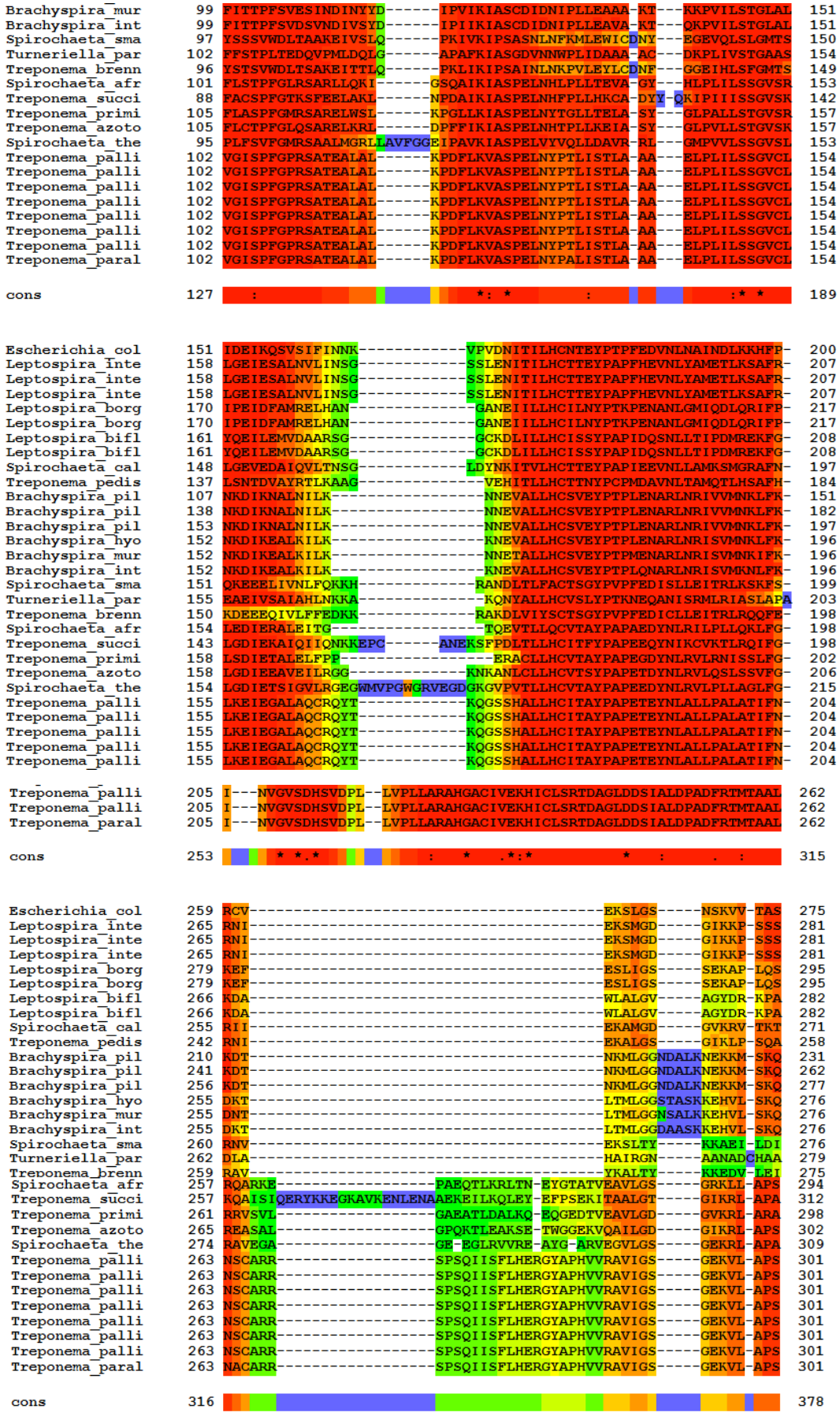


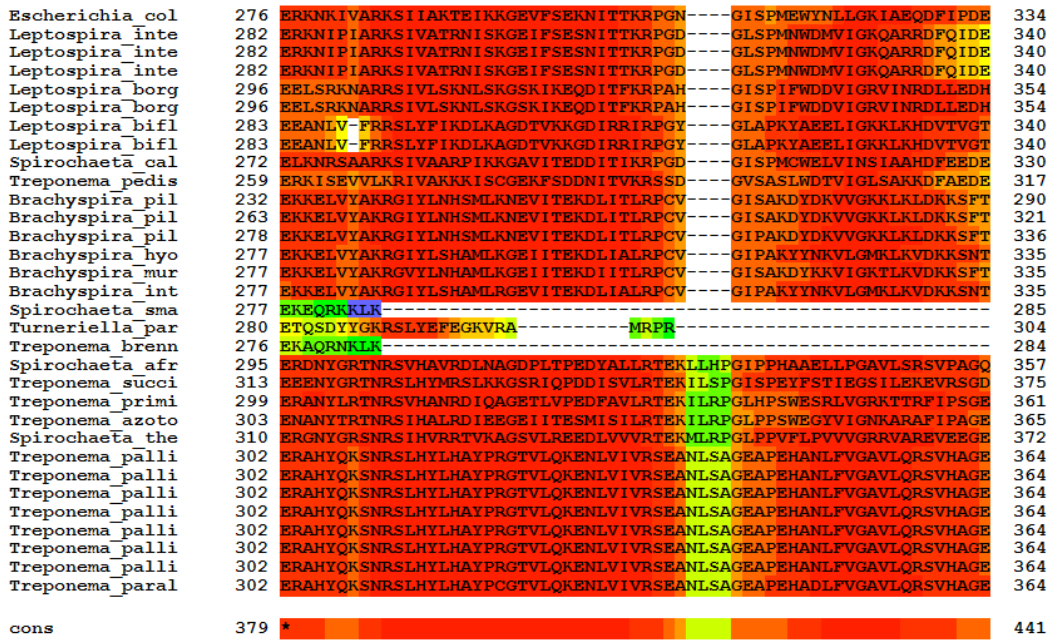

Leptospira_inte 119 CIEFTMRKK--T----------NSVVSVCEAQENP---FWMF-RLDDDS-FKMT 155
Leptospira_inte 119 CIEFTMRKK--T----------NSVVSVCEAQENP---FWMF-RLDDDS-FKMT 155
Leptospira_inte 119 CIEFTMRKK--T----------NSVVSVCEAQENP---FWMF-RLDDDS-FKMT 155
Leptospira_borg 112 VAGLKASH-PEWKMSTAAVPLIDFS-HGEDPNRVKVIIDRNGKA---IYFS-RSL--------- 160
Leptospira_borg 112 VAGLKASH-PEWKMSTAAVPLIDFS-HGEDPNRVKVIIDRNGKA---IYFS-RSL--------- 160
Leptospira_bifl 113 VVGLKTKH-RNWEMTTAAVPFTSAE-DPKDPNKVKVVFDNKGRA---NYFS-RSP--------- 161
Leptospira_bifl 113 VVGLKTKH-RNWEMTTAAVPFTSAE-DPKDPNKVKVVFDNKGRA---NYFS-RSP--------- 161
Brachyspira_pil 111 VIDAFDDEN--VDIATLKTKIDDYSPLIKDENAVKVVCDFKDYA---MYFS-RAT--------- 159
Brachispira_int 111 LIDAFDDEN--VDIATLKTRIEDDSQLIKDENAVKVVTDINNYA---MYFS-RAT--------- 159
Brachyspira_pil 111 VIDAFDDEN--VDIATLKTKIDDYSPLIKDENAVKVVCDYKDYA---MYFS-RAT--------- 159
Brachyspira_pil 111 VIDAFDDEN--VDIATLKTKIDDYSPLIKDENAVKVVCDYKDYA---MYFS-RAT--------- 159
Brachyspira_hyo 111 LIDAFDDEN--VDIATLKTKIDDDSPLIKDENAVKVVTDINNYA---MYFS-RAT--------- 159
Brachyspira_pil 111 VIDAFDDEN--VDIATLKTKIDDYSPLIKDENAVKVVCDFKDYA---MYFS-RAT--------- 159
Brachyspira_mur 112 IIDVFDDSS--VDIATLKVKIEDDSSLIKDENAVKVVTDINDYA---MYFS-RAA--------- 160
Turneriella_par 111 AVALFREHGEDCHIATTAVPFSNEA-DEANENMVKVVLAHDARA---LYFS-RSR--------- 160
Spirochaeta_sma 120 ALRKYEIDN--Y----------DSMLSVVRFKRFL---WGDS-G----A-PINY 152
Treponema_brenn 119 GLELYNSRK--F----------DSILTCIRNYRFF---WNDD-G----T-SKNY 151
Spirochaeta_cal 123 AWNLFIQGG--Y----------DNVITAMPSRKSP---YFNIVEVDENGNVHLS 161
Treponema_pedis 124 AIELYIDEL--DMVVSVKT--S---------DIAAVICEENSNGFLEMSFS-K---------- 162
cons 127 : 189
Leptospira_inte 156 KLLNGEDVA-RRQDSPKVFTLNGSIYLSEVNYFREKKKFITDDTLAYLMSKENS--------- 208
Leptospira_inte 156 KLLNGEDVA-RRQDSPKVFTLNGSIYLSEVNYFREKKKFITDDTLAYLMSKENS--------- 208
Leptospira_inte 156 KLLNGEDVA-RRQDSPKVFTLNGSIYLSEVNYFREKKKFITDDTLAYLMSKENS--------- 208
Leptospira_borg 161 -----IPSQ-FKQ--TVPLYRHLGIYGYDRDFL----------LKYNS---- 190
Leptospira_borg 161 -----IPSQ-FKQ--TVPLYRHLGIYGYDRDFL----------LKYNS---- 190
Leptospira_bifl 162 -----IPAS-FKG--EAKYHRHLGIYAYERDFL----------MNYNQ---- 191
Leptospira_bifl 162 -----IPAS-FKG--EAKYHRHLGIYAYERDFL----------MNYNQ---- 191
Brachyspira_pil 160 -----IPYKRFDQDILVKYYKHIGVYAFRREAL----------LKIEN---- 192
Brachispira_int 160 -----IPHKRFDQNISAEYYKHIGVYAFRKDIL----------LKIEK---- 192
Brachyspira_pil 160 -----IPYKRFDQDILVKYYKHIGVYAFRREAL----------LKIEN---- 192
Brachyspira_pil 160 -----IPYKRFDQDILVKYYKHIGVYAFRREAL----------LKIEN---- 192
Brachyspira_hyo 160 -----IPHKRFDQNISAEYYKHIGVYAFRKDIL----------LKIEK---- 192
Brachyspira_pil 160 -----IPYKRFDQDILVKYYKHIGVYAFRREAL----------LKIEN---- 192
Brachyspira_mur 161 -----IPHRRFDQNINVKYYKHIGIYAFRKDIL----------LKIEN---- 193
Turneriella_par 161 -----IPFYRKDVRASLPALKHQGLYAYSREAL----------LKLAK---- 193
Spirochaeta_sma 153 DYH---NRP-RRQKYKGCLLENGAFYINKVANIIKYKNRLCGNIGYYLMPEYTWLEIDEEDDW 211
Treponema_brenn 152 EYM---NRP-RRQNFDGMFMENGAFYINSVANIKKFRNRLSGKIGIYEMPEYTATEIDEPDDW 210
Spirochaeta_cal 162 KTLE-KPIV-RRQDAPKCFDMNASIYIWKRETLLGCDSVITENTGLYIMPEERS--------- 213
Treponema_pedis 163 ------NAT-RRQDVKPYYEYNGAVYVINPKSLKDKGMFHFTKIKKTVMDTIRS--------- 209
cons 190 : ..* : 252
Leptospira_inte 209 ---------- 208
Leptospira_inte 209 ---------- 208
Leptospira_inte 209 ---------- 208
Leptospira_borg 191 ----------LPKSNLEESESLEQLRA--IEAGYGIGVYLA 219
Leptospira_borg 191 ----------LPKSNLEESESLEQLRA--IEAGYGIGVYLA 219
Leptospira_bifl 192 ----------LPPSDWETVESLEQLRA--LQNGSTIGVYLS 220
Leptospira_bifl 192 ----------LPPSDWETVESLEQLRA--LQNGSTIGVYLS 220
Brachyspira_pil 193 ----------LQECEYEKIEKLEQLRW--LYHGMKIKVLET 221
Brachispira_int 193 ----------LEECKYEMIEKLEQLRW--LYNGIKIKVLET 221
Brachyspira_pil 193 ----------LQECEYEKIEKLEQLRW--LYHGMKIKVLET 221
Brachyspira_pil 193 ----------LQECEYEKIEKLEQLRW--LYHGMKIKVLET 221
Brachyspira_hyo 193 ----------LEECRYELIEKLEQLRW--LYNGMKIKVLET 221
Brachyspira_pil 193 ----------LQECEYEKIEKLEQLRW--LYHGMKIKVLET 221
Brachyspira_mur 194 ----------LKECDYESIEKLEQLRW--LYNGMKIKVLET 222
Turneriella_par 194 ----------LPVAALEEAESLEQLRF--LYNGFNIFVHIA 222
Spirochaeta_sma 212 IFAEKIMKKYRLD-GETKTEEIKIVFSDVDGCLTDAGMYYSENGEELKKFSTYDGMAVELLKN 273
Treponema_brenn 211 IVLENLMRRHVLVKQTSDYSKIKLLISDVDGVLTDGGMYYSENGDELKKFNTRDGMAFQLLRE 273
Spirochaeta_cal 214 ---------- 213
Treponema_pedis 210 ---------- 209
cons 253 315


Leptospira_inte 209 ----IDIDDMIDW---------- 217
Leptospira_inte 209 ----IDIDDMIDW---------- 217
Leptospira_inte 209 ----IDIDDMIDW---------- 217
Leptospira_borg 220 QEAGLSVDTPADLEVVIEDFKKRK---------- 243
Leptospira_borg 220 QEAGLSVDTPADLEVVIEDFKKRK---------- 243
Leptospira_bifl 221 DKANLGVDSPADLEVVITEFKKKG---------- 244
Leptospira_bifl 221 DKANLGVDSPADLEVVITEFKKKG---------- 244
Brachyspira_pil 222 NHFIHGIDTKEDLEIVQEYLKKEA---------- 245
Brachispira_int 222 KKFLHGIDTREDLEFVQNYLRNFA---------- 245
Brachyspira_pil 222 NHFIHGIDTKEDLEIVQEYLKKEA---------- 245
Brachyspira_pil 222 NHFIHGIDTKEDLEIVQEYLKKEA---------- 245
Brachyspira_hyo 222 KKFLHGIDTREDLEFVQNYLRNFA---------- 245
Brachyspira_pil 222 NHFIHGIDTKEDLEIVQEYLKKEA---------- 245
Brachyspira_mur 223 KKFLHGIDTMEDLEFVQNYLRNFA---------- 246
Turneriella_par 223 REDSPGVDVPEDIKRVEQILQH---------- 244
Spirochaeta_sma 274 KGIRTGIITGEKTELVKRRAEKIKCNYLYIGIQNK--LKIIKEICNNENVNMENVAYLGDDIN 334
Treponema_brenn 274 HGIKTAIITSENTKMVEYRAKKLNIDFLKQGKRNGGKLEAAQEICRNLGISLSEVAFVGDDVN 336
Spirochaeta_cal 214 ----IDIDTELDF---------- 222
Treponema_pedis 210 ----IDIDTNFDW---------- 218
cons 316 : . 378
Leptospira_inte 218 ----------KLAEIL----------LL----------DKV 228
Leptospira_inte 218 ----------KLAEIL----------LL----------DKV 228
Leptospira_inte 218 ----------KLAEIL----------LL----------DKV 228
Leptospira_borg 244 ----------WIT 246
Leptospira_borg 244 ----------WIT 246
Leptospira_bifl 245 ----------LL 246
Leptospira_bifl 245 ----------LL 246
Brachyspira_pil 246 ----------LIKLG---NNA 253
Brachispira_int 246 ----------QSELNNENLYK 256
Brachyspira_pil 246 ----------LIKLG---NNA 253
Brachyspira_pil 246 ----------LIKLG---NNA 253
Brachyspira_hyo 246 ----------QSELNNENLYK 256
Brachyspira_pil 246 ----------LIKLG---NNA 253
Brachyspira_mur 247 ----------ESELNNENIYK 257
Turneriella_par 245 ----------AV 246
Spirochaeta_sma 335 DIQVLKAVGHPFCPINARNDIKRIPGIIILKSAGGDGAFREMAERIITSLSGEDYEY 391
Treponema_brenn 337 DIELLEHVGFAFCPADAAETVKAVCSVCELSAKG--------GEGVIREIV-GFLLV 384
Spirochaeta_cal 223 ----------AFVEFL----------FS----------KRLV 234
Treponema_pedis 219 ----------QFAEFI----------LH--------RQDAGDLAK 235
cons 379 435

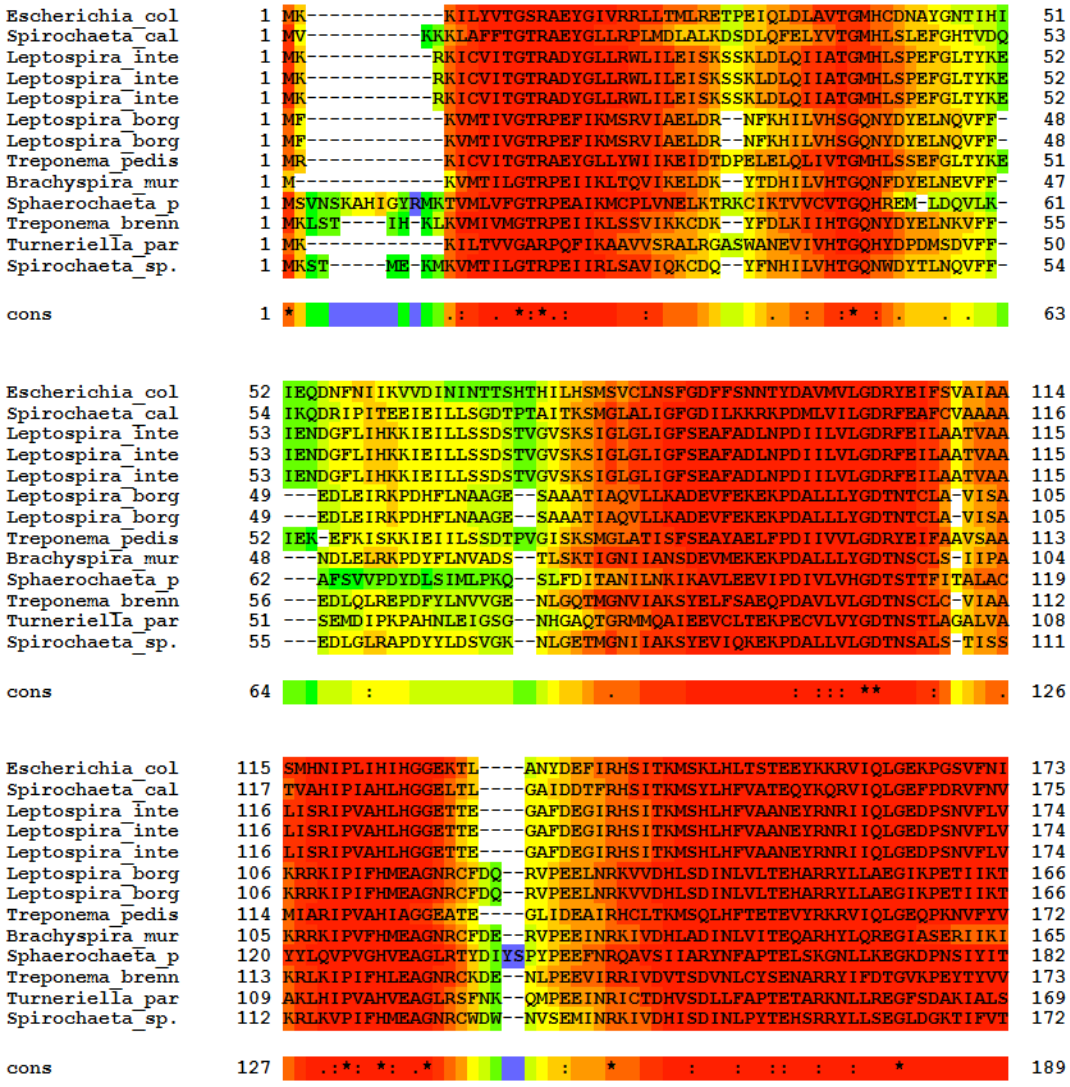
ANEXO 8 (neuB) - score : 88

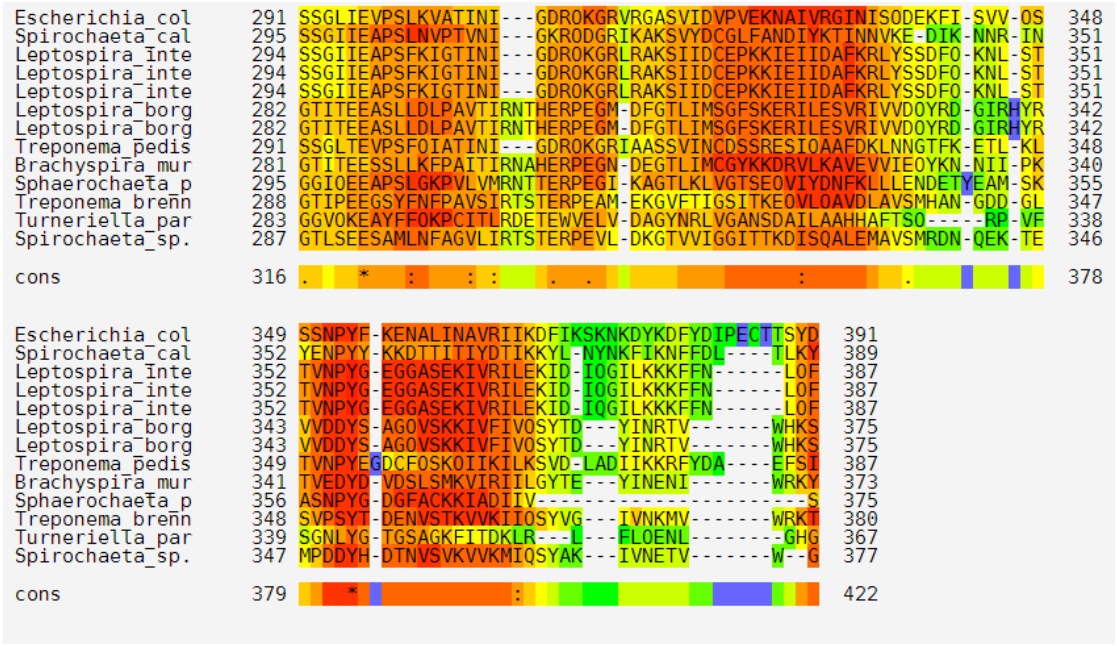



Brachyspira_mur 99 FITTPFSVESINDINYYD------IPVIKIASCDIDNIPLLEAAA-KT---KKPVILSTGLAL 151
Brachyspira_int 99 FITTPFSVDSVNDIVSYD------IPIIKIASCDIDNIPLLEAVA-KT---QKPVILSTGLAL 151
Spirochaeta_sma 97 YSSSVWDLTAAKEIVSLQ------PKIVKIPSASNLNFKMLEWICDNY---EGEVQLSLGMTS 150
Turneriella_par 102 FFSTPLTEDQVPMLDQLG------APAFKIASGDVNNWPLIDAAA-AC---DKPLIVSTGAAS 154
Treponema_brenn 96 YSTSVWDLTSAKEITTLQ------PKLIKIPSAINLNKPVLEYLCDNF---GGEIHLSFGMTS 149
Spirochaeta_afr 101 FLSTPFGLRSARLLQKI------GSQAIKIASPELNHLPLLTEVA-GY---HLPLILSSGVSR 153
Treponema_succi 88 FACSPFGTKSFEELAKL------NPDAIKIASPELNHFPLLHKCA-DYY-QKIPIIISSGVSK 142
Treponema_primi 105 FLASPFGMRSARELWSL------KPGLLKIASPELNYTGLLTELA-SY---GLPALLSTGVSR 157
Treponema_azoto 105 FLCTPFGLQSARELKRL------DPFFIKIASPELNHTPLLKEIA-SY---GLPVLLSTGVSK 157
Spirochaeta_the 95 PLFSVFGMRSAALMGRLLAVFGGEIPAVKIASPELVYVQLLDAVR-RL---GMPVVLSSGVSL 153
Treponema_palli 102 VGISPFGPRSATEALAL------KPDFLKVASPELNYPTLISTLA-AA---ELPLILSSGVCL 154
Treponema_palli 102 VGISPFGPRSATEALAL------KPDFLKVASPELNYPTLISTLA-AA---ELPLILSSGVCL 154
Treponema_palli 102 VGISPFGPRSATEALAL------KPDFLKVASPELNYPTLISTLA-AA---ELPLILSSGVCL 154
Treponema_palli 102 VGISPFGPRSATEALAL------KPDFLKVASPELNYPTLISTLA-AA---ELPLILSSGVCL 154
Treponema_palli 102 VGISPFGPRSATEALAL------KPDFLKVASPELNYPTLISTLA-AA---ELPLILSSGVCL 154
Treponema_palli 102 VGISPFGPRSATEALAL------KPDFLKVASPELNYPTLISTLA-AA---ELPLILSSGVCL 154
Treponema_palli 102 VGISPFGPRSATEALAL------KPDFLKVASPELNYPTLISTLA-AA---ELPLILSSGVCL 154
Treponema_paral 102 VGISPFGPRSATEALAL------KPDFLKVASPELNYPALISTLA-AA---ELPLILSSGVCL 154
cons 127 189
Escherichia_col 151 IDEIKQSVSIFINNK----------VPVDNITILHCNTEYPTPFEDVNLNAINDLKKHFP- 200
Leptospira_inte 158 LGEIESALNVLINSG----------SSLENITILHCTTEYPAPFHEVNLYAMETLKSAFR- 207
Leptospira_inte 158 LGEIESALNVLINSG----------SSLENITILHCTTEYPAPFHEVNLYAMETLKSAFR- 207
Leptospira_inte 158 LGEIESALNVLINSG----------SSLENITILHCTTEYPAPFHEVNLYAMETLKSAFR- 207
Leptospira_borg 170 IPEIDFAMRELHAN----------GANEIILLHCILNYPTKPENANLGMIQDLQRIFP- 217
Leptospira_borg 170 IPEIDFAMRELHAN----------GANEIILLHCILNYPTKPENANLGMIQDLQRIFP- 217
Leptospira_bifl 161 YQEILEMVDAARSG----------GCKDLILLHCISSYPAPIDQSNLLTIPDMREKFG- 208
Leptospira_bifl 161 YQEILEMVDAARSG----------GCKDLILLHCISSYPAPIDQSNLLTIPDMREKFG- 208
Spirochaeta_cal 148 LGEVEDAIQVLTNSG----------LDYNKITVLHCTTEYPAPIEEVNLLAMKSMGRAFN- 197
Treponema_pedis 137 LSNTDVAYRTLKAAG----------VEHITLLHCTTNYPCPMDAVNLTAMQTLHSAFH- 184
Brachyspira_pil 107 NKDIKNALNILK----------NNEVALLHCSVEYPTPLENARLNRIVVMNKLFK- 151
Brachyspira_pil 138 NKDIKNALNILK----------NNEVALLHCSVEYPTPLENARLNRIVVMNKLFK- 182
Brachyspira_pil 153 NKDIKNALNILK----------NNEVALLHCSVEYPTPLENARLNRIVVMNKLFK- 197
Brachyspira_hyo 152 NKDIKEALKILK----------KNEVALLHCSVEYPTPLENARLNRISVMNKLFK- 196
Brachyspira_mur 152 NKDIKEALKILK----------NNETALLHCSVEYPTPMENARLNRISVMNKIFK- 196
Brachyspira_int 152 NKDIKEALKILK----------KNEVALLHCSVEYPTPLQNARLNRISVMKNLFK- 196
Spirochaeta_sma 151 QKEEELIVNLFQKKH----------RANDLTLFACTSGYPVPFEDISLLEITRLKSKFS- 199
Turneriella_par 155 EAEIVSAIAHLNKKA----------KQNYALLHCVSLYPTKNEQANISRMLRIASLAPA 203
Treponema_brenn 150 KDEEEQIVLFFEDKK----------RAKDLVIYSCTSGYPVPFEDICLLEITRLRQQFE- 198
Spirochaeta_afr 154 LEDIERALEITG----------TQEVTLLQCVTAYPAPAEDYNLRILPLLQKLFG- 198
Treponema_succi 143 LGDIEKAIQIIQNKKEPC------ANEKSFPDLTLLHCITFYPAPEEQYNIKCVKTLRQIFG- 198
Treponema_primi 158 LSDIETALELFPF----------ERACLLHCVTAYPAPEGDYNLRVLRNISSLFG- 202
Treponema_azoto 158 LGDIEEAVEILRGG----------KNKANLCLLHCVTSYPAPETDYNLRVLQSLSSVFG- 206
Spirochaeta_the 154 LGDIETSIGVLRGEGWMVPGWGRVEGDGKGVPVTLLHCVTAYPAPEEDYNLRVLPLLAGLFG- 215
Treponema_palli 155 LKEIEGALAQCRQYT----------KQGSSHALLHCITAYPAPETEYNLALLPALATIFN- 204
Treponema_palli 155 LKEIEGALAQCRQYT----------KQGSSHALLHCITAYPAPETEYNLALLPALATIFN- 204
Treponema_palli 155 LKEIEGALAQCRQYT----------KQGSSHALLHCITAYPAPETEYNLALLPALATIFN- 204
Treponema_palli 155 LKEIEGALAQCRQYT----------KQGSSHALLHCITAYPAPETEYNLALLPALATIFN- 204
Treponema_palli 155 LKEIEGALAQCRQYT----------KQGSSHALLHCITAYPAPETEYNLALLPALATIFN- 204
Treponema_palli 205 I---NVGVSDHSVDPL--LVPLLARAHGACIVEKHICLSRTDAGLDDSIALDPADFRTMTAAL 262
Treponema_palli 205 I---NVGVSDHSVDPL--LVPLLARAHGACIVEKHICLSRTDAGLDDSIALDPADFRTMTAAL 262
Treponema_paral 205 I---NVGVSDHSVDPL--LVPLLARAHGACIVEKHICLSRTDAGLDDSIALDPADFRTMTAAL 262
cons 253 315
Escherichia_col 259 RCV----------EKSLGS-----NSKVV-TAS 275
Leptospira_inte 265 RNI----------EKSMGD-----GIKKP-SSS 281
Leptospira_inte 265 RNI----------EKSMGD-----GIKKP-SSS 281
Leptospira_inte 265 RNI----------EKSMGD-----GIKKP-SSS 281
Leptospira_borg 279 KEF----------ESLIGS-----SEKAP-LQS 295
Leptospira_borg 279 KEF----------ESLIGS-----SEKAP-LQS 295
Leptospira_bifl 266 KDA----------WLALGV-----AGYDR-KPA 282
Leptospira_bifl 266 KDA----------WLALGV-----AGYDR-KPA 282
Spirochaeta_cal 255 RII----------EKAMGD-----GVKRV-TKT 271
Treponema_pedis 242 RNI----------EKALGS-----GIKLP-SQA 258
Brachyspira_pil 210 KDT----------NKMLGGNDALKNEKKM-SKQ 231
Brachyspira_pil 241 KDT----------NKMLGGNDALKNEKKM-SKQ 262
Brachyspira_pil 256 KDT----------NKMLGGNDALKNEKKM-SKQ 277
Brachyspira_hyo 255 DKT----------LTMLGGSTASKKEHVL-SKQ 276
Brachyspira_mur 255 DNT----------LTMLGGNSALKKEHVL-SKQ 276
Brachyspira_int 255 DKT----------LTMLGGDAASKKEHVL-SKQ 276
Spirochaeta_sma 260 RNV----------EKSLTY-----KKAEI-LDI 276
Turneriella_par 262 DLA----------HAIRGN-----AANADCHAA 279
Treponema_brenn 259 RAV----------YKALTY-----KKEDV-LEI 275
Spirochaeta_afr 257 RQARKE----------PAEQTLKRLTN-EYGTATVEAVLGS-----GRKLL-APS 294
Treponema_succi 257 KQAISIQERYKKEGKAVKENLENAAEKEILKQLEY-EFPSEKITAALGT-----GIKRL-APA 312
Treponema_primi 261 RRVSVL----------GAEATLDALKQ-EQGEDTVEAVLGD-----GVKRL-ARA 298
Treponema_azoto 265 REASAL----------GPQKTLEAKSE-TWGGEKVQAILGD-----GIKRL-APS 302
Spirochaeta_the 274 RAVEGA----------GE-EGLRVVRE-AYG-ARVEGVLGS-----GEKRL-APA 309
Treponema_palli 263 NSCARR----------SPSQIISFLHERGYAPHVVRAVIGS-----GEKVL-APS 301
Treponema_palli 263 NSCARR----------SPSQIISFLHERGYAPHVVRAVIGS-----GEKVL-APS 301
Treponema_palli 263 NSCARR----------SPSQIISFLHERGYAPHVVRAVIGS-----GEKVL-APS 301
Treponema_palli 263 NSCARR----------SPSQIISFLHERGYAPHVVRAVIGS-----GEKVL-APS 301
Treponema_palli 263 NSCARR----------SPSQIISFLHERGYAPHVVRAVIGS-----GEKVL-APS 301
Treponema_palli 263 NSCARR----------SPSQIISFLHERGYAPHVVRAVIGS-----GEKVL-APS 301
Treponema_palli 263 NSCARR----------SPSQIISFLHERGYAPHVVRAVIGS-----GEKVL-APS 301
Treponema_paral 263 NACARR----------SPSQIISFLHERGYAPHVVRAVIGS-----GEKVL-APS 301
cons 316 378


Escherichia_col 276 ERKNKIVARKSIIAKTEIKKGEVFSEKNITTKRPGN----GISPMEWYNLLGKIAEQDFIPDE 334
Leptospira_inte 282 ERKNIPIARKSIVATRNISKGEIFSESNITTKRPGD----GLSPMNWDMVIGKQARRDFQIDE 340
Leptospira_inte 282 ERKNIPIARKSIVATRNISKGEIFSESNITTKRPGD----GLSPMNWDMVIGKQARRDFQIDE 340
Leptospira_inte 282 ERKNIPIARKSIVATRNISKGEIFSESNITTKRPGD----GLSPMNWDMVIGKQARRDFQIDE 340
Leptospira_borg 296 EELSRKNARRSIVLSKNLSKGSKIKEQDITFKRPAH----GISPIFWDDVIGRVINRDLLEDH 354
Leptospira_borg 296 EELSRKNARRSIVLSKNLSKGSKIKEQDITFKRPAH----GISPIFWDDVIGRVINRDLLEDH 354
Leptospira_bifl 283 EEANLV-FRRSLYFIKDLKAGDTVKKGDIRRIRPGY----GLAPKYAEELIGKKLKHDVTVGT 340
Leptospira_bifl 283 EEANLV-FRRSLYFIKDLKAGDTVKKGDIRRIRPGY----GLAPKYAEELIGKKLKHDVTVGT 340
Spirochaeta_cal 272 ELKNRSAARKSIVAARPIKKGAVITEDDITIKRPGD----GISPMCWELVINSIAAHDFEEDE 330
Treponema_pedis 259 ERKISEVVLKRIVAKKKISCGEKFSDDNITVKRSSD----GVSASLWDTVIGLSAKKDFAEDE 317
Brachyspira_pil 232 EKKELVYAKRGIYLNHSMLKNEVITEKDLITLRPCV----GISAKDYDKVVGKKLKLDKKSFT 290
Brachyspira_pil 263 EKKELVYAKRGIYLNHSMLKNEVITEKDLITLRPCV----GISAKDYDKVVGKKLKLDKKSFT 321
Brachyspira_pil 278 EKKELVYAKRGIYLNHSMLKNEVITEKDLITLRPCV----GIPAKDYDKVVGKKLKLDKKSFT 336
Brachyspira_hyo 277 EKKELVYAKRGIYLSHAMLKGEIITEKDLIALRPCV----GIPAKYYNKVLGMKLKVDKKSNT 335
Brachyspira_mur 277 EKKELVYAKRGVYLNHAMLKGEIITEKDIITLRPCV----GISAKDYKKVIGKTLKVDKKSFT 335
Brachyspira_int 277 EKKELVYAKRGIYLSHAMLRGEVITEKDLIALRPCV----GIPAKYYNKVLGMKLKVDKKSNT 335
Spirochaeta_sma 277 EKEQRKKLK---------- 285
Turneriella_par 280 ETQSDYYGKRSLYEFEGKVRA----------MRPR---------- 304
Treponema_brenn 276 EKAQRNKLK---------- 284
Spirochaeta_afr 295 ERDNYGRTNRSVHAVRDLNAGDPLTPEDYALLRTEKLLHPGIPPHAAELLPGAVLSRSVPAGQ 357
Treponema_succi 313 EEENYGRTNRSLHYMRSLKKGSRIQPDDISVLRTEKILSPGISPEYFSTIEGSILEKEVRSGD 375
Treponema_primi 299 ERANYLRTNRSVHANRDIQAGETLVPEDFAVLRTEKILRPGLHPSWESRLVGRKTTRFIPSGE 361
Treponema_azoto 303 ENANYTRTNRSIHALRDIEEGEIITESMISILRTEKILRPGLPPSWEGYVIGNKARAFIPAGE 365
Spirochaeta_the 310 ERGNYGRSNRSIHVRRTVKAGSVLREEDLVVVRTEKMLRPGLPPVFLPVVVGRRVAREVEEGE 372
Treponema_palli 302 ERAHYQKSNRSLHYLHAYPRGTVLQKENLVIVRSEANLSAGEAPEHANLFVGAVLQRSVHAGE 364
Treponema_palli 302 ERAHYQKSNRSLHYLHAYPRGTVLQKENLVIVRSEANLSAGEAPEHANLFVGAVLQRSVHAGE 364
Treponema_palli 302 ERAHYQKSNRSLHYLHAYPRGTVLQKENLVIVRSEANLSAGEAPEHANLFVGAVLQRSVHAGE 364
Treponema_palli 302 ERAHYQKSNRSLHYLHAYPRGTVLQKENLVIVRSEANLSAGEAPEHANLFVGAVLQRSVHAGE 364
Treponema_palli 302 ERAHYQKSNRSLHYLHAYPRGTVLQKENLVIVRSEANLSAGEAPEHANLFVGAVLQRSVHAGE 364
Treponema_palli 302 ERAHYQKSNRSLHYLHAYPRGTVLQKENLVIVRSEANLSAGEAPEHANLFVGAVLQRSVHAGE 364
Treponema_palli 302 ERAHYQKSNRSLHYLHAYPRGTVLQKENLVIVRSEANLSAGEAPEHANLFVGAVLQRSVHAGE 364
Treponema_paral 302 ERAHYQKSNRSLHYLHAYPCGTVLQKENLVIVRSEANLSAGEAPEHADLFVGAVLQRSVHAGE 364
cons 379 * 441

ANEXO 9 (neuC) - score : 73




Escherichia_col 291 SSGLIEVPSLKVATINI---GDROKGRVRGASVIDVPVEKNAIVRGINISODEKFI-SVV-OS 348
Spirochaeta_cal 295 SSGIIEAPSLNVPTVNI---GKRODGRIKAKSVYDCGLFANDIYKTINNVKE-DIK-NNR-IN 351
Leptospira_inte 294 SSGIIEAPSFKIGTINI---GDROKGRLRAKSIIDCEPKKIEIIDAFKRLYSSDFO-KNL-ST 351
Leptospira_inte 294 SSGIIEAPSFKIGTINI---GDROKGRLRAKSIIDCEPKKIEIIDAFKRLYSSDFO-KNL-ST 351
Leptospira_inte 294 SSGIIEAPSFKIGTINI---GDROKGRLRAKSIIDCEPKKIEIIDAFKRLYSSDFO-KNL-ST 351
Leptospira_borg 282 GTITEEASLLDLPAVTIRNTHERPEGM-DFGTLIMSGFSKERILESVRIVVDOYRD-GIRHYR 342
Leptospira_borg 282 GTITEEASLLDLPAVTIRNTHERPEGM-DFGTLIMSGFSKERILESVRIVVDOYRD-GIRHYR 342
Treponema_pedis 291 SSGLTEVPSFOIATINI---GDROKGRIAASSVINCDSSRESIOAAFDKLNNGTFK-ETL-KL 348
Brachyspira_mur 281 GTITEESSLLKFPAITIRNAHERPEGN-DEGTLIMCGYKKDRVLKAVEVVIEOYKN-NII-PK 340
Sphaerochaeta_p 295 GGIOEEAPSLGKPVLVMRNTTERPEGI-KAGTLKLVGTSEOVIYDNFKLLLENDETYEAM-SK 355
Treponema_brenn 288 GTIPEEGSYFNFPAVSIRTSTERPEAM-EKGVFTIGSITKEOVLOAVDLAVSMHAN-GDD-GL 347
Turneriella_par 283 GGVOKEAYFFOKPCITLRDETEWVELV-DAGYNRLVGANSDAILAAHHAFTSO-----RP-VF 338
Spirochaeta_sp. 287 GTLSEESAMLNFAGVLIRTSTERPEVL-DKGTVVIGGITTKDISQALEMAVSMRDN-QEK-TE 346
cons 316 . * : : : . . : 378
Escherichia_col 349 SSNPYF-KENALINAVRIIKDFIKSKNKDYKDFYDIPECTTSYD 391
Spirochaeta_cal 352 YENPYY-KKDTTITIYDTIKKYL-NYNKFIKNFFDL----TLKY 389
Leptospira_inte 352 TVNPYG-EGGASEKIVRILEKID-IOGILKKKFFN------LOF 387
Leptospira_inte 352 TVNPYG-EGGASEKIVRILEKID-IOGILKKKFFN------LOF 387
Leptospira_inte 352 TVNPYG-EGGASEKIVRILEKID-IQGILKKKFFN------LOF 387
Leptospira_borg 343 VVDDYS-AGOVSKKIVFIVOSYTD---YINRTV-------WHKS 375
Leptospira_borg 343 VVDDYS-AGOVSKKIVFIVOSYTD---YINRTV-------WHKS 375
Treponema_pedis 349 TVNPYEGDCFOSKOIIKILKSVD-LADIIKKRFYDA----EFSI 387
Brachyspira_mur 341 TVEDYD-VDSLSMKVIRIILGYTE---YINENI-------WRKY 373
Sphaerochaeta_p 356 ASNPYG-DGFACKKIADIIV----------------------S 375
Treponema_brenn 348 SVPSYT-DENVSTKVVKIIOSYVG---IVNKMV-------WRKT 380
Turneriella_par 339 SGNLYG-TGSAGKFITDKLR---L---FLOENL--------GHG 367
Spirochaeta_sp. 347 MPDDYH-DTNVSVKVVKMIQSYAK---IVNETV-------W--G 377
cons 379 * : 422
Escherichia_col 174 GSLGAENALSLHL-PNKQELELKYGSL-LKRYFVVVFHPETLSTQSVNDQIDELLSAISFFKN 234
Spirochaeta_cal 176 GAIGIDNILKLKL-LDKNEIQNVISDKIKYPYVVITFHPETIDCQSAEDQIKEILKALDKF-N 236
Leptospira_inte 175 GGMGIDGIKKSNL-LKKEELESSLNIKFKKNNLLITFHPVTLEESTAKVQMIELLLSLETLSP 236
Leptospira_inte 175 GGMGIDGIKKSNL-LKKEELESSLNIKFKKNNLLITFHPVTLEESTAKVQMIELLLSLETLSP 236
Leptospira_inte 175 GGMGIDGIKKSNL-LKKEELESSLNIKFKKNNLLITFHPVTLEESTAKVQMIELLLSLETLSP 236
Leptospira_borg 167 GSHMDEVLKYYKVKIEKSDILEKLKLE-KNKFFIVSFHREENVD--TIENLKKLLESLNAICE 226
Leptospira_borg 167 GSHMDEVLKYYKVKIEKSDILEKLKLE-KNKFFIVSFHREENVD--TIENLKKLLESLNAICE 226
Treponema_pedis 173 GGAITEAKKRLRL-LSKQELEQQLKFKFWNRTILVTFHPLTLSY-TTKTEFENLLYILNQY-G 232
Brachyspira_mur 166 GSSMKEIFEIYKDKINNSNILDKLSLE-KNKYFVVSSHREEVVD--IESNLRSLIESLNALSE 225
Sphaerochaeta_p 183 GNTAIDALKTTVRS---GYTHPELDWASDSRLIMITAHRRENLG--E--PMHNMFRAIRRIIE 238
Treponema_brenn 174 GSPMAEVLNGCIDRISSSSILSELNLE-ANKYILLSAHREENID--MESNFLALMNAVNAMAE 233
Turneriella_par 170 GDVMLDAALFYGERIT-SDYLKTIGVE-PQGYVLATIHRAENTD--SPERLATILDALAEISK 228
Spirochaeta_sp. 173 GSPMREVLDSHMEGIQKSDVLDRLNLE-KNKYILVSAHREENID--NEEHFMSLMTAINDMAE 232
cons 190 * : .: * : :: : 252
Escherichia_col 235 TH---DFIFIGSNADTGSDIIQRKVKYF-CKEY---KFRYLISIRSEDYLAMIKYSCGLIGNS 290
Spirochaeta_cal 237 DM---FLIFTKANSDAGGRIINTLIEEF-VSRNPD-RSILFSSLGQLKYLSLIKHSLMVIGNS 294
Leptospira_inte 237 EI---GLIFTMPNADTDGRIIFELVKEF-TSSH-S-NAWYFTSLGQTRYLSCLQFVDAVVGNS 293
Leptospira_inte 237 EI---GLIFTMPNADTDGRIIFELVKEF-TSSH-S-NAWYFTSLGQTRYLSCLQFVDAVVGNS 293
Leptospira_inte 237 EI---GLIFTMPNADTDGRIIFELVKEF-TSSH-S-NAWYFTSLGQTRYLSCLQFVDAVVGNS 293
Leptospira_borg 227 EY-GLPVIVST--H-PR---TRKRLEDFKEVEMNP-LITFLKPFGFFDYVQLQMSAFCILSDS 281
Leptospira_borg 227 EY-GLPVIVST--H-PR---TRKRLEDFKEVEMNP-LITFLKPFGFFDYVQLQMSAFCILSDS 281
Treponema_pedis 233 DL---GIIFTKANSDTDGRVINQLIDEY-VFNNSH-RAVVFASLGHLRYLSTLQYVEAVVGNS 290
Brachyspira_mur 226 KY-DYKIIFST--H-PR---TRRLLDKLENIKVHS-NIIFSKPFGFFDYIALQKNSFCTLSDS 280
Sphaerochaeta_p 239 EHPDVKAIYPI--H-FNPI-VREAAK-Q-ELGGCE-RIRIIEPLDVLDFHNFLAQSYLIITDS 294
Treponema_brenn 234 NY-DMPVLYSC--H-PR---SRKMIEKR-GFKFDK-RVIQHQPLGFFDYNKLQQNAFCVVSDS 287
Turneriella_par 229 RI---PVILPL--H-PR---TRKLLEQRSWSNPRPQDLRIINPIGYLQMVCLEKNCHMIATDS 282
Spirochaeta_sp. 233 FY-QMPVIYST--H-PR---SKKFIEQR-SFQFHP-LVQSLKPFGFLDYNKLQMNAYCVLSDS 286
cons 253 : . .: :* 315

REFERENCIAS

Almagro-Moreno, S., Boyd, E.F. (2009) Insigths into the evolution of sialic acid catabolism among bacteria, BMC Evolutionary Biology 9:118

Blix, G. (1936) Uber die kohlenhydratgruppen des submaxillaris mucins. Z.Physiol. Chem. 240: 43-54.

Blix, Svennerholus & Werner (1952) Acta Chemica Scandinava 6(3):358-362

Blood, D. C., Hernderson, J. A., Radostits, (1982) O. M. Medicina veterinaria. Editorial Interamericana, 5ta. Edición, 594 – 605, 1982.

Boffil, P., Rivas, A., Ramirez, W., Montañez, J., Martinez, A., Quincoses, T., González, L. R., Fustes, E. (1989) Manual de Enfermedades Infecciosas. Tomo I. Ed. I.S.A.C.H., M.E.S.

Buchrieser, C., Rusniok, C., Kunst, F., Cossart, P. and Glaser, P., (2003) Comparison of the genome sequences of Listeria monocytogenes and Listeria innocua: clues for evolution and pathogenicity. FEMS Immunol. Med. Microbiol., 35:207–13.

Bulach, DM., Zuerner, RL., Wilson, P., Seemann, T., Mcgrath, A., Cullen, P.A., Davis, J., Johnson,M., Kuczek E., Alt, D.P., Peterson-Burch, B., Coppel, R.L., Rood, J.I., Davies, J.K., Adler, B. (2006) Genome reduction in Leptospira borgpetersenii reflects limited transmission potential. Proc. Natl. Acad. Sci. U.S.A. 103 (39): pp. 14560–5.

Cespedes, Z.M. (2005) Leptospirosis: Enfermedad Zoonótica Emergente. Rev. Perú. med. exp. salud pública. vol.22, no.4 p.290-307.

Chen, F., Mackey, A.J., Stoeckert, C.J. Jr., Roos, D.S. (2006) OrthoMCL-DB: querying a comprehensive multi-species collection of ortholog groups. Nucleic Acids Res. 1; 34(Database issue):D363-8.

Chen, F., Mackey A.J., Vermunt, J.K., Roos, D.S. (2007) Assessing performance of orthology detection strategies applied to eukaryotic genomes. 18; 2(4):e383.

Chu, D., Roobol, J., Blomfield, I.C., (2008) A theoretical interpretation of the transient sialic acid toxicity of a nanR mutant of *Escherichia coli*. J Mol Biol. 18;375(3):875-89.

Díaz, D.V. (2005) Asociación Venezolana de Producción Animal - Leptospirosis. (4) , pp 300-306

Dongen, S.van (2000) Performance criteria for graph clustering and markov cluster experiments. National Research Institute for Mathematics and Computer Science

Fernández, V. (2007) Estudios Moleculares sobre la neuraminidasa de *Mannheimia haemolytica* Secretariado de Publicaciones Universidad de León, León, España.

Ferrero, M.A., Martínez-Blanco, H., Lopez-Velasco, F.F., Ezquerro-Sáenz, C., Navasa, N., Lozano, S., Rodríguez-Aparicio, L.B. (2007) Purification and characterization of GlcNAc-6-P 2-epimerase from *Escherichia coli* K92. Acta Biochim Pol. 54(2):387-99.

Hakobyan, S., Harris, C.L., Tortajada, A., Goicochea de Jorge, E., García-Layana, A., Fernández-Robredo, P., Rodríguez de Córdoba, S., Morgan, B.P. (2008) "Measurement of factor H variants in plasma using variant-specific monoclonal antibodies: application to assessing risk of age-related macular degeneration." Invest Ophthalmol Vis Sci. 49(5):1983-90

Johnson, R.C., Harris, V.G. (1967) Differentiation of pathogenic and saprophytic *Leptospires* I. Growth at low temperatures. J. Bacteriol. 94 (1): pp. 27–31.

Kalivoda, K.A., Steenbergen, S.M., Vimr, E.R., Plumbridge, J. (2003) Regulation of sialic acid catabolism by the DNA binding protein NanR in *Escherichia coli*. J Bacteriol. 185(16):4806-15.

Kiefelt, M.J., Wilson, J.C., Bennett, S., Gredley, M., Von Itzstein, M. (2000) Synthesis and evaluation of C-9 modified N-acetylneuraminic acid derivatives as substrates for N-acetylneuraminic acid aldolase. Bioorg Med Chem. 8(3):657-64.

Klenk, E. (1941) Neuraminsäure, das Spaltprodukt eines neuen Gehirnlipoids. Zeitschr. f. Physiol. Chem. 268:50-58.

Koonin, E.V. (2005) Orthologs, paralogs, and evolutionary genomics. Annu Rev Genet, 39:309–38.

Lawrence, J.G. (2005) Horizontal and vertical gene transfer: the life history of pathogens. Contrib. Microbiol. 12:255–71.

Li, L., Stoeckert, CJ. Jr., Roos, DS. (2003) OrthoMCL: identification of ortholog groups for eukaryotic genomes. Genome Res. 13(9):2178-89.

Levett PN 2001. Leptospirosis. *Clin. Microbiol. Rev.* 14 (2): pp. 296–326.

Mathews, C. K., Van Holde, K.E et Ahern, K.G (2003). *Bioquímica* (3 edición)

Martinez, J., Steenbergen, S., Vimr, E. (1995) Derived structure of the putative sialic acid transporter from Escherichia coli predicts a novel sugar permease domain. JBacteriol. 177(20):6005-10.

Miller, B.G., Raines, R.T. (2004) Identifying latent enzyme activities: substrate ambiguity within modern bacterial sugar kinase Biochemistry. 1;43(21):6387-92.

Motro, Y., Dunn, D.S., La, T., Phillips, N.D., Hampson, D.J., Bellgard, M.I. (2008) Intestinal *spirochaetes* of the genus *Brachyspira* share a partially conserved 26 kilobase genomic region with *Enterococcus faecalis* and *Escherichia coli*. Microbiol Insights.1:3–11.

Mulligan, C., Leech, A.P., Kelly, D.J., Thomas, G.H. (2012) The membrane proteins SiaQ and SiaM form an essential stoichiometric complex in the sialic acid tripartite ATP-independent periplasmic (TRAP) transporter SiaPQM (VC1777-1779) from *Vibrio cholerae*.J Biol Chem. Jan 27;287(5):3598-608. doi: 10.1074/jbc.M111.281030.

Nascimento, A.L., Ko, A.I., Martins, E.A., Monteiro-Vitorello, C.B., Ho, P.L., Haake, D.A., Verjovski-Almeida, S., Hartskeerl, R.A., Marques, M.V., Oliveira, M.C. and other authors. (2004) Comparative genomics of two *Leptospira interrogans* serovars reveals novel insights into physiology and pathogenesis. J. Bacteriol., 186:2164–72.

Nunziato, P.W., Wright, L.F., Vann, W.F. Y Silver, R.P. (1995) Nucleotide sequence and genetic analysis of the neuD and neuB genes in region 2 of the polysialic acid gene cluster of *Escherichia coli* K1. J. Bacteriol. J. 177: 321-319.

Petersen, M., Fessner, W., Frosch, M., Y Luneberg, E. (2000) The siaA gene involved in capsule polisaccharide biosynthesis of *Neisseria meningitidis* B codes for N-acylglucosamine-6-phosphate 2-epimerase activity. FEMS Microbiol. Lett. 184: 161-164.

Picardeau, M., Lobry, J.R., Hinnebusch, B.J. (2000) Analyzing DNA strand compositional asymmetry to identify candidate replication origins of *Borrelia burgdorferi* linear and circular plasmids. Genome Res.,10, 1594–1604.

Picardeau, M., Brenot, A., Saint Girons, I. (2001) First evidence for gene replacement in *Leptospira spp.* Inactivation of *L. biflexa* flaB results in non-motile mutants deficient in endoflagella. Mol. Microbiol. 40 (1): pp. 189–99.

Picardeau, M., (2013) Diagnosis and epidemiology of leptospirosis. Med.Mal Infect, 43(1):1–9.

Plumbridge, J., Vimr, E. (1999) Convergent pathways for utilization of the amino sugars N-acetylglucosamine, N-acetylmannosamine, and N-acetylneuraminic acid by *Escherichia coli*. J Bacteriol. 181(1):47-54.

Pumarola, S.T., Jiménez de Anta Losada, M.T. (2002) Leptospirosis. Medicine. 8(69): 3688-92.

Reglero, A.G., Bravo, I., Fernández-Martínez, V. (2007) Ácidos siálicos: distribución, metabolismo y función biológica An. R. Acad. Nac. Farm., 73 (4): 833-871

Remm, M., Strom, C.E., Sonnhammer, E.L. (2001) Automatic clustering of orthologs and in-paralogs from pairwise species comparisons. J Mol Biol, 314(5):1041–52.

Ricaldi, J.N., Matthias, M.A., Vinetz, J.M., Lewis, A.L. (2012) Expression of sialic acids and other nonuosonic acids in *Leptospira.* BMC Microbiology, 12:1611.

Rocha, E.P. & Danchin, A. (2003) Essentiality, not expressiveness, drives gene-strand bias in bacteria. Nat. Genet., 34:377–8.

Roobol, D., Blomfield, I.C., (2008) A theoretical interpretation of the transient sialic acid toxicity of a nanR mutant of Escherichia coli. J Mol Biol. 18;375(3):875-89.

Schmidt, H., Hensel, M. (2004) Pathogenicity islands in bacterial patho- genesis. Clin. Microbiol. Rev., 17:14–56.

Sohanpal, B.K., EL-labany, S., Lahooti, M., Plumbridge, J.A., Blomfield, I.C. (2004) Integrated regulatory responses of fimB to N-acetylneuraminic (sialic) acid and GlcNAc in *Escherichia coli* K-12 Proc Natl Acad Sci U S A. 16;101(46):16322-7.

Stoughton, D.M., Zapata, G., Picone, R., Vann, W.F. (1999) Identification of Arg-12 in the active site of Escherichia coli K1 CMP-sialic acid synthetase. Biochem J. 15;343 Pt 2:397-402.

Tamames, J. (2001) Evolution of gene order conservation in prokaryotes. Genome Biol. 2, RESEARCH0020.

Tanner, M. E. (2005) The enzymes of sialic acid biosynthesis. Biorg. Chem. 33: 216-228.

Vann, W.F., Daines, D.A., Murkin, A.S., Tanner, M.E., Chaffin, D.O., Rubens, C.E., Vionnet, J., Silver, R.P. (2004) The NeuC protein of Escherichia coli K1 is a UDP N-acetylglucosamine 2-epimerase. J Bacteriol 186(3);706-12. PMID: 14729696

Vann, W.F., Tavarez, J.J., Crowley, J., Vimr, E., Silver, R.P. (1997) Purification and characterization of the *Escherichia coli* K1 neuB gene product N-acetylneuraminic acid synthetase. Glycobiology 7(5);697-701. PMID: 9254051

Vann, W.F., Silver, R.P., Abeijon, C., Chang, K., Aaronson, W., Sutton, A., Finn, C.W., Lindner, W., Kotsatos, M. (1987) Purification, properties, and genetic location of *Escherichia coli* cytidine 5'-monophosphate N-acetylneuraminic acid synthetase. J Biol Chem. 25;262(36):17556-62.

Verma, A., Hellwage, J., Artiushin, S., Zipfel, P.F., Kraiczy, P., Timoney, S.B. (2006) LfhA, a novel factor H-binding protein of *Leptospira interrogans*. Infect. Immun. 74 (5): pp. 2659–66.

Vimr, E.R., Troy, F.A. (1985) Identification of an inducible catabolic system for sialic acids (nan) in *Escherichia coli*.J Bacteriol. 164(2):845-53.

Wolf, Y.I., Rogozin, I.B., Kondrashov, A.S. and Koonin, E.V. (2001) Genome alignment, evolution of prokaryotic genome organization, and prediction of gene function using genomic context. Genome Res., 11:356–72.

Zapata, G., Crowley, J.M., Vann, Y.W.F. (1992) Sequence and expresion of the *Escherichia coli* NeuC gene product. J. Bacteriol. 174: 315-319.