



UNIVERSIDAD NACIONAL AUTÓNOMA
DE MÉXICO

Facultad de Medicina

**Licenciatura en Investigación
Biomédica Básica**

Evolución de la Triosafosfato Isomerasa

T E S I S

QUE PARA OBTENER EL TÍTULO DE:
**LICENCIADA EN INVESTIGACIÓN BIOMÉDICA
BASICA**

P R E S E N T A:
FÁTIMA ALEJANDRA PARDO ÁVILA

Director de tesis:
Dr. Daniel Alejandro Fernández Velasco

Junio 2011

Facultad de Medicina





Universidad Nacional
Autónoma de México



UNAM – Dirección General de Bibliotecas
Tesis Digitales
Restricciones de uso

DERECHOS RESERVADOS ©
PROHIBIDA SU REPRODUCCIÓN TOTAL O PARCIAL

Todo el material contenido en esta tesis esta protegido por la Ley Federal del Derecho de Autor (LFDA) de los Estados Unidos Mexicanos (México).

El uso de imágenes, fragmentos de videos, y demás material que sea objeto de protección de los derechos de autor, será exclusivamente para fines educativos e informativos y deberá citar la fuente donde la obtuvo mencionando el autor o autores. Cualquier uso distinto como el lucro, reproducción, edición o modificación, será perseguido y sancionado por el respectivo titular de los Derechos de Autor.

Este trabajo se realizó con el apoyo económico de:

CONACYT 102182
Beca PAPIIT IN206510

AGRADECIMIENTOS

A la Universidad Nacional Autónoma de México

Agradezco a todos los miembros actuales y pasados del Laboratorio de Físicoquímica e Ingeniería de Proteínas, que con su trabajo diario ayudaron a que esta tesis se llevara a cabo.

Agradezco a la Dra. Alejandra Vázquez Lobo y al Dr. León Patricio Martínez Castilla por su ayuda en el análisis filogenético de la TIM.

Se agradece la ayuda técnica de:

M. en C. Laura Ileana Álvarez Añorve y Dra. Isabel Velázquez

AGRADECIMIENTOS

A mis padres, a mi tía, mi hermana y Erik, por su apoyo durante todos estos años.

A Román, porque sin tu apoyo gran parte de este trabajo no hubiera sido posible.

A Diego y Carlos, por estar conmigo en buenos y malos momentos, es un honor ser amiga de personas como ustedes.

A mis compañeros de la LIBB, aprendí mucho de ustedes.

A mis pasados tutores, la Dra Georgina Garza Ramos y el Dr. Luis Mendoza Sierra. Este trabajo también es resultado de lo que ustedes me han enseñado.

Al Dr Alejandro Fernández, por estar siempre dispuesto a enseñar, porque juntos logramos hacer esta tesis y nos divertimos mucho en el proceso.

A Haven, por la terapia, las discusiones y las clases de Biología.

A Mariana Schulte-Sasse, por tu valiosa ayuda en el laboratorio, tu entusiasmo y especialmente por tu amistad.

A Renan, por ser un gran compañero de laboratorio, dispuesto siempre a ayudar.

A Olivia y Andrea I.

A Luis B, por las asesorías constantes, apoyo incondicional y ayuda, me siento muy contenta y agradecida de que seas parte de mi vida.

INDICE

Resumen	2
Introducción	3
<i>Plegamiento de Proteínas</i>	3
<i>La triosafosfato isomerasa</i>	6
<i>Reconstrucción de Proteínas Ancestrales</i>	13
Hipótesis	16
Objetivos	16
Metodología y resultados	16
<i>Esquema general de la reconstrucción de proteínas ancestrales</i>	16
<i>Trabajo Teórico</i>	17
<i>Obtención de secuencias</i>	17
<i>Alineamiento Múltiple de Secuencias</i>	17
<i>Obtención del mejor modelo de sustitución.</i>	19
<i>Obtención de árbol filogenético</i>	22
<i>Reconstrucción de Estados Ancestrales</i>	27
<i>Trabajo Experimental</i>	32
<i>Síntesis y clonación de genes</i>	32
<i>Expresión y Purificación de Tim's ancestrales</i>	33
<i>Expresión de TIM63, TIM55 y TIM52</i>	33
<i>Montaje de protocolo de purificación de TIM63</i>	33
<i>Actividad de Triosafosfato Isomerasa</i>	34
<i>Propiedades espectroscópicas de TIM63</i>	35
<i>Fluorescencia intrínseca</i>	35
<i>Dicroísmo Circular</i>	36
<i>Desnaturalización por temperatura</i>	37
Resultados Experimentales	38
<i>Expresión de TIM63, TIM55 y TIM52</i>	38
<i>Montaje de protocolo de purificación de TIM63</i>	39
<i>Propiedades espectroscópicas de TIM63</i>	39
<i>Fluorescencia intrínseca</i>	39
<i>Dicroísmo Circular</i>	40
<i>Desnaturalización por temperatura</i>	41
Discusión	42
Bibliografía	47

Resumen

La triosafosfato isomerasa (TIM) es una enzima glucolítica. Presenta una estructura de barril (β/α)₈ oligomérico obligado, es considerada un catalizador perfecto ya que la catálisis está limitada solamente por difusión. La estructura tridimensional de la TIM está muy conservada en los tres dominios de la vida, sin embargo las vías de plegamiento varían entre especies, desde procesos de dos estados hasta procesos con intermediarios monoméricos y/o diméricos. Para elucidar cómo apareció esta variedad de comportamientos, proponemos examinar la evolución de la TIM mediante la inferencia de proteínas ancestrales. La reconstrucción de proteínas ancestrales consiste en inferir, a partir de las secuencias de proteínas que se conocen actualmente, cuál era la secuencia de la proteína que tenían los organismos (ahora extintos) que dieron origen a los que existen actualmente. Utilizando métodos de máxima verosimilitud y el árbol filogenético propuesto por Cicarelli inferimos secuencias de las TIMs ancestrales de eucariontes, procariontes y arqueas. Para los eucariontes se obtuvo una alta probabilidad de reconstrucción (>0.8) para un alto porcentaje ($>70\%$) de aminoácidos. Los sitios con baja probabilidad se localizan en las hélices, en el exterior del barril; en el núcleo del barril y en el sitio catalítico se observó lo opuesto. Resultados similares se obtuvieron para los procariontes, mientras que para arqueas se obtuvieron pocos sitios con alta probabilidad de reconstrucción. Por lo anterior, elegimos tres secuencias de proteínas ancestrales de eucariontes a diferentes profundidades del árbol para hacer reconstrucciones experimentales. Se sintetizaron los genes que codifican estas proteínas en un vector de clonación y los subclonamos en un vector de expresión pET28b(+). Sobreexpresamos de manera soluble las proteínas ancestrales. Iniciamos la caracterización de una de las proteínas ancestrales, la TIM63, una proteína de hace 1.37 billones de años. Esta proteína fue activa, tiene estructura terciaria de barril (β/α)₈ y tiene una T_m aparente de 64.6 °C. A partir de estos datos y de una caracterización más detallada que se realizará posteriormente, queremos dar una explicación de los componentes fisicoquímicos que han moldeado la evolución de la TIM.

Introducción

Plegamiento de Proteínas

El plegamiento de las proteínas es el proceso por el cual una cadena polipeptídica sin estructura específica altera su conformación para alcanzar su estado nativo. Las proteínas son sintetizadas en los ribosomas como cadenas lineales típicamente de unos cientos de aminoácidos, en un orden específico, a partir de información codificada dentro del DNA celular. Para que funcionen, es necesario que estas cadenas se plieguen en una estructura tridimensional nativa única que es característica para cada proteína. El proceso de plegamiento puede ocurrir rápidamente y es un proceso específico. Un polipéptido recién formado debe ser capaz de encontrar su camino al plegamiento correcto rápidamente en vez del gran número de estructuras alternativas. Descubrir cómo ocurre esto es uno de los mayores retos en la biología estructural. (Pain 2000)

La pregunta principal es ¿Cómo encuentra una proteína el estado de menor energía en un tiempo razonable? El número de posibles conformaciones de una cadena polipeptídica es muy grande. Por ejemplo, para una cadena de 100 aminoácidos, si asumimos que cualquier residuo puede adoptar sólo uno de dos posibles conformeros, existirían 1.26×10^{30} conformaciones distintas. Si sólo 10^{-11} s fueran requeridos para convertir una conformación en otra (el mínimo tiempo compatible con los movimientos de átomos requeridos de acuerdo a las leyes físicas), una búsqueda sistemática de todas las posibles conformaciones para encontrar la que tiene la menor energía tomaría más de 10^{19} años. La mayoría de las proteínas pueden encontrar su estado nativo en un segundo. Esta aparente incompatibilidad se conoce como la paradoja de Levinthal. (Pain 2000)

Se ha descrito una correlación entre las constantes de plegamiento de proteínas, obtenidas al estudiar la velocidad de renaturalización de proteínas previamente desnaturalizadas en urea o guanidinio, y su orden de contacto relativo (OCR), definido como la suma de la distancia que existe entre pares de residuos que interactúan dividido entre el número de residuos. El OCR es una propiedad de la topología del estado nativo (Plaxco, Simons et al. 1998). Esta correlación predice que las proteínas que tienen un plegamiento que contiene relativamente muchos contactos entre regiones cercanas entre sí en la cadena

polipeptídica tenderán a plegarse más rápido que aquellas que tienen plegamientos más complejos, involucrando en promedio, interacciones entre regiones más distantes dentro de la secuencia (Pain 2000). La correlación entre el OCR y la velocidad de plegamiento llevó a la hipótesis de que la topología de las proteínas, no sólo su secuencia específica de aminoácidos, puede ser el factor más importante para determinar su vía de plegamiento. Para comprobar esta hipótesis, se han llevado a cabo estudios encaminados a dilucidar las vías de plegamiento de proteínas con estructuras similares pero con secuencias diversas. (Zarrine-Afsar, Larson et al. 2005).

También hay que recordar que las proteínas son parte de organismos que están sujetos a fuerzas evolutivas y por lo tanto, las proteínas son el resultado de un proceso evolutivo natural. Al estudiar una proteína es difícil separar las propiedades que son una consecuencia general de la fisicoquímica de la cadena polipeptídica de aquello que refleja el efecto de la selección natural. (Scalley-Kim and Baker 2004)

Para saber si existe un efecto de la selección natural sobre el plegamiento, el grupo de Baker estudió cinco proteínas que habían sido diseñadas *in silico* y que por lo tanto, nunca habían estado sometidas a la selección natural. Cuatro de éstas proteínas eran optimizaciones computacionales de plegamientos observados en la naturaleza, la quinta proteína era Top7 una proteína diseñada para tener un tipo de plegamiento no encontrado aún en la naturaleza. Encontraron que en general, las proteínas rediseñadas se pliegan y se despliegan más rápido que sus contrapartes naturales. El que las proteínas optimizadas presenten una energía de activación pequeña, a diferencia de las naturales, sugiere que la selección natural pudo haber actuado sobre las últimas con una presión de selección sobre alguna característica que llevara a una alta barrera de energía de desplegamiento, . (Scalley-Kim and Baker 2004)

Por lo tanto, en un estudio comparativo es importante tomar en cuenta cuál es la naturaleza de la similitud estructural de las proteínas que se comparan, es decir, si comparten un origen evolutivo común. Para estudiar el papel de la topología, independiente de los detalles de la secuencia, se puede optar por comparar proteínas que tengan una estructura terciaria muy similar, pero que no estén relacionadas evolutivamente. Las proteínas que son homólogas, aunque tengan baja identidad en la secuencia, podrían aún tener interacciones terciarias conservadas debido a la ancestría común. (Zarrine-Afsar, Larson et al. 2005) A continuación se presentan dos estudios comparativos de plegamiento.

El primer estudio se realizó con dos tipos de proteínas que pertenecen a la familia de las globinas que unen grupos hemo, las leghemoglobinas¹ y las mioglobinas. A pesar de tener una baja identidad en la secuencia presentan una estructura terciaria muy parecida. Se estudió la vía de plegamiento para estas proteínas y se observó que en ambos casos, las proteínas formaban intermediarios con hélices en etapas tempranas del plegamiento, sin embargo, los detalles de la vía fueron diferentes y parecen ser altamente dependientes de los detalles de la secuencia. (Nishimura, Prytulla et al. 2000)

El otro estudio se llevó a cabo con dos proteínas que presentan el tipo de plegamiento barril (β/α)₈, también conocido como barril TIM. Los barriles TIM, consisten en una hoja β cerrada elíptica formada por ocho hebras β paralelas, la cual está rodeada por ocho α hélices. Las hélices y las hebras están conectadas entre sí por medio de asas, de tal manera que se tiene 8 asas ($\beta\alpha$), que conectan el extremo carboxilo de las hebras β con las hélices α , y 8 asas ($\alpha\beta$) que conectan el extremo carboxilo de las hélices α con el extremo amino de las hebras β (figura 1). (Reardon 1995)

El barril TIM es una estructura ubicua que se encuentra en arqueas, procariontes y eucariontes. Este patrón de plegamiento está altamente representado, aproximadamente el 10% de las enzimas con estructuras cristalográficas reportadas lo presentan. Actualmente la base de datos SCOP, que clasifica a las proteínas en función de su estructura, distingue 26 superfamilias de barriles TIM. Los barriles TIM son funcionalmente versátiles ya que pueden catalizar cinco de los seis tipos de reacciones enzimáticas conocidas. (Murzin 2005)

A pesar de que este tipo de plegamiento está muy conservado, las proteínas que lo presentan no tienen alto grado de identidad en la secuencia. Sin embargo, alineamientos de secuencias basados en estructura revela la presencia de grupos de residuos fisicoquímicamente similares en posiciones estructuralmente similares. Por lo tanto, el plegamiento parece no estar determinado por los detalles de la secuencia sino por características generales como la distribución de los residuos polares y no polares. (Wallace 2002)

Las características comunes de los barriles TIM hacen pensar en un origen evolutivo común. Para este patrón de plegamiento, por ejemplo, el sitio activo está localizado en la cara C-terminal del barril- β central (la región formada por las asas $\beta\alpha$ figura 1), a pesar de no haber una preferencia estructural evidente en favor de la cara C-terminal sobre la N-

¹ Hemoglobinas de los nódulos de las raíces de plantas leguminosas fijadoras de nitrógeno

terminal. Además, aproximadamente dos tercios de las familias de enzimas barril (β/α)₈ utilizan sustratos o cofactores que contienen al menos un grupo fosfato; de hecho, el sitio de unión a fosfato, que consiste en las asas $\beta\alpha$ 7 y $\beta\alpha$ 8, es considerado el motivo estructural más conservado entre las enzimas barril (β/α)₈ (Vega 2003). Sin embargo, existen ciertas familias de barriles (β/α)₈ que no se pueden agrupar, lo que sugiere que pudo haber convergencia evolutiva (Nagano 2002).

Las vías de plegamiento de la indol-3-glicerol fosfato sintasa de *Sulfolobus solfataricus* y la subunidad alfa de la triptófano sintasa de *Escherichia coli*, han sido comparadas para probar la hipótesis de que los mecanismos de plegamiento están más conservados que las secuencias. Estas proteínas presentan una identidad de sólo el 13%, sin embargo, se encontró que existe un intermediario estable en la vía, parecido en propiedades termodinámicas y estructura secundaria para ambas proteínas. La comparación mostró que los pasos finales en la reacción de plegamiento están mejor conservados que los eventos tempranos. Los primeros pasos del plegamiento parecen ser más sensibles a las diferencias de las secuencias en este par de barriles TIM (Wallace 2002).

Los estudios presentados en esta sección muestran dos puntos de vista que se deben tomar en cuenta al estudiar el plegamiento de proteínas. Si bien las propiedades fisicoquímicas de los aminoácidos que conforman una proteína determinan sus propiedades, hay que recordar que las proteínas se encuentran en un medio sujeto a la selección natural. Las fuerzas evolutivas y las determinantes fisicoquímicas deben ser tomadas en cuenta si se quiere entender el problema del plegamiento de proteínas.

La triosafosfato isomerasa

Este trabajo se centra en la proteína que da nombre al tipo de plegamiento de barril (β/α)₈, la triosafosfato isomerasa, también llamada TIM. Las triosafosfato isomerasas de diferentes especies presentan un alto grado de similitud estructural, sin embargo el grado de identidad en la secuencia no es particularmente alto, lo cual la hace un modelo interesante para un estudio comparativo de plegamiento (figura 2).

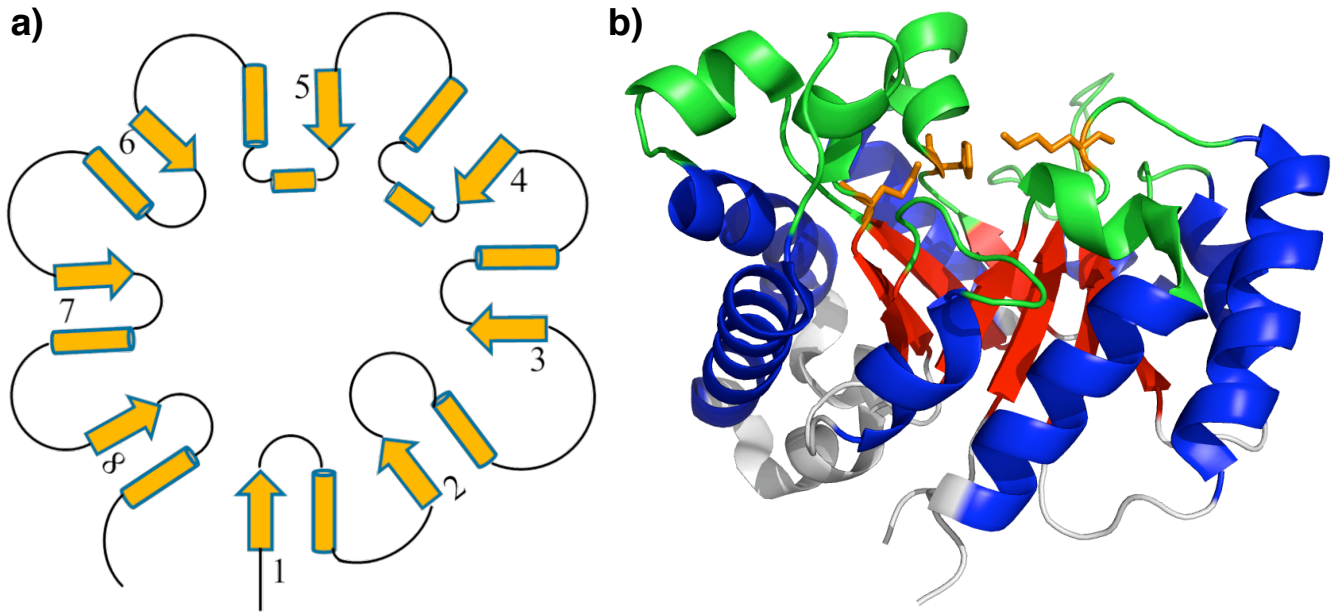


Figura 1. Estructura de un barril (β/α)₈. A) Esquema de la estructura secundaria de un barril (β/α)₈, como flechas se señalan las hebras β y como cilindros las hélices α . B) Estructura terciaria de un barril (β/α)₈, en rojo se muestran las hebras β , en azul las hélices α , en gris las asas $\alpha\beta$ y en verde las asas $\beta\alpha$. Entrada de PDB 1TRE

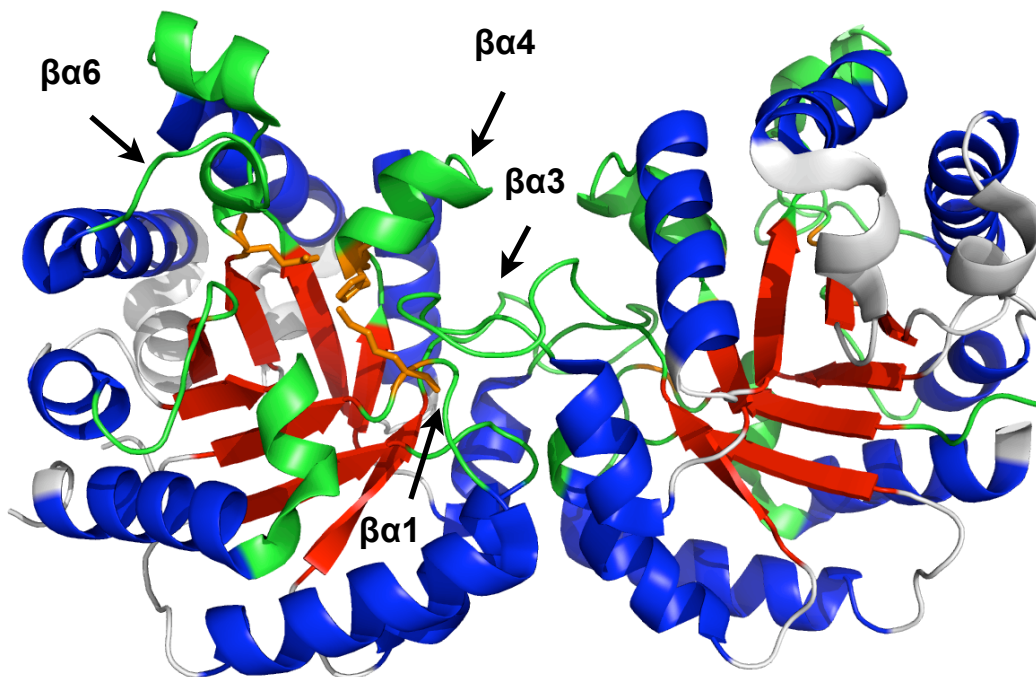


Figura 2. Triosafosfato isomerasa. Código de colores igual a la figura 1, en naranja se muestran los residuos catalíticos y se señalan las asas en las que se encuentran. También se señala el asa 3, componente principal en la dimerización.

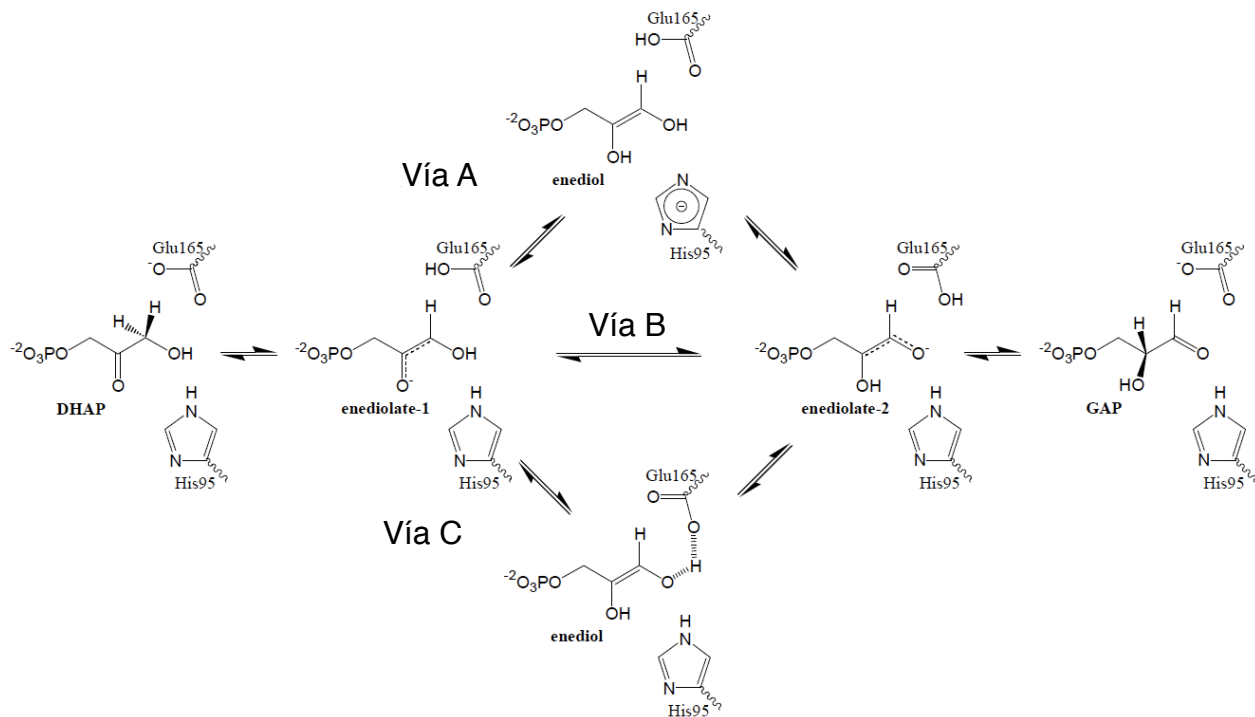


Figura 3. Mecanismos catalíticos propuestos para TIM (modificado de (Peimbert 2008)). Para todas las vías el Glu165 actúa como la base catalítica. La transferencia del protón del O1 al O2 se lleva a cabo por la His95 en la vía A, directamente sin enzima en la vía B y por el Glu165 en la vía C.

La TIM es una enzima ubicua, se ha encontrado en todos los organismos en los que se ha buscado, participa en la glucólisis, la gluconeogénesis, la síntesis de triglicéridos y el ciclo de Calvin, llevando a cabo la interconversión entre dihidroxiacetona fosfato y gliceraldehído-3-fosfato mediante la formación de un cis-enediol(ato) (figura 3). La isomerización ocurre sin cofactores y la enzima aumenta la reacción en una tasa de 10^9 en relación a la reacción no enzimática.

Los residuos que forman la triada catalítica son Lys12, His95 y Glu165 (según la numeración de la TIM de *Saccharomyces cerevisiae*) los cuales se encuentran localizados en las asas β 1, 4 y 6, respectivamente (Figura 2) (Peimbert 2008).

Estudios estructurales muestran que el carboxilato del Glu165 está posicionado precisamente para actuar como la base catalítica en la abstracción de protones de la dihidroxiacetona fosfato (DHAP) o del gliceraldehído-3-fosfato (G3P). La His95 parece tener varios papeles catalíticos. Estudios de resonancia magnética nuclear (NMR) indican que el imidazol neutro de His95 actúa como un electrófilo, polarizando el grupo carbonilo del sustrato. Después de que el protón es abstraído del sustrato, la “naciente” carga negativa en el carbonilo del sustrato es estabilizada por un puente de hidrógeno con la His95. Además,

se ha propuesto que la His95 puede actuar como ácido (la sustitución H95Q disminuye el paso de la enolización y fuerza un cambio en el mecanismo catalítico)(Peimbert 2008).

Se han propuesto tres vías por las cuales se puede llevar a cabo la reacción, en todas ellas la primer transferencia involucra la abstracción de un protón del C-1 de DHAP al Glu165 para formar un enediolato. El mecanismo clásico, propuesto por Albery y Knowles propone que un enediol se forma posteriormente por la transferencia de un protón de la His95 al oxígeno O2 del sustrato, este enediol es después transformado al enediolato por la transferencia de un protón del O1 del sustrato a la His95. En la vía “B”, propuesta por Alagona, Ghio y Kollman, el enediolato es formado por la transferencia directa de O1 A O2. Esta vía no ha recibido soporte por estudios experimentales. La vía “criss-cross”, originalmente propuesta para la mutante H95Q y tratada teóricamente por Perakyla y Pakkanen propone que el Glu165 lleva a cabo todas las transferencias de protones, i.e. de la posición C1 de DHAP al O2 del enediolato y después del O1 al C2 (Figura 3)(Peimbert 2008).

El asa $\beta\alpha$ 6 puede moverse aproximadamente 7 Å y de esta manera, encontrarse en dos conformaciones llamadas “abierta” y “cerrada”. Cuando se une un ligante, se favorece la conformación “cerrada”; en esta conformación el asa 6 cubre el sitio activo, lo que hace que el ligante sea inaccesible al solvente, además de llevar al Glu165 a la posición catalítica. En solución el intermediario enediol se convierte fácilmente a metilglioxal y fosfatos. Los aminoácidos del asa 6 estabilizan al enediol y lo mantienen deshidratado, favoreciendo la catálisis de la enzima. Como evidencia, en una mutante sin asa 6 que tiene como sustrato G3P, el intermediario enediol sale del sitio activo y se descompone en metilglioxal 5.5 veces más rápido que lo que tarda la reacción de isomerización (Pompliano, Peyman et al. 1990; Knowles 1991).

El factor limitante de la reacción catalizada por la TIM es el tiempo que tarda el sustrato en llegar al sitio activo de la enzima, es decir, la catálisis está limitada por difusión, por esta razón es considerada un catalizador perfecto.(Albery and Knowles 1976)

Las TIMs de diferentes especies presentan actividades catalíticas similares, sin embargo, se han encontrado grandes diferencias en la afinidad que tienen por el ácido 2-fosfoglicólico (PGA), análogo del estado de transición (Tabla 1).

Tabla 1. Parámetros cinéticos para la TIM de diferentes organismos.

Especie	PGA Ki mM	Km (G3P) (mM)	kcat (min⁻¹)
<i>Vibrio marinus</i> (Alvarez, Zeelen et al. 1998)	89	1.9 ± 0.2	4.2 ± 0.1 x 10 ⁵
<i>Escherichia coli</i> (Alvarez, Zeelen et al. 1998)	6	1.03 ± 0.1	5.4 ± 0.1 x 10 ⁵
<i>Giardia lamblia</i> (López-Velázquez, Molina-Ortiz et al. 2004)	0.043	0.53 ± 0.03	2.9 ± 0.2 x 10 ⁵
<i>Saccharomyces cerevisiae</i> (López-Velázquez, Molina-Ortiz et al. 2004)	0.03	1.27	2.6 x 10 ⁵
<i>Tripanosoma brucei</i> (López-Velázquez, Molina-Ortiz et al. 2004)	0.024	0.35	1 x 10 ⁶
<i>Homo sapiens</i> (Mainfroid, Terpstra et al. 1996)	0.0074	0.49	2.7 x 10 ⁵

PGA Ki (constante de inhibición para el PGA), Km (constante de Michaelis-Menten), kcat (constante catalítica, $kcat = V_{max}/[E]$).

A la fecha (mayo 2011) se han resuelto 109 estructuras cristalográficas de TIM, pertenecientes a 23 organismos diferentes. Se ha observado que en su forma silvestre la TIM es un oligómero (un dímero en organismos mesófilos y tetrámero en arqueas y algunos hipertermófilos) formado por monómeros plegados como barriles (β/α)₈ de aproximadamente 250 residuos. Aunque cada subunidad cuenta con un sitio activo, la enzima presenta actividad sólo cuando se encuentra como oligómero, por lo que se dice que es un oligómero obligado (Mainfroid, Terpstra et al. 1996). La interfase está formada principalmente por residuos de las asas $\beta\alpha$ 1, 2 y 3. El asa $\beta\alpha$ 3, generalmente la más larga de la TIM, contribuye mayoritariamente a la interacción entre subunidades (Lolis, Alber et al. 1990).

Se ha estudiado la vía de plegamiento para TIMs de diferentes organismos y se han encontrado comportamientos diversos. Las transiciones de desplegamiento han sido descritas por modelos que aumentan en complejidad, desde equilibrio hasta desplegamiento cinéticamente controlado; y desde reacciones de dos estados hasta reacciones multi-estado que involucran intermediarios monoméricos y/o diméricos (tabla 2). Los intermediarios más comunes son monómeros, que han sido observados cuando las proteínas son desnaturalizadas químicamente. La caracterización del desplegamiento de TIM de diversos organismos muestra que el paso de disociación contribuye con ~75% de la estabilidad conformacional del dímero. (Tellez, Blancas-Mejia et al. 2008)

Así mismo, estudios realizados a partir de monómeros de TIM obtenidos por derivatización química, el análisis de intermediarios en la vía de plegamiento y el estudio de mutaciones que impiden la dimerización, han mostrado que los monómeros son marginalmente estables y que aproximadamente el 70% del ΔG de plegamiento está dado por la asociación de los monómeros, como se puede observar en la figura 4, donde se compara la contribución de la asociación en el plegamiento de TIMs de diferentes especies (Peimbert 2008). Las razones de que la TIM sea un oligómero obligado no están totalmente entendidas.

Tabla 2. Vías de plegamiento reportadas para la TIM de diferentes especies.

Enzima	Vía de desnaturalización	Agente desnaturalizante	Referencia	Técnicas
BsTIM	$N_2 \Rightarrow 2D$	GdnHCl	(Mainfroid, Mande et al. 1996)	IF
ScTIM	$N_2 \Leftrightarrow 2D$	Temperatura	(Benítez-Cardoza, Rojo-Domínguez et al. 2001)	IF, DC, DSC
OcTIM	$N_2 \Leftrightarrow 2D$	GdnHCl	(Rietveld and Ferreira 1998)	IF
GgTIM	$N_2 \Rightarrow 2D$	Temperatura	(Shi 2008)	IF, DC, DSC
LmTIM	$N_2 \Leftrightarrow 2D$	GdnHCl	(Lambeir, Backmann et al. 2000)	IF, DC
HsTIM	$N_2 \Leftrightarrow 2D$	Urea	(Mainfroid, Terpstra et al. 1996)	IF, DC
ScTIM	$N_2 \Leftrightarrow 2M \Leftrightarrow 2D$	GdnHCl	(Nájera, Costas et al. 2003) (Morgan, Wilkins et al. 2000)	IF, DC
EhTIM	$N_2 \Leftrightarrow 2M \Leftrightarrow 2D$	GdnHCl	(Vázquez-Pérez and Fernández-Velasco 2007)	IF
TmTIM	$N_2 \Rightarrow A \Rightarrow 2D$	GdnHCl	(Beaucamp, Hofmann et al. 1997)	IF
PfTIM	$N_2 \Rightarrow N_2^* \Rightarrow A \Rightarrow 2D$	GdnHCl	(Gokhale, Ray et al. 1999)	IF, CD
TbTIM	$\begin{array}{c} A \quad A \\ \uparrow \quad \uparrow \\ N_2 \Leftrightarrow N_2^* \Leftrightarrow 2M \Leftrightarrow 2D \end{array}$	GdnHCl	(Chávez-Cárdenas, Fernández-Velasco et al. 2002)	IF, CD
TcTIM	$N_2 \Leftrightarrow N_2^* \Leftrightarrow 2M \Leftrightarrow 2D$	GdnHCl	(Chávez-Cárdenas, Pérez-Hernández et al. 2005)	IF, DC

N2 es el dímero nativo, N2 intermediario dimérico, M intermediario monomérico, D el estado desnaturalizado y A estados agregados. (Sc: Saccharomyces cerevisiae, Gg: Gallus gallus, Lm: Leishmania mexicana, Hs: Homo sapiens, Tb: Trypanosoma brucei, Tm: Thermotoga maritima, Pf: Plasmodium Falciparum, Tc: Trypanosoma cruzi, Sp: Schizosaccharomyces Pombe). IF: intensidad de fluorescencia, DSC: calorimetría diferencial de barrido, DC: dicroísmo circular.*

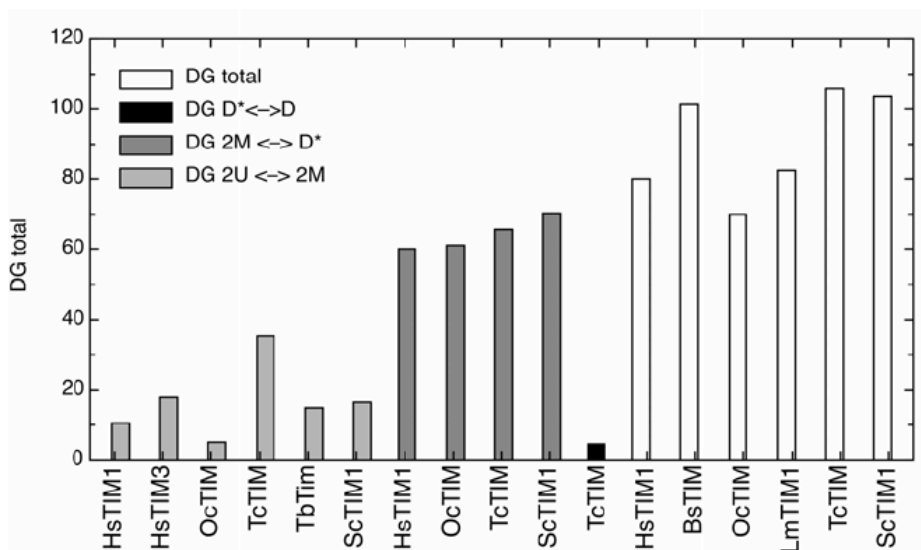


Figura 4. Comparación de energías libres determinadas experimentalmente para diferentes pasos de la vía de plegamiento. $DG = \Delta G$, $D^* \leftrightarrow D$ = intermediario dimérico a nativo, $2M \leftrightarrow D^*$ intermediario monomérico a dimérico (asociación), $2U \leftrightarrow 2M$ intermediario monomérico a desnaturalizado. Tomado de .

Hacer una comparación de los resultados que se observan en la tabla 2 es difícil, ya que no es posible determinar si las diferencias observadas en los mecanismos de plegamiento se deben a propiedades particulares de las TIM estudiadas o a las diferentes técnicas y condiciones experimentales empleadas. Un estudio comparativo de TIMs estudiadas bajo las mismas condiciones descartaría la variación debido a la metodología, y podría revelarnos patrones entre especies así como interacciones terciarias conservadas.

Para estudiar la contribución evolutiva en el plegamiento de la TIM proponemos estudiar no sólo las proteínas extantes (las que se encuentran en taxa con organismos que están vivos, lo contrario a organismos extintos), sino también los estados evolutivos intermedios, es decir, las proteínas ancestrales a partir de las cuales se generaron las que conocemos actualmente. Estudiar los ancestros nos permitirá ver cómo se fueron seleccionando determinadas regiones y el efecto que tenían sobre el plegamiento. Incluso poner a prueba si hay correlaciones en las regiones que estabilizan diferentes intermediarios en la vía.

Estudiar ancestros nos permitirá contestar preguntas relacionadas no sólo con el plegamiento, sino también preguntas inaccesibles por otros medios, como la antigüedad de la perfección catalítica o del estado oligomérico obligado. Un estudio bioquímico riguroso de proteínas ancestrales promete nuevas visiones de las determinantes fisicoquímicas que han moldeado la evolución de esta proteína.

Reconstrucción de Proteínas Ancestrales

La causa principal de la evolución son los cambios en los genes debido a las mutaciones. Un gen o secuencia de DNA con una mutación causada por sustituciones de nucleótidos, inserciones, deleciones o recombinaciones, puede esparcirse dentro de una población por deriva génica y/o selección natural y eventualmente puede ser fijada en una especie. Si este gen mutante produce un carácter morfológico o fisiológico nuevo, este carácter será heredado a todos los individuos descendientes a menos que el gen vuelva a mutar y se produzca otro carácter.

Si se establece un árbol filogenético para un grupo de especies, es posible rastrear el origen de los caracteres específicos que observamos, además de ayudar a entender el mecanismo evolutivo por el que se originaron (Nei 2000).

Conociendo la relación filogenética y comparando las secuencias de proteínas actuales (extantes) es posible inferir la secuencia de aminoácidos de proteínas ancestrales. Esto fue planteado por primera vez por Emile Zuckerkandl y Linus Pauling en 1963, a pesar de que en ese tiempo, llevar a cabo una reconstrucción no era metodológicamente posible. Este artículo reconoció la importancia de reunir los conocimientos obtenidos con la química estructural, la biología molecular y la evolución (Zuckerkandl and Pauling 1965).

Con los avances en secuenciación masiva, biología molecular, teoría evolutiva y el aumento en el poder de cómputo, se ha hecho posible el estudio de proteínas ancestrales en el laboratorio. Esto permite poner a prueba hipótesis acerca de las secuencias, afinidades o interacciones de las proteínas reconstruidas, así como del ambiente en que estas se encontraban; y de esta manera, entender los orígenes, evolución y distribución de la vida y las biomoléculas que la constituyen (Cole and Gaucher 2010).

La reconstrucción de proteínas ancestrales ha permitido responder preguntas a diferentes niveles. Desde el fenotipo de la proteína, como descubrir cuál era el color del ancestro de proteínas similares a la proteína verde fluorescente (GFP)(Ugalde, Chang et al. 2004), hasta hacer una aproximación de la temperatura de la tierra en el precámbrico por medio de la determinación de la temperatura media de desnaturalización (T_m) de factores de elongación (EF-TU) ancestrales (Gaucher, Govindarajan et al. 2008).

Comparar proteínas extantes ayuda a identificar los aminoácidos importantes para cumplir una determinada función, sin embargo, la reconstrucción de proteínas da información no

accesible de otra manera. Por ejemplo, este tipo de trabajos han ayudado a explicar por qué intentar cambiar la función de una proteína (o el sustrato que reconoce) haciendo mutaciones puntuales no siempre es suficiente (Ortlund, Bridgham et al. 2007), una reconstrucción de receptores de corticoides que reconocían diferentes ligantes permitió identificar los aminoácidos que permitían reconocer diferencialmente los ligantes y el orden en que se debían realizar mutaciones para poder pasar de una función a otra. Por otro lado, al hacer una reconstrucción se pueden rastrear los cambios necesarios para obtener la nueva función, incluso identificar aminoácidos implicados que no podrían ser reconocidos de otra manera, esto permitió en el caso de las proteínas similares a GFP reconocer qué aminoácidos eran los implicados en que la fluorescencia fuese roja o verde (Field and Matz 2010).

La tabla 3 muestra la variedad de trabajos realizados con esta aproximación. De particular interés es la reconstrucción de un ancestro de la enzima imidazol glicerol fosfato sintasa (HisF), un barril $(\beta/\alpha)_8$. HisF presenta una simetría de orden cuatro, lo que sugiere que el barril se pudo haber formado a partir de duplicaciones de fragmentos $(\beta\alpha)_2$. La inferencia de los estados ancestrales de HisF realizada por Richter y cols. apoya esta hipótesis, ya que mostró que los fragmentos $(\beta\alpha)_2$ del barril eran más parecidos entre sí en las proteínas ancestrales que en las proteínas HisF extantes. Se generó una proteína $(\beta\alpha)_4$, que corresponde a la mitad de un barril, a partir de la expresión de fragmentos $(\beta\alpha)_2$. La proteína fue estable, lo que demostró que era factible el origen de HisF por duplicaciones de los fragmentos $(\beta\alpha)_2$ (Richter, Bosnali et al. 2010).

Tabla 3. Ejemplos de proteínas ancestrales reconstruidas.

Genes extantes	Molécula reconstruida	Edad aproximada (millones de años)	Referencia
Proteínas verde fluorescentes	Proteínas fluorescentes ancestrales	20	(Ugalde, Chang et al. 2004)
Ribonucleasa digestivas	RNasa digestivas en los primeros rumiantes	40	(Jermann 1995)
Alcohol deshidrogenasas de levadura	Enzima al origen de la fermentación	80	(Thomson, Gaucher et al. 2005)
Receptores de glucocorticoides	Receptor corticoide ancestral	450	(Ortlund, Bridgham et al. 2007)
Galectinas: Congerina I y II	Ancestro común de congerinas I y II	?	(Konno, Kitagawa et al. 2011)
Factores de Elongación	Último ancestro común de eubacteria	3500	(Gaucher, Govindarajan et al. 2008)
Proteínas HisF	Fragmento ($\beta\alpha$)2	~ 4200 a 3500	(Richter, Bosnali et al. 2010)
Tiorredoxinas	Tiorredoxinas del último ancestro común de bacterias, de arqueas y de eucariontes	~ 4200 a 3500	(Perez-Jimenez, Fernandez et al. 2011)

En algunos estudios se reconstruyeron varias proteínas ancestrales, en la tabla se colocan las proteínas más ancestrales para cada estudio.

Hipótesis

Se puede inferir la secuencia de una TIM ancestral soluble, estable y activa, utilizando información evolutiva.

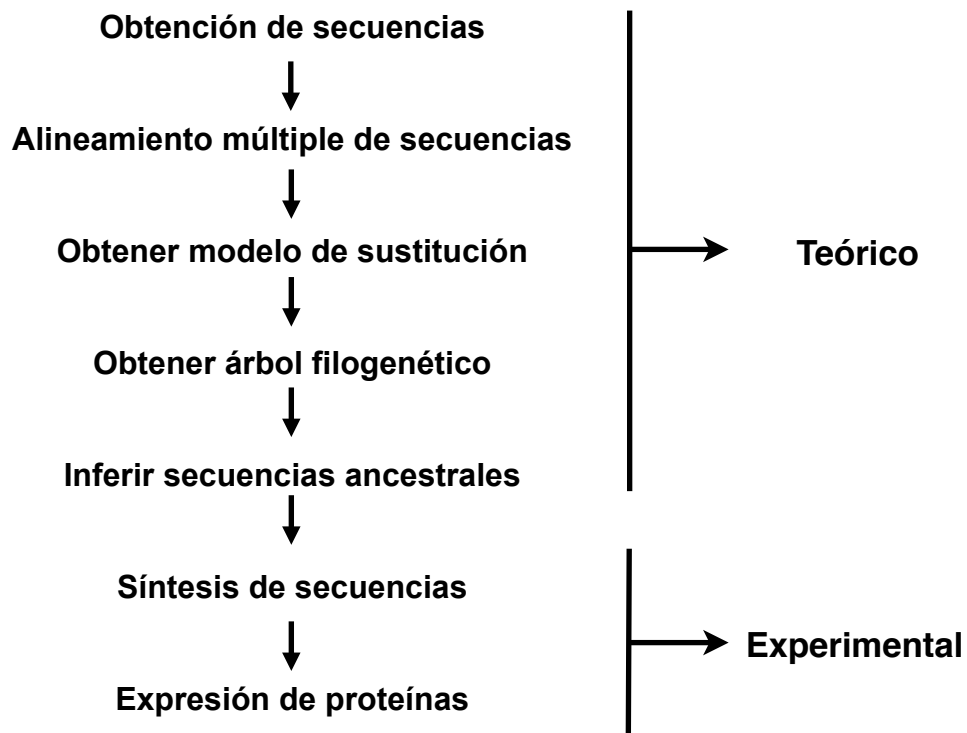
Objetivos

- Inferir las secuencias de TIM ancestrales.
- Expresar TIM ancestrales en el laboratorio.
- Iniciar la caracterización de una TIM ancestral.

Metodología y resultados

Esquema general de la reconstrucción de proteínas ancestrales

Los pasos necesarios para reconstruir una proteína ancestral se pueden observar en el esquema 1. Cada uno de los pasos se explicarán a continuación.



Esquema 1. Pasos a seguir en la reconstrucción de proteínas ancestrales

Debido a la naturaleza de los resultados, para la parte teórica se presenta la explicación de la metodología junto a los resultados que se obtuvieron. Para la parte experimental, se presenta primero la metodología y posteriormente los resultados.

Trabajo Teórico

Obtención de secuencias.

Las secuencias de nucleótidos que codifican TIM se obtuvieron de la base de datos del NCBI. Se obtuvieron 95 secuencias de procariontes, 44 de eucariontes y 17 de arqueas. El número de secuencias de TIM en el NCBI es mucho mayor, sin embargo sólo se seleccionaron las secuencias de organismos que habían sido parte de los estudios filogenéticos de Cicarelli y Lutzoni, por razones que se explicarán más adelante. Para esto, buscamos en el NCBI la secuencia anotada como codificante para la TIM de los organismos que se utilizaron en los estudios antes mencionados. El nombre de los organismos, la clave que se les asignó en este trabajo, así como el código de acceso de las secuencias se encuentran en el anexo 1.

Alineamiento Múltiple de Secuencias

El alineamiento de secuencias es el procedimiento por el cual se compararan dos o más secuencias de DNA o proteínas, buscando patrones que estén en el mismo orden en las secuencias. Caracteres idénticos o similares se localizan en la misma columna, mientras que los caracteres que no lo son se observan como espacios. En un alineamiento óptimo, los caracteres no idénticos alineados así como los espacios se colocan de tal manera que la mayor cantidad de caracteres idénticos o similares se encuentren alineados. Las secuencias obtenidas fueron traducidas a proteína utilizando el código genético universal utilizando el programa SE-AL. La secuencia de TIM de diferentes especies es en general de aproximadamente 250 aminoácidos, sin embargo las secuencias de arqueas tienden a ser menores (~200 aminoácidos). Al hacer un alineamiento preliminar con los organismos de los 3 dominios, no obtuvimos resultados satisfactorios ya que se observó un número grande de espacios, por lo que dividimos la secuencias en 3 subgrupos: arqueas, procariontes y eucariontes. Para cada dominio se realizó un alineamiento inicial utilizando el programa Clustal-X, posteriormente el alineamiento fue refinado manualmente utilizando el programa SE-AL, este programa permitió realizar el alineamiento para la secuencia de nucleótidos y de los aminoácidos que codifican. Los alineamientos obtenidos se muestran en el anexo 2.

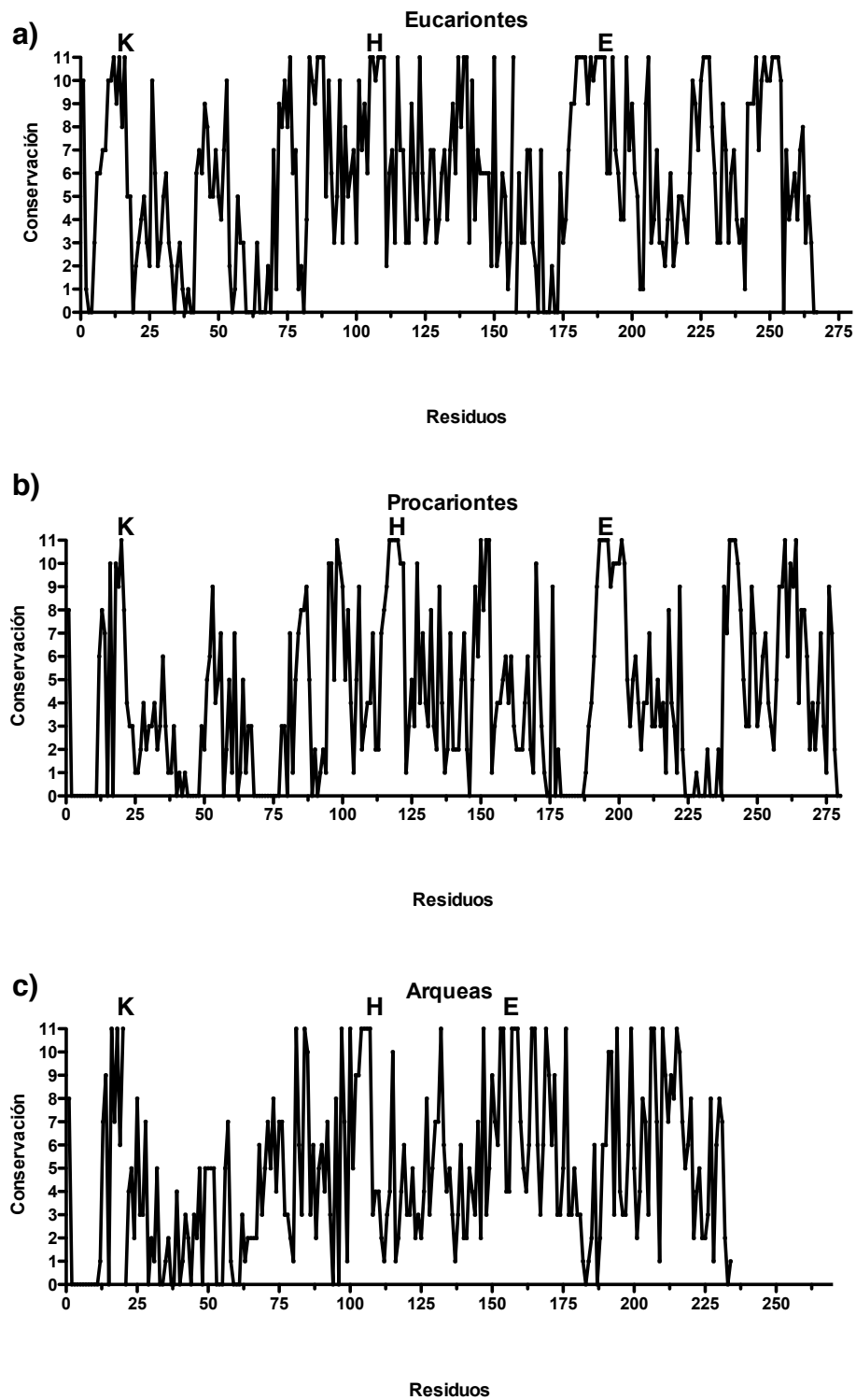


Figura 5 (bis). Conservación de residuos en alineamientos de TIM. A) eucariontes; b) procariontes; c) arqueas. Se grafica la conservación para cada una de las posiciones dentro de las secuencias de TIM de los diferentes dominios, el valor de conservación se calcula en base al alineamiento, con el programa AMAS de Jalview, los valores van de 0 (no conservado) a 11 (muy conservado). Se muestran las regiones donde se encuentran los residuos catalíticos K, H, E; se puede observar que los residuos adyacentes tienen un índice de conservación elevado.

Se calculó un índice de conservación para cada residuo de la proteína, a partir de los alineamientos y con el algoritmo AMAS de Jalview. (Waterhouse, Procter et al. 2009), este índice de conservación tiene una escala que va de 0 (no conservado) a 11 (muy conservado, se encuentra el mismo aminoácido en todas las secuencias).

En la figura 5 se grafica el índice de conservación para cada una de las posiciones dentro de las secuencias, se señalan las posiciones de los aminoácidos catalíticos. Se puede observar que para los tres dominios, la región con los menores índices es la que se encuentra entre los residuos 25 y 75. Esta región coincide con la zona que tiene más espacios dentro del alineamiento. Se observa también que las regiones cercanas a los aminoácidos catalíticos tienen mayores índices de conservación.

Obtención del mejor modelo de sustitución.

Al comparar dos secuencias de nucleótidos podemos preguntarnos cómo surgieron las diferencias entre ellas. Las secuencias que observamos actualmente son resultado de las sustituciones que sufrió la secuencia ancestral común.

Las sustituciones pueden clasificarse mediante diferentes criterios. Aquellas que involucran el cambio de purinas por purinas o pirimidinas por pirimidinas se llaman transiciones y las sustituciones que van de purina a pirimidina (o viceversa) se llaman transversiones (Page and Holmes 1998).

Las sustituciones en regiones codificantes también pueden clasificarse según el efecto que tienen sobre la proteína que codifican. Por la degeneración del código genético, hay sustituciones que no cambian el aminoácido codificado, este tipo se conoce como sustituciones sinónimas, mientras que las que sí lo cambian se llaman no-sinónimas (Page and Holmes 1998).

También se pueden obtener seis tipos de sustituciones al analizar la relación entre los tipos/números de cambios y el efecto que producen . En algunos casos, al comparar las secuencias veríamos una sola diferencia aunque realmente hubo dos cambios, o en otros es posible que no veamos diferencia aunque sí haya habido cambios. Por lo tanto, contar sólo el número de diferencias entre dos secuencias subestima la magnitud real de cambio evolutivo (Page and Holmes 1998)(figura 6).

Para lidiar con esto, se han desarrollado una gran variedad de métodos para medir la distancia evolutiva, todos con varias suposiciones sobre la naturaleza del proceso evolutivo.

Cuando se trabaja con aminoácidos, en vez de usar un modelo de sustitución, se utiliza una matriz de sustitución, que indica con qué frecuencia se observan cambios entre pares de aminoácidos. Hay una gran variedad de matrices, la decisión sobre qué matriz particular usar depende del tipo de reconstrucción que se quiera hacer y la naturaleza de las secuencias. (figura 8)

Para saber cuál es la matriz que se ajusta mejor a nuestros alineamientos de aminoácidos, usamos el programa Prottest. Este programa selecciona el modelo de evolución de proteínas que mejor se ajusta a un grupo de secuencias alineadas. Los modelos que evalúa son matrices de sustitución empíricas (como WAG, LG, mtREV, Dayhoff, DCMut, JTT, VT, Blosum62, CpREV, RtREV, MtMam, MtArt, HIVb, and HIVw) que indican las tasas relativas de cambio entre pares de aminoácidos y modelos que toman en cuenta otras características, como si hay sitios invariables (posiciones en el alineamiento que permanecen siempre iguales), si hay heterogeneidad en la tasa de sitios (que no todas las posiciones dentro de la secuencia tengan la misma tasa de cambio) o si hay algún efecto por la frecuencia observada de los aminoácidos dentro de las secuencias. Estas consideraciones toman en cuenta las limitaciones evolutivas impuestas por la conservación de la estructura y función de la proteína. Prottest usa el criterio de información de Akaike ($AIC = -2(\ln \text{verosimilitud} - \# \text{parámetros})$) y el criterio de información bayesiano ($BIC = G - g \ln N$, donde G es el cociente de verosimilitud, g los grados de libertad y N el tamaño de la muestra) para determinar cuál modelo ajusta mejor los datos, se usan este tipo de estadísticos ya que al buscar cuál es el mejor modelo, es importante comparar no solamente la verosimilitud que se obtiene, sino también tomar en cuenta cómo afecta a la verosimilitud el número de parámetros que se utilizan en el modelo .

Para los tres dominios obtuvimos que la mejor matriz de sustitución era la matriz LG, una matriz creada con el software XRATE a partir de 3912 alineamientos de la base de datos Pfam, consistiendo de ~50,000 secuencias y ~6.5 millones de residuos. Esta matriz, además, muestra una clara mejora en verosimilitud en comparación con las matrices WAG y JTT, cuando se hacen pruebas con otros alineamientos de proteínas. (Le and Gascuel 2008)

Por otro lado, se utilizó el programa ModelTest para saber cuál era el modelo que mejor se ajustaba a nuestro alineamiento de nucleótidos, se obtuvo que el mejor modelo (según el criterio de selección de Akaike) era el GTR+G (modelo general de tiempo reversible), donde se asume que cada una de las sustituciones(A<->T, A<->C, A<->G, T<->C, T<->G, G<->C) tiene una tasa de cambio diferente y que la probabilidad de sustitución no es la misma para todos los sitios.

Obtención de árbol filogenético

El análisis filogenético de un grupo de secuencias de ácidos nucleicos o de proteínas relacionadas consiste en determinar cómo los miembros del grupo pudieron haber divergido durante la evolución. Las relaciones evolutivas son representadas usando un grafo llamado árbol. Las secuencias extantes son localizadas en los extremos de las ramas, la manera en que se unen las ramas refleja el grado en que las diferentes secuencias se relacionan. Por ejemplo, dos secuencias que son similares se localizarán en ramas adyacentes y se unirán en un rama común debajo de ellas. Por otro lado, secuencias poco relacionadas estarán más alejadas en el árbol. El objetivo del análisis filogenético es descubrir el arreglo de las ramas, así como la longitud de estas, que mejor represente la relación entre las secuencias (Page and Holmes 1998).

Hay diferentes metodologías por las que se pueden obtener árboles: máxima verosimilitud, máxima parsimonia y métodos bayesianos. Intentamos hacer una reconstrucción por máxima verosimilitud y por métodos bayesianos.

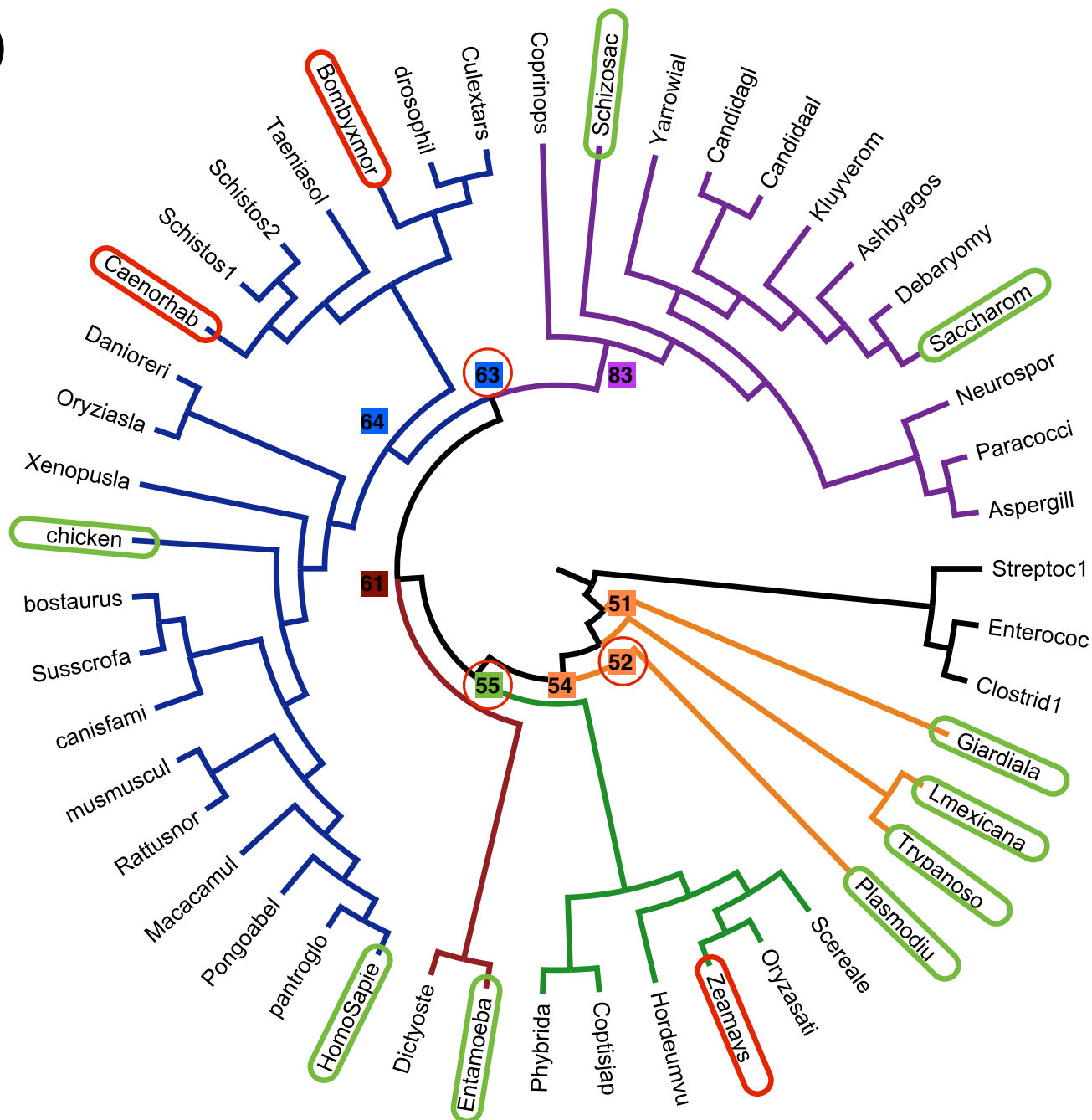
La reconstrucción por métodos bayesianos fue realizada con el programa MrBayes, siguiendo el modelo de sustitución de nucleótidos que habíamos elegido por los resultados del programa Modeltest, el modelo GTR 4x4. Obtuvimos un árbol, sin embargo no era convincente por no presentar al grupo de insectos como monofilético. Este problema también lo observó el grupo de Höhne (Knobloch, Schmidt et al. 2010), al hacer la inferencia del árbol filogenético para TIMs de insectos, también observaron que los insectos no aparecían como monofiléticos y notaron que esto se debía a que había una saturación en la tercera posición de los codones. Para tomar esto en cuenta, en vez de usar un modelo de sustitución de nucleótidos, hicimos una prueba preliminar con un modelo de sustitución de codones y obtuvimos resultados preliminares que mostraban una buena reconstrucción filogenética, aunque es necesario hacer un análisis más cuidadoso al respecto, que por dificultades metodológicas (como el tiempo que toma el proceso) no hemos podido llevar a cabo.

Hicimos una reconstrucción por máxima verosimilitud, utilizando el programa RaxML, sin embargo tampoco obtuvimos filogenias confiables, pues presentaban problemas similares a los de las filogenias obtenidas con MrBayes.

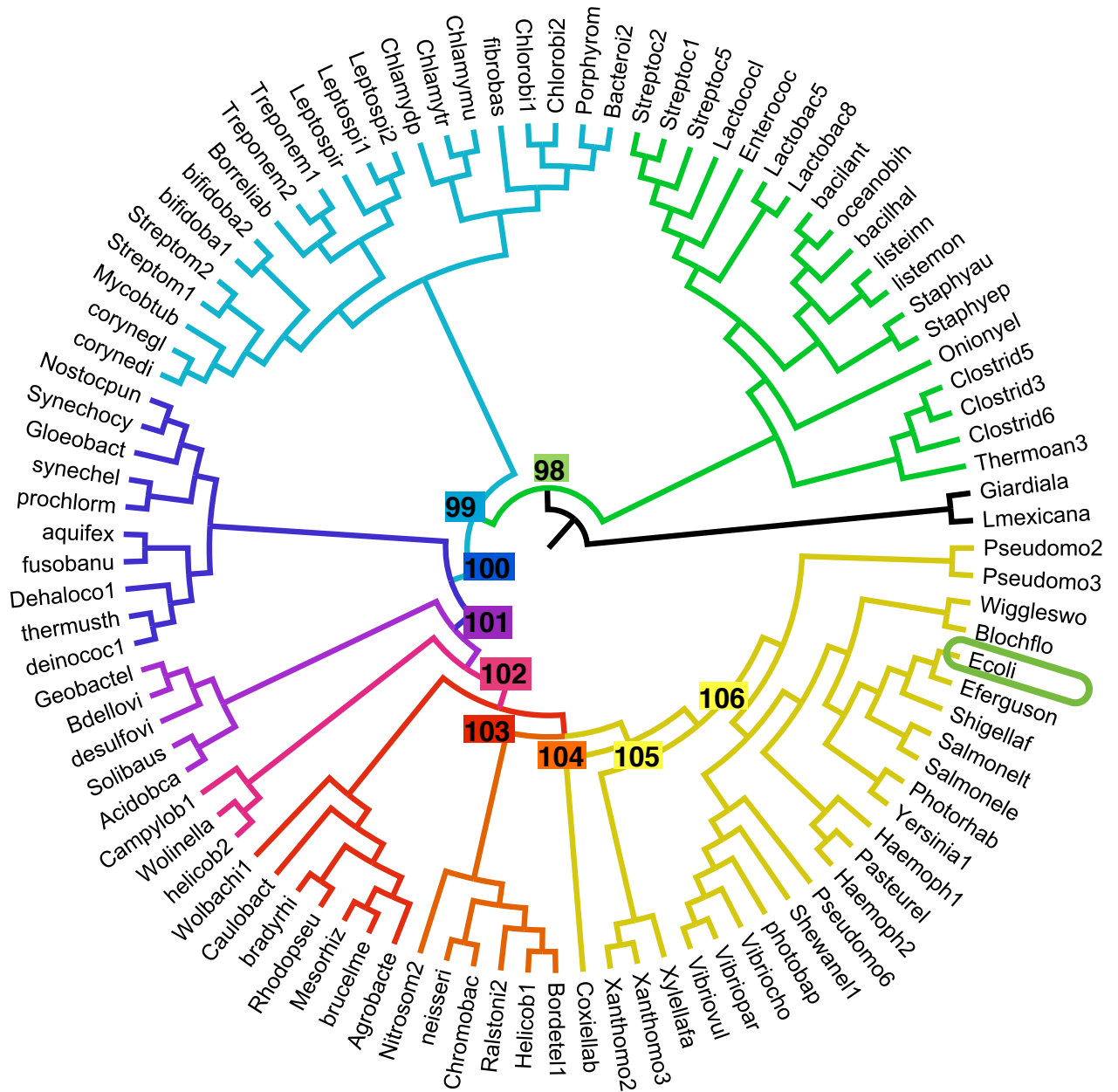
Los resultados anteriores sugieren que la TIM no es un buen marcador filogenético. Sin embargo, por ser parte del metabolismo central, supusimos que la historia evolutiva de la TIM era la misma que la historia de los organismos que la contienen. Decidimos trabajar con el árbol propuesto por Ciccarelli (Ciccarelli, Doerks et al. 2006), este árbol se reconstruyó a partir de 191 organismos y utilizando 36 genes que no presentan transferencia horizontal. En el caso de eucariontes, incluimos secuencias de hongos y utilizamos la filogenia reportada por Lutzoni, para añadir las secuencias al grupo de los eucariontes (Lutzoni, Kauff et al. 2004). En todos los casos, copiamos la topología de los árboles reportados y los acoplamos a los alineamientos que hicimos, con ayuda del programa MacClade. (Figura 8)

Figura 8. Árboles utilizados para la reconstrucción, adaptado de Cicarelli et al y Lutzoni et al. A) eucariontes, b) procariontes, c) arqueas. Los número señalan los nombres de los nodos que se analizaron. Los organismos rodeados de verde representan los organismos de los cuáles se ha estudiado la TIM, rodeados en rojo se encuentran los ancestros hipotéticos de los cuáles se reconstruyó experimentalmente la secuencia en este trabajo.

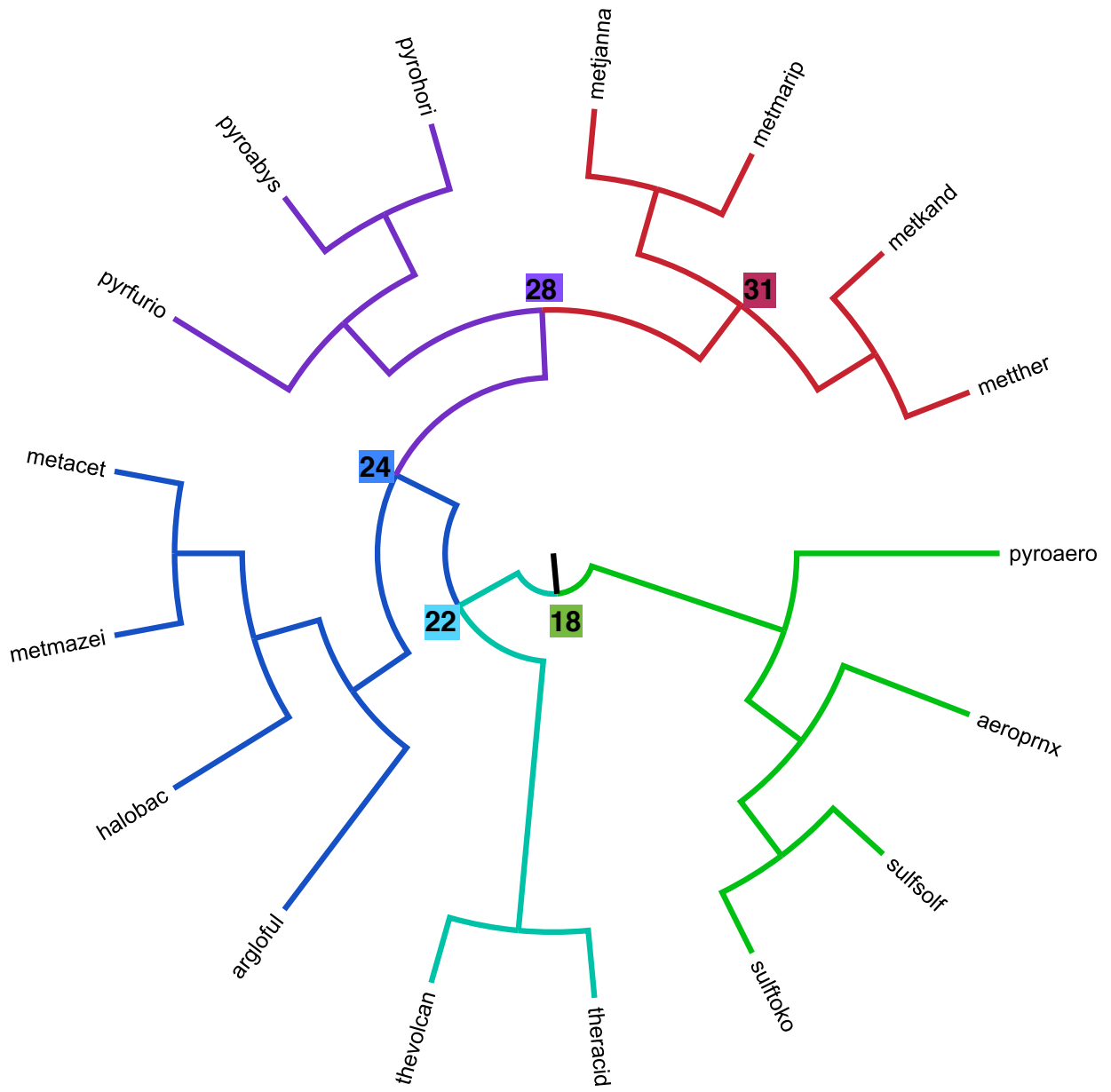
a)



b)



c)



Reconstrucción de Estados Ancestrales.

La reconstrucción de estados ancestrales puede realizarse también por diferentes metodologías: máxima parsimonia, máxima verosimilitud y métodos bayesianos.

El primer concepto para entender la teoría detrás de las reconstrucciones es el de una **secuencia ancestral probabilística**. Cada sitio en una secuencia en cada punto a lo largo de las ramas del árbol (en cada intermediario evolutivo) puede ser representada como una matriz de probabilidad de 20x1, donde cada elemento de la matriz es la probabilidad de que cada uno de los 20 aminoácidos se encuentre en el sitio en la secuencia ancestral representada por el punto. Estas probabilidades suman la unidad (Benner, Sassi et al. 2007).

El modelo de reconstrucción por máxima parsimonia o del mínimo cambio se centra en que ningún cambio por un aminoácido es más probable que otro cambio (todos los aminoácidos tienen la misma probabilidad de cambio) y que un cambio es más probable que dos cambios. Por lo tanto, si podemos suponer que la secuencia contemporánea surgió sin que la ancestral sufriera ningún cambio, entonces ningún cambio es inferido. Si para explicar la diferencia entre la secuencia ancestral y la extante es suficiente suponer un cambio, se infiere 1 cambio (y se desechan todas las soluciones que impliquen 2 cambios)(Benner, Sassi et al. 2007). Por seguir un modelo evolutivo tan simplificado, la máxima parsimonia tiene varias desventajas, como la imposibilidad de resolver posiciones ambiguas y la inhabilidad de proveer medidas de confiabilidad estadísticamente robustas.

Sin embargo, como mencionamos en el apartado anterior, tener en cuenta un modelo o matriz de sustitución puede acercarse más a lo que sucede en la naturaleza. Pensemos en dos secuencias de proteínas homólogas, y que las dos presenten (por ejemplo) una metionina en la posición 39. Aunque ambas secuencias tengan el mismo aminoácido en la misma posición, es posible que otro aminoácido (por ejemplo Isoleucina) haya ocupado la posición 39 en el último ancestro común. Si este hubiera sido el caso, al menos dos cambios independientes serían requeridos para que la metionina ocupara el lugar 39 en las dos proteínas extantes. La probabilidad de un cambio se supone que es pequeña en comparación con la probabilidad de ningún cambio, la verosimilitud de que este ancestro hubiera tenido una Ile en la posición 39 es considerablemente menor a la verosimilitud de que el ancestro hubiera tenido una Met en ese sitio, pero la verosimilitud no es cero.

Puede construirse una teoría que asigne una verosimilitud numérica basada en este modelo. Se pueden considerar datos empíricos o teoría para darle un peso a las probabilidades que varios otros aminoácidos hayan estado en la posición 39 durante la historia de la secuencia. Por ejemplo, Lys es (en varias situaciones) químicamente más diferente (comparada con Met) que Ile. Por lo tanto la historia podría dar una menor probabilidad a un cambio Lys-a-Met que un cambio Ile-a-Met. De esta manera, la probabilidad de que el sitio 39 haya sufrido 2 cambios Lys-a-Met es mucho menor que la probabilidad de que haya sufrido 2 cambios Ile-a-Met. Esto significa que la probabilidad de que la secuencia ancestral haya tenido una Lys es menor que la probabilidad de que haya tenido una Ile, dado que las 2 secuencias extantes tienen Met en el sitio 39. En este caso, el ancestro probabilístico podría tener valores diferentes a cero para todos los aminoácidos en la posición 39 (Benner, Sassi et al. 2007).

Este análisis es llamado de máxima verosimilitud, porque integra características, incluyendo la distancia del nodo a las hojas (secuencias extantes), en un modelo probabilístico. Por lo tanto, la probabilidad de que dos cambios ocurran a lo largo de una rama del árbol es menor si la línea es corta y más alta si la longitud de la rama es mayor. Esto surge directamente del hecho de que la longitud de la rama está determinada por el número de cambios por sitio (Benner, Sassi et al. 2007).

La proteína se reconstruye a partir de los aminoácidos que tienen una mayor probabilidad para cada sitio. No hay un consenso para decir cuándo un aminoácido pasa el límite a partir del cual se puede decir que es el único aminoácido probable, pero es claro, que si la probabilidad de que el sitio 39 tenga una Met es mayor a 0.81 mientras la probabilidad de cualquier otro aminoácido es menor a 0.01, se puede argumentar que sólo es necesario reconstruir la proteína que tiene Met en el sitio 39. (Benner, Sassi et al. 2007)

En inferencias bayesianas, la cantidad de interés es una distribución de probabilidad posterior, en vez de un punto o un mejor estimado, como en MP y ML, y donde un sumatoria explícita de las diferentes probabilidades para los parámetros no esenciales es llevada a cabo (Williams, Pollock et al. 2006). Este tipo de análisis se implementa con programas como MRBAYES, el cual hace la reconstrucción de estados ancestrales para ciertos nodos preseleccionados al mismo tiempo que hace la reconstrucción filogenética.

Hicimos reconstrucciones de las secuencias de eucariontes por el método de parsimonia, sin embargo para el caso del nodo más antiguo hubo 50 posiciones que no se pudieron resolver, mientras que para el nodo más cercano al presente hubo 2 posiciones sin resolver. Si pensamos que para cada posición puede haber como mínimo 2 aminoácidos posibles, eso nos daría que para el nodo más antiguo tendríamos 2^{50} posibles secuencias, lo que haría imposible estudiarlas en el laboratorio, por lo que decidimos explorar otros métodos de reconstrucción.

No pudimos llevar a cabo la reconstrucción por métodos bayesianos porque para ello es necesario hacer la reconstrucción del árbol filogenético a la vez que se reconstruyen los estados ancestrales.

Por estas razones, decidimos hacer la reconstrucción por máxima verosimilitud. Para lo cual se utilizó el programa PAML version 3.2 (Yang, Kumar et al. 1995). Se corrió el programa codeml, utilizando la matriz de sustitución LG y la opción de reconstrucción marginal. Para cada uno de los nodos del árbol se infirió la secuencia respectiva, aunque para el análisis mostrado a continuación sólo se utilizaron los nodos que se encuentran señalados en la figura 8, ya que estos nos permiten conocer las características de los ancestros a diferentes profundidades de los árboles.

Para cada una de las posiciones de las secuencias reconstruidas se obtuvo una probabilidad de reconstrucción. En la figura 9 se observa un histograma del número de posiciones con una determinada probabilidad. Observamos que en el caso de los eucariontes, los nodos más recientes tienen un mayor número de sitios que se reconstruyen con una alta probabilidad. Para los procariontes se observa que esto no se cumple, sino que los nodos con una mayor probabilidad son los de una profundidad intermedia. Para las arqueas se observan en general bajas probabilidades, posiblemente porque se usaron pocas secuencias, aunque un estudio más detallado se llevará a cabo en trabajos posteriores.

El alineamiento de las secuencias ancestrales probabilísticas de las TIM de eucariontes que se encuentran señaladas en la figura 8, comparadas con secuencias de TIM de eucariontes extantes se muestran en el anexo 3.

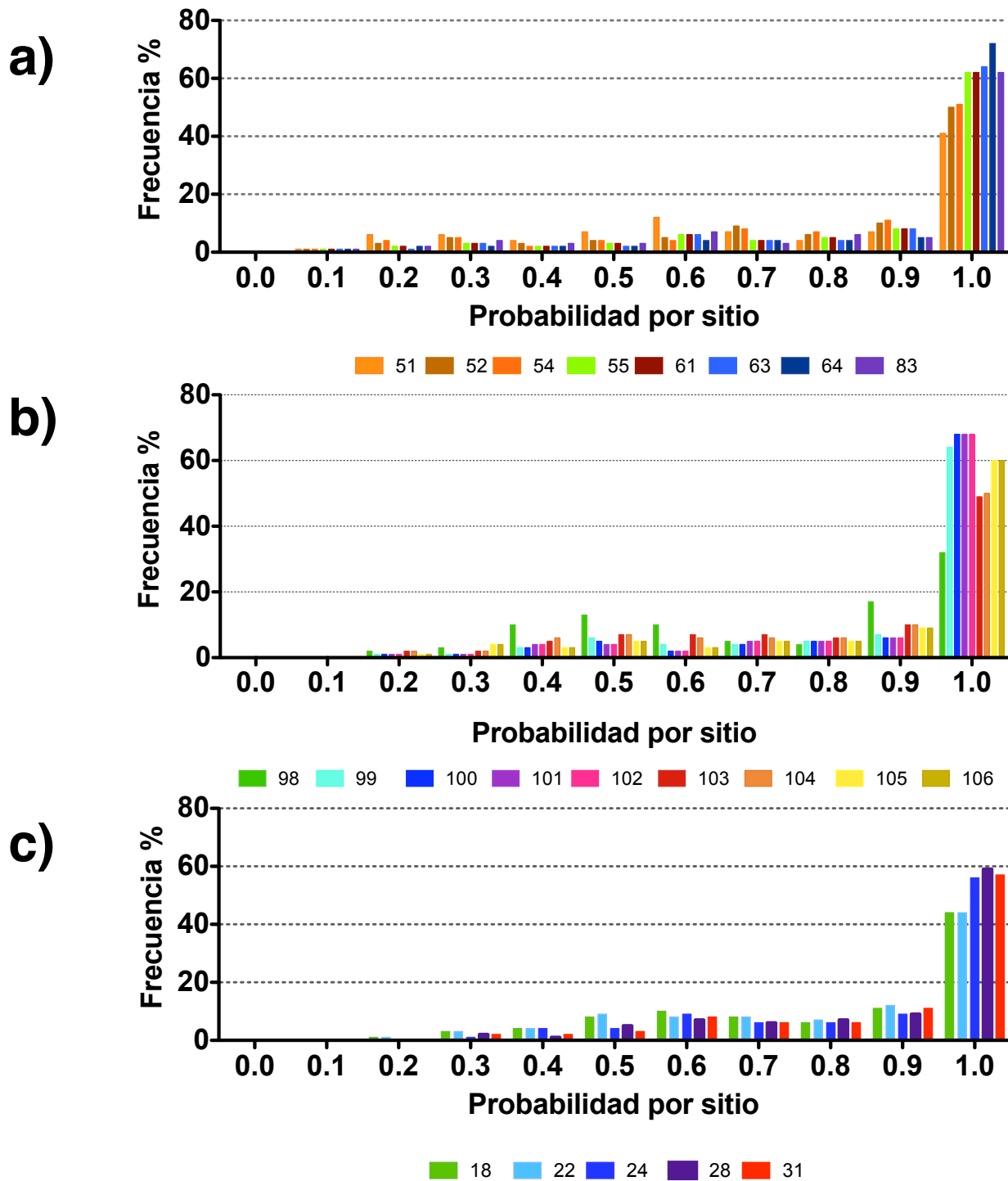


Figura 9. Distribución de las probabilidades por sitio. A) eucariontes, b) procariontes, c) arqueas. Los números representan los nombres de los nodos señalados en la figura 9 para cada uno de los árboles.

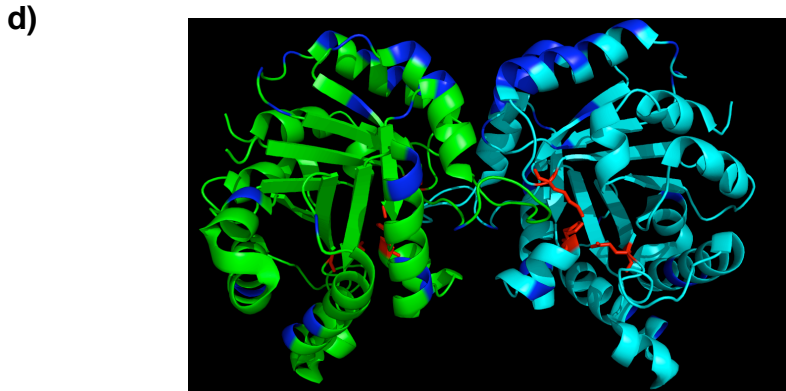
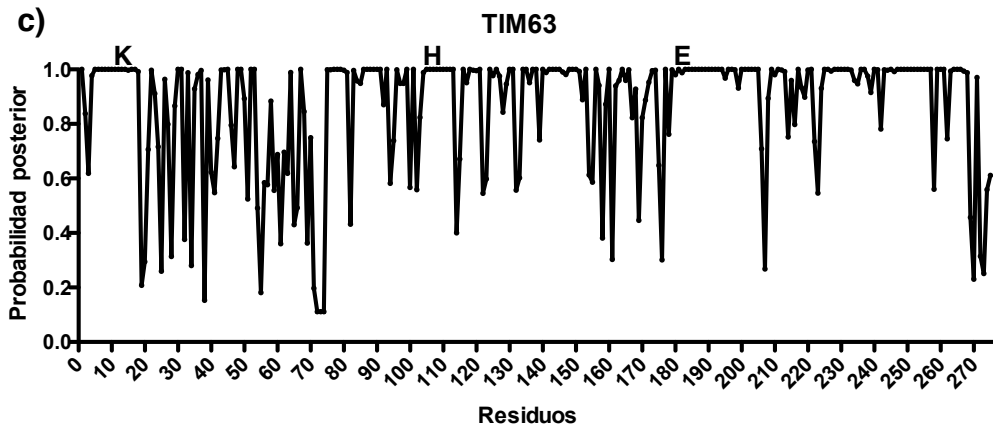
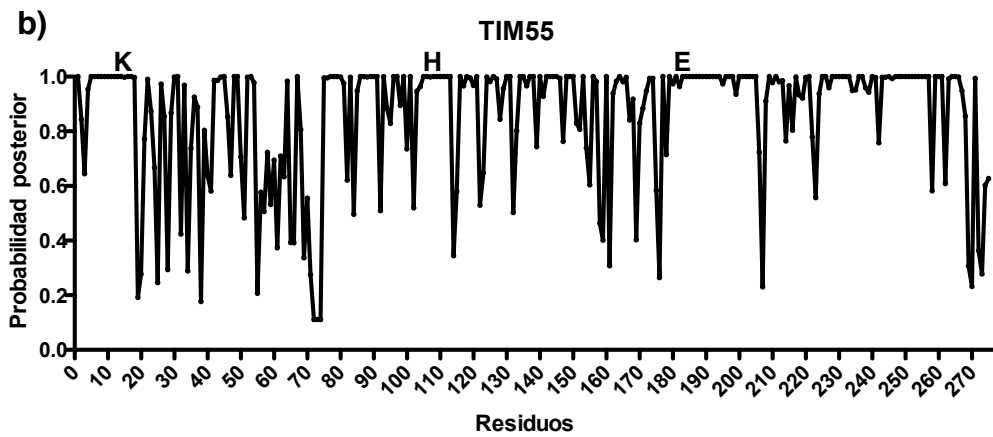
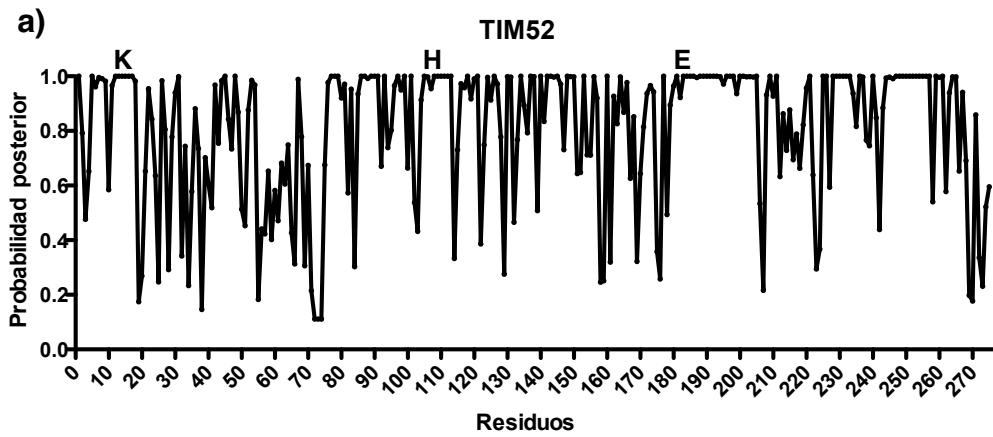


Figura 10. Probabilidad posterior por sitio de las secuencias reconstruidas. A) TIM52, B) TIM55, C) TIM63. Se puede observar que la región que tenía una baja conservación (entre los residuos 25 y 75) es la que se reconstruye con una menor probabilidad posterior. D) Los residuos que se reconstruyeron con una probabilidad posterior <0.8 para TIM63 se marcan en azul en una estructura representativa de TIM (PDB: 1YPI). Los residuos catalíticos están en rojo. Se muestra que la mayoría de estos residuos se encuentran en las hélices- α .

Trabajo Experimental

Síntesis y clonación de genes

Decidimos empezar por la expresión de las proteínas ancestrales de eucariontes ya que las TIMs ancestrales inferidas para este dominio fueron las que mostraron una mayor proporción de sitios reconstruidos con una alta probabilidad, y los sitios con baja probabilidad se encontraban en la parte externa del barril; además, la mayoría de las TIMs caracterizadas pertenecen al dominio de los eucariontes. Para tener una idea completa de la evolución de la proteína se eligieron tres proteínas ancestrales a diferentes profundidades del árbol (nodos 52, 55 y 63 que se señalan en la Figura 9). En la figura 11 se observa la probabilidad posterior asociada a cada sitio de la secuencia reconstruida para las tres TIMs ancestrales que escogimos, tal como se obtiene del programa PAML, en esta figura se pueden observar las regiones donde se localizan los sitios reconstruidos con una baja probabilidad.

Los genes se mandaron sintetizar con la compañía Epoch, las secuencias se flanquearon por los sitios de corte de NdeI y BamHI y la secuencia fue optimizada por la compañía para su expresión en *E. coli*. Los genes fueron clonados en el plásmido pBluescript II SK(-), un esquema de las construcciones se observa en el anexo 4.

Las secuencias fueron subclonadas en el vector de expresión pet28b(+). Este plásmido fue elegido porque coloca una etiqueta de Histidinas en el amino terminal de la proteína y tiene un sitio de corte por trombina, lo que permitirá quitar las colas de histidina.

Al digerir el plásmido pet28b(+) con la enzima BamHI se observó actividad estrella, es decir, la enzima cortaba inespecíficamente. Como se puede observar en el anexo 3, los genes de TIM63 y TIM52 resultaron también estar flanqueados por el sitio de corte que reconoce la enzima XhoI, del lado donde se encontraba el sitio de BamHI, por lo que decidimos usar XhoI. Para el caso de TIM55, tuvimos que usar los sitios NdeI y BamHI como se había planeado originalmente.

Para todas las condiciones, se digirió pet28b(+) durante 2.5 horas en una reacción de 50 ul (20 ul de plásmido purificado por columna de Qiagen, Buffer (4) 5 ul, BSA 10x 5 ul, NdeI 2 ul, XhoI ó BamHI 2 ul y agua 16 ul). Las secuencias que codifican para TIM se digirieron por 6 horas en una reacción de 50 ul (10 ul de plásmido purificado por lisis alcalina, Buffer (4) 5 ul, BSA 10x 5 ul, NdeI 3 ul, XhoI ó BamHI 3 ul y agua 24 ul). Los productos de las digestiones fueron cortados de gel y purificados por kit de Qiagen.

La reacción de ligación se llevó a cabo para TIM63 y TIM55 con una relación inserto:pet28 de 5:1, en un volumen de 10 ul, para TIM52 se usó una relación inserto:pet28 de 15:1 y las reacciones se incubaron por 2 días a 16°C. Se transformaron células quimiocompetentes DH5α con 10 ul de la reacción de ligación, plaqueamos y obtuvimos células resistentes. Purificamos DNA y mandamos secuenciar a la compañía Laragen para verificar las secuencias deseadas.

Expresión y Purificación de Tim's ancestrales

Expresión de TIM63, TIM55 y TIM52

A partir de una colonia aislada de cada una de las TIMs ancestrales se inocularon 10 ml de medio LB+Kanamicina y se incubaron toda la noche a 37°C, 250 rpm. Con los 10 ml se inocularon matraces con 100 ml de medio LB+Kanamicina y se crecieron a 37°C y 250 rpm hasta alcanzar una $DO_{600} \sim 0.6-0.8$, las células se indujeron con 1mM de IPTG y se incubaron durante 13 hrs a 37°C y 250 rpm. Las células se cosecharon por centrifugación a 5000 rpm por 20 minutos, las células fueron resuspendidas en amortiguador de fosfatos (NaH_2PO_4 35 mM, NaCl 300 mM, pH 6.3) con imidazol 5 mM y fueron lisadas por sonicación. Se centrifugaron durante 15 minutos a 15000 rpm y se recuperó el sobrenadante, el cual fue pasado por una columna de Níquel equilibrada previamente con buffer de fosfatos con imidazol 5 mM. La columna se lavó con amortiguador con imidazol 15 mM, y las proteínas se eluyeron con 25 ml de amortiguador con imidazol 75 mM, 15 ml de amortiguador con imidazol 130 mM y 15 ml de amortiguador con imidazol 275 mM y 15 ml de amortiguador con imidazol 500 mM.

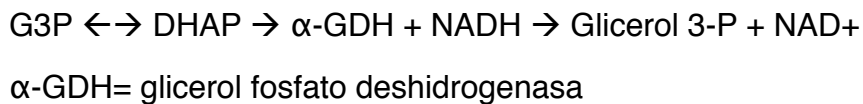
Montaje de protocolo de purificación de TIM63

A partir de una colonia aislada de TIM63 se inocularon 10 ml de medio LB+Kanamicina y se incubaron toda la noche a 37°C, 250 rpm. Con los 10 ml se inoculó un matraz con 250 ml de medio LB+Kanamicina y se crecieron a 37°C y 250 rpm hasta alcanzar una $DO_{600} \sim 0.6-0.8$, las células se indujeron con 0.8 mM de IPTG y se incubaron durante 13 hrs a 37°C y 250 rpm. Las células se cosecharon por centrifugación a 5000 rpm por 20 minutos, las células fueron resuspendidas en amortiguador de fosfatos (NaH_2PO_4 35 mM, NaCl 300 mM,

pH 6.3) con imidazol 5 mM y fueron lisadas por sonicación. Se centrifugaron durante 15 minutos a 15000 rpm y se recuperó el sobrenadante, el cual fue pasado por una columna de Níquel equilibrada previamente con buffer de fosfatos con imidazol 5 mM. La columna se lavó con amortiguador con imidazol 15 mM, y las proteínas se eluyeron con 40 ml de amortiguador con imidazol 75 mM, 15 ml de amortiguador con imidazol 130 mM y 15 ml de amortiguador con imidazol 275 mM y 15 ml de amortiguador con imidazol 500 mM. Las fracciones que eluyeron a 75, 130 y 275 mM de imidazol se concentraron y posteriormente se dializaron contra amortiguador de Trietanolamina 10 mM, EDTA 1mM, DTT 1 mM, pH 7.6. Los pasos de la purificación se siguieron midiendo concentración de proteína y actividad.

Actividad de Triosafosfato Isomerasa

La TIM cataliza la reacción de isomerización de gliceraldehído 3-fosfato (G3P) a dihidroxiacetona fosfato (DHAP). Se puede medir la cantidad de G3P que se convierte a DHAP mediante un ensayo acoplado a otra reacción enzimática que permite monitorear un cambio de absorbancia contra tiempo:



El NADH tiene una absorbancia específica de 340 nm. Cada molécula de DHAP que se forma a partir del G3P, es transformado por la $\alpha\text{-GDH}$ a glicerol 3-P y se libera un NAD^+ . Se mide la disminución de la absorbancia del NADH en el tiempo.

La reacción se llevó a cabo en un amortiguador de Trietanolamina 100 mM, EDTA 10 mM, pH 7.4. La reacción se midió en un volumen final de 1 ml de mezcla de reacción que contenía G3P 1 mM(final), 2 ul de NADH 1mM, 50 ul de $\alpha\text{-GDH}$ 0.14 mg/gl y 50 ng/ml de proteína (final).

Propiedades espectroscópicas de TIM63

Fluorescencia intrínseca

La absorción de luz por una molécula causa la excitación de un electrón, lo que ocasiona que el electrón pase de un estado basal a un estado excitado. Aunque el proceso de absorción es extremadamente rápido, tomando cerca de 10^{-15} s, la secuencia de eventos que regresan la molécula en el estado excitado al estado basal es considerablemente más lenta, tomando de 10^{-14} s a varios segundos. El regreso al estado basal por medio de la luminiscencia molecular es uno de los procesos más lentos en estados electrónicamente excitados. (Sharma 1999)

Al llegar al estado excitado, la molécula puede estar vibracionalmente excitada. La molécula entonces comenzará a vibrar con una frecuencia característica del estado excitado, liberando el exceso de su energía vibracional en la forma de radiación en el infrarrojo o en la forma de energía cinética impartida a otras moléculas con las que colisiona. La molécula desciende al estado vibracional más bajo del estado electrónicamente excitado, un proceso llamado relajación vibracional. Una vez que esto ha pasado, la molécula puede perder energía solamente al ir a un nivel de energía electrónica más bajo. Esto puede ocurrir de varias maneras. En el caso de que el estado vibracional más elevado de un estado electrónico menor sobrelape con el nivel vibracional más bajo de un estado electrónico mayor, entonces la molécula sigue perdiendo energía por medio de relajación vibracional, en un proceso que se conoce como conversión interna. (Sharma 1999)

Por otro lado, si la diferencia en energía entre los estados electrónicos mayores y menores es relativamente grande, puede ocurrir otro proceso que consiste en el descenso molecular del estado vibracional más bajo del estado electrónico mayor a cualquiera de los niveles vibracionales del estado electrónico menor. El exceso de energía, en este caso, es liberado como una onda de luz visible o ultravioleta cuya frecuencia depende de la diferencia en energía entre los estados vibracionales. Siguiendo la transición radiativa, la molécula sufre relajación vibracional hacia el estado vibracional más bajo del estado electrónico más bajo. La deactivación radiativa del estado electrónicamente excitado se conoce como fluorescencia. (Sharma 1999)

De los grupos químicos más comunes encontrados en macromoléculas biológicas, sólo los triptofanos, tirosinas, NADH y FAD fluorescen significativamente para ser usados en espectroscopía de fluorescencia. Todas estas moléculas consisten de anillos aromáticos planos. (Guzmán-Luna 2006)

Los espectros de fluorescencia son extremadamente sensibles a la polaridad del solvente y a las propiedades relacionadas con la facilidad para hacer puentes de hidrógeno. Por lo que los espectros de fluorescencia de una molécula nos permiten conocer el ambiente de grupos individuales en las estructuras macromoleculares. (Sharma 1999) Y en el caso de las proteínas, nos permite obtener información acerca de su estructura terciaria. Por ejemplo, un triptófano que está expuesto al agua tiene un máximo de fluorescencia a una longitud de onda de 350 nm, mientras que un triptófano totalmente oculto emite a cerca de 330 nm (Creighton 1997)

La fluorescencia intrínseca de las proteínas es usada para seguir la exposición de los residuos aromáticos que ocurre por desnaturalización o disociación generada por agentes químicos, temperatura, presión, etc. (Guzmán-Luna 2006)

La TIM63 tiene 5 triptófanos, que se localizan en las posiciones 11, 91, 159, 170, 193 y 5 tirosinas que se encuentran en las posiciones 46, 68, 116, 166 y 210.

Las mediciones de fluorescencia fueron realizadas en un espectrofluorómetro ISS PCI excitando a longitudes de onda de 280 o 295 nm y midiendo la emisión en el rango de 310 a 400 nm, la concentración de proteína fue de 50 ug/ml. A 280 nm esperamos ver la contribución de todos los residuos aromáticos de la proteína, mientras que a 295 esperamos ver sólo la contribución de los triptófanos.

Dicroísmo Circular

La base del dicroísmo circular es que la luz polarizada circularmente a la izquierda y a la derecha es absorbida de manera diferente por moléculas quirales. Los L-aminoácidos en polipéptidos y proteínas interactúan diferencialmente con rayos de luz polarizada a la izquierda o a la derecha, lo que causa que estos dos rayos viajen a diferentes velocidades a través de estas moléculas, rotando la luz polarizada.

Los cromóforos proteicos pueden dividirse en tres clases: el enlace peptídico, aminoácidos con cadenas laterales aromáticas y grupos prostéticos. En el UV-lejano (debajo de 250 nm) las características espectrales están determinadas primordialmente por la conformación de la

cadena principal (esqueleto polipeptídico), especialmente su estructura secundaria. Las hélices- α exhiben dos mínimos característicos a 208 y 222 nm, un máximo menos intenso centrado en 215 nm es típico de las hebras- β . Debido a que las señales de los cromóforos individuales es aditiva, los espectros de DC en la región de 190-240 nm representan una combinación de los espectros de los distintos elementos de estructura secundaria y por tanto arrojan información importante respecto al contenido total de estructura secundaria de una proteína (Creighton 1997; Guzmán-Luna 2006).

El DC es una técnica particularmente útil para monitorear cambios conformacionales en las proteínas, por ejemplo, en el desplegamiento térmico o en el inducido por un desnaturizante.

La medición de DC se realizó en un equipo Chirascan. A una concentración de 0.250 mg/ml, en amortiguador Trietanolamina 10 mM, pH 7.6, y se hicieron dos barridos.

Desnaturalización por temperatura

La desnaturalización por temperatura de la TIM63 fue determinada siguiendo la disminución en la señal de DC a 222 nm. Se elevó la temperatura en una tasa de 1°C /min, de 20°C a 90°C. Los experimentos se realizaron a una concentración de proteína de 0.250 mg/mL, en amortiguador Trietanolamina 10 mM (pH 7.6).

Para ver la transición de desplegamiento térmico, se graficó el cambio de elipticidad a 222 nm (Figura 15). El trazo fue ajustado a una sigmoide con pendiente para obtener la temperatura media de desnaturalización aparente ($T_{m_{app}}$). El cambio de señal también se monitoreó al enfriar de 90 a 20 °C.

Resultados Experimentales

Expresión de TIM63, TIM55 y TIM52

Las tres proteínas se encontraron de manera soluble y fue posible purificarlas usando la columna de níquel (Figura 12).

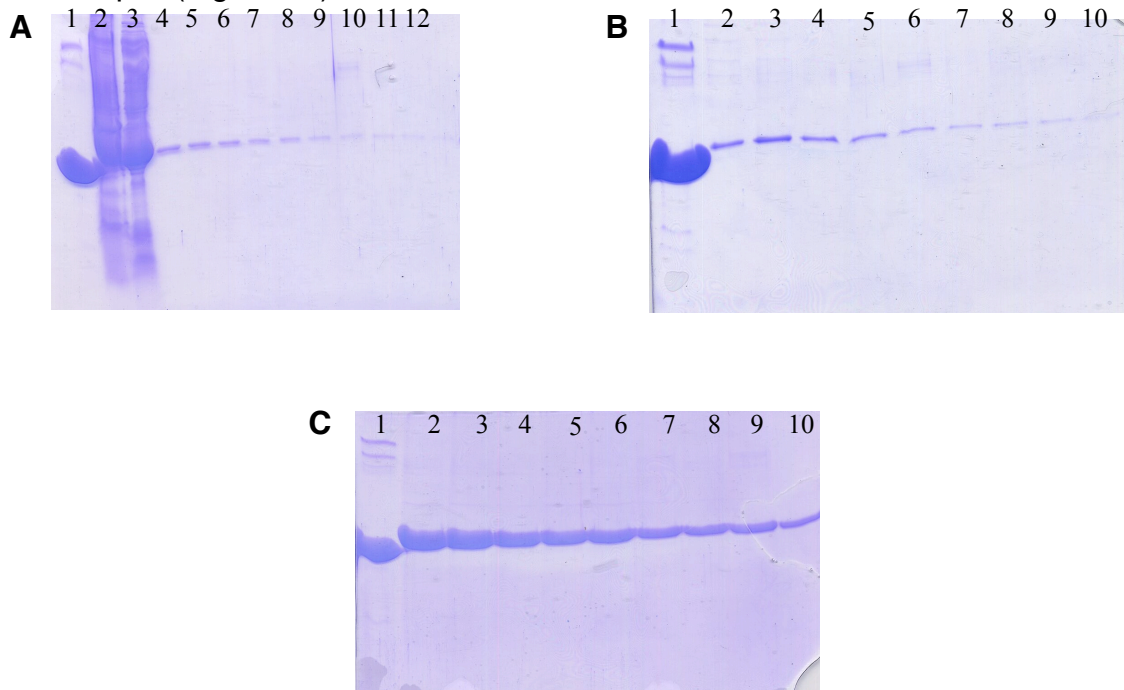


Figura 11. TIM52, TIM55 y TIM63 purificadas por columna de afinidad. A) TIM52: carril 1: scTIM, carril 2: extracto total, carril 3: sobrenadante, carriles 4-12: fracciones que eluyen a concentraciones crecientes de imidazol B)TIM55 C)TIM63, para B y C: carril 1: ScTIM, carril 2-10: fracciones que eluyen a concentraciones crecientes de imidazol.

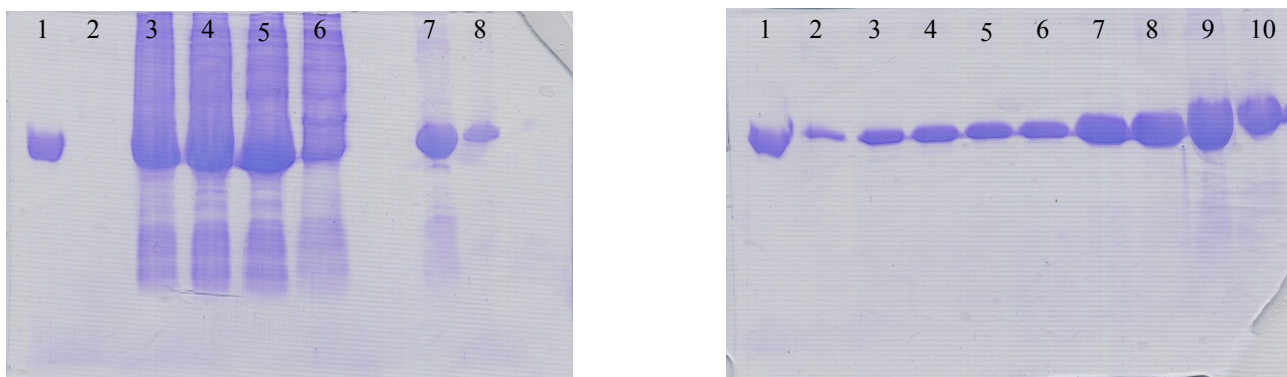


Figura 12. Purificación TIM 63. Izquierda, carril 1: TIM63, carril 2: medio donde crecieron las células, carril 3: células resuspendidas, carril 4: extracto crudo, carril 5: sobrenadante después de sonicar, carril 6: pellet. Carriles 7-8: fracciones eluidas de la columna de níquel a 500 mM imidazol. Derecha, carril 1: TIM63, carriles 2-6: fracciones que eluyeron a 75 mM de imidazol, carriles 7-8: fracciones que eluyeron a 130 mm imidazol, carriles 9-10: fracciones que eluyeron a 275 mM imidazol

Montaje de protocolo de purificación de TIM63

Los pasos de la purificación se siguieron por geles de SDS-acrilamida, midiendo concentración de proteína y actividad (como se describe en la metodología). Los resultados se muestran en la tabla 4 y en la figura 13. Se observa que la actividad específica no aumenta significativamente en los pasos finales de la purificación, lo que sugiere que la enzima se está inactivando en el proceso o que al medir la actividad en las primeras fracciones, hay una contribución de la TIM endógena.

Tabla 4. Tabla de purificación de TIM63

Paso	Volumen (ml)	Proteína total (mg)	Actividad total (umol/min)	Actividad específica (umol/min/mg)
Células resuspendidas	10	83.75	7540.19	90.03
Extracto crudo	10	80.03	6691.03	83.6
sobrenadante	10	73.15	6821.45	93.25
Elución a 75 mM	5.5	9.30	597.90	64.31
Elución a 130 mM	10	21.21	2046.46	96.46
Elución a 275 mM	10	19.26	1981.55	102.89
Proteína final		49.77	4625.91	

Propiedades espectroscópicas de TIM63

Fluorescencia intrínseca

Se tomaron espectros de emisión, excitando a 280 y 295 nm, en ambos casos observamos que los espectros presentan un máximo de emisión (λ_{max}) en 328.5 nm (Figura 14). Observamos también, que la intensidad obtenida excitando a 280 o a 295 nm es similar, como a 295 nm se excitan preferencialmente triptófanos, podemos afirmar que la mayor contribución en el espectro observado al excitar a 280 nm, proviene de la contribución de los triptófanos.

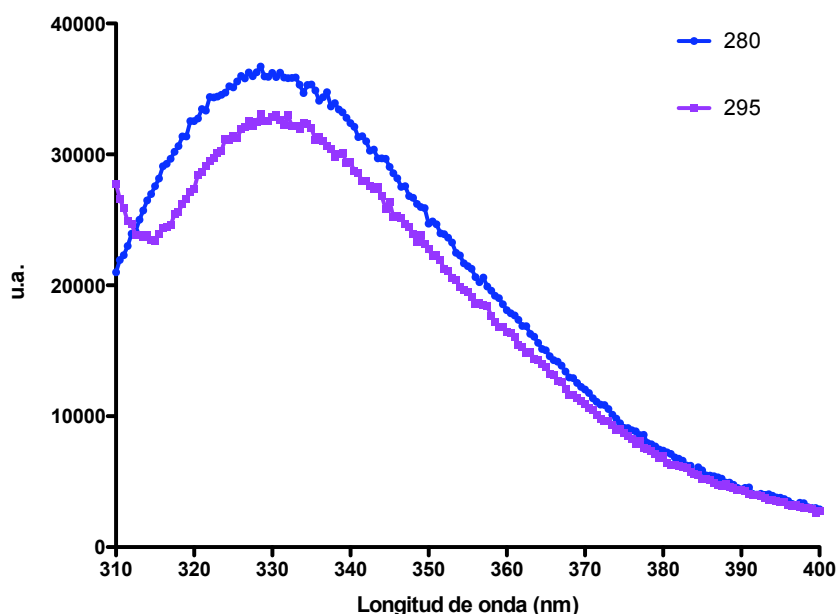


Figura 13. Espectros de fluorescencia de TIM63. En morado se observa el espectro de emisión de fluorescencia al excitar a 295 nm y en azul al excitar a 280 nm. La λ_{max} para ambos espectros es de 328.5 nm.

Dicroísmo Circular

El espectro de DC en el UV lejano mostrado en la figura 15 es el promedio de 2 barridos continuos y se corrigió restándole la señal del amortiguador. El espectro tiene los mínimos característicos cerca de 210 y 220 nm observados en proteínas que contienen hebras β y hélices α .

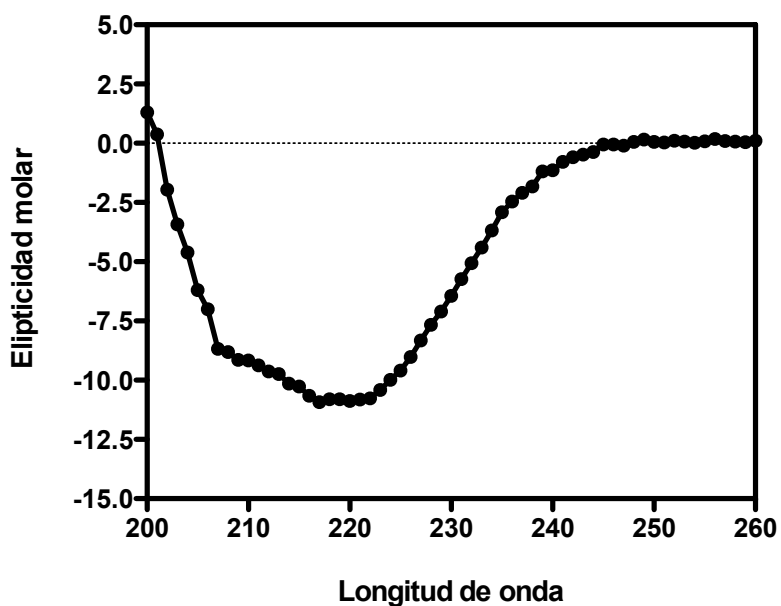


Figura 14. Espectro de dicroísmo circular de TIM63.

Desnaturalización por temperatura

Para analizar la transición de desplegamiento térmico, se graficó el cambio de elipticidad a 222 nm (Figura 16). El trazo se ajustó a una sigmoide con pendiente y se obtuvo un temperatura media de desnaturalización aparente ($T_{m_{app}}$) de 64.6 °C. Posteriormente se siguió el cambio de señal al enfriar de 90 a 20 °C, no se recuperó la señal original, indicando que la desnaturalización fue irreversible.

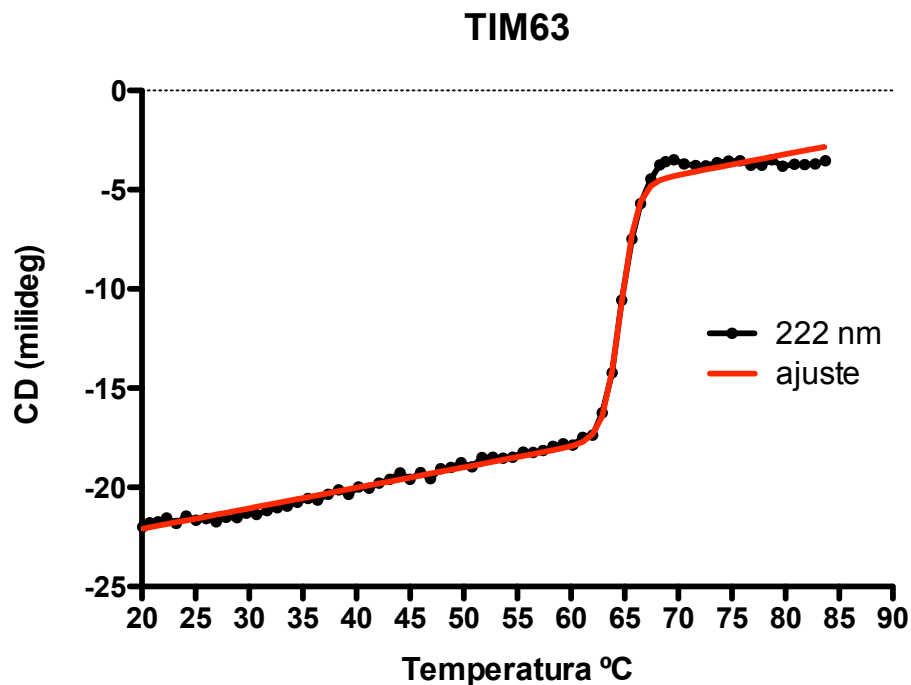


Figura 15. Transición térmica de TIM63. Se grafica el cambio de señal a 222 nm en un intervalo de temperatura de 20 a 90 °C. La concentración de proteína fue de 0.250 mg/ml en amortiguador Trietanolamina 10 mM.

Discusión

El tipo de plegamiento de los barriles TIM puede, en general, tolerar un considerable grado de diversidad en las secuencias, sin embargo varias posiciones de aminoácidos dentro de la secuencia de los barriles se encuentran conservadas. En el caso de la TIM varias regiones están bajo una presión selectiva significativa. Esto no aplica sólo a los residuos del sitio activo sino a regiones que estabilizan al barril o contribuyen a la dimerización. Esta presión se hace evidente con la saturación de transiciones y transversiones en la tercera posición de los codones, fenómeno observado en el artículo de Höhne donde concluyen que la saturación lleva a que al hacer reconstrucciones filogenéticas, los nodos tengan un menor soporte que aquellos que son reconstruidos a partir de RNA 18S (Knobeloch, Schmidt et al. 2010). Nuestros resultados son consistentes con los resultados de Höhne y muestran que la TIM por sí sola no es un buen marcador filogenético.

Se hacen varias suposiciones al hacer una reconstrucción ancestral, una muy importante es la del modelo o matriz de sustitución que se utiliza para inferir los estados ancestrales. La habilidad de modelar la evolución se ha mejorado sustancialmente durante la última década, aunque es aceptado que los modelos no son perfectos y falta mucho para que los modelos capturen exactamente todos los procesos evolutivos (Gaucher, Kratzer et al. 2010). Sin embargo, esto no debe verse como un paso limitante, una manera de afrontar el problema es mediante la prueba de diferentes modelos para ver los que mejor se ajusten a los datos que se tienen, como lo que hace el programa Prottest. De esta manera, aunque no se tengan modelos perfectos, al menos sabemos que estamos usando el modelo que mejor se ajusta a los datos experimentales.

Realizamos la inferencia de los estados ancestrales para eucariontes, procariontes y arqueas a diferentes profundidades de los árboles. En general se obtuvo que una gran proporción (>60%) de sitios se reconstruyen con una alta probabilidad posterior, sin embargo es necesario recordar que la probabilidad posterior de que una secuencia completa sea inferida correctamente es el producto de las probabilidades para cada uno de los sitios. Por ejemplo, para una proteína de 500 aminoácidos, si cada sitio se reconstruye con una probabilidad de 0.95, la probabilidad de que la secuencia sea correcta es $\sim 7.27 \times 10^{-12}$ (0.95^{500}). Por lo tanto, las secuencias ancestrales que obtenemos no son el verdadero ancestro, pero son nuestra mejor aproximación y asumimos que a pesar de que la secuencia

no sea exactamente la del ancestro, dará un buen acercamiento al fenotipo ancestral. (Thornton 2004)

Para conocer el fenotipo (propiedades fisicoquímicas y cinéticas) de las proteínas inferidas es necesario realizar la reconstrucción experimental en el laboratorio. Iniciamos la reconstrucción de los ancestros de los eucariontes porque eran secuencias con una alta proporción de sitios reconstruidos con una alta probabilidad, de igual forma observamos que los sitios que tenían una baja probabilidad se encontraban en la parte externa del barril, alejados del sitio catalítico y de la región relacionada con la dimerización. La importancia de que regiones como la relacionada con la dimerización tengan una alta probabilidad de reconstrucción, radica en que si observamos un efecto (en este caso en la asociación) podamos concluir que es una característica de la proteína ancestral y no un artefacto de la metodología. Las proteínas que se eligieron corresponden al último ancestro común de los eucariontes (TIM52), que vivió en el proterozoico hace cerca de 1.60 Giga años (Gyr), el último ancestro común de animales y hongos (TIM63), que vivió hace cerca de 1.37 Gyr, y un ancestro intermedio entre estos dos sucesos (TIM55)(Perez-Jimenez, Inglés-Prieto et al. 2011). De esta manera queremos observar los cambios de fenotipo que fue sufriendo la proteína en el tiempo, hasta ser la proteína que se conoce actualmente.

Las tres proteínas ancestrales pueden sobreexpresarse en *E. coli* y se encuentran de manera soluble. Las proteínas fueron subclonadas en Pet28b(+) para añadir una etiqueta de histidinas que facilitara su purificación, esta etiqueta consta de 20 aminoácidos dentro de los cuales, aparte de la etiqueta de histidinas se encuentra un sitio de corte por trombina, lo que permitirá posteriormente retirar la etiqueta.

En la purificación se obtuvo un buen rendimiento en la cantidad total de proteína, sin embargo la actividad específica no aumentó con la purificación, esto indica que la enzima se está inactivando, por lo que es necesario mejorar el proceso de purificación. Sin embargo, la actividad específica, obtenida hasta el momento, de TIM63, si bien no es tan alta como la de las TIM silvestres, sí es mayor que para otras TIMs diseñadas.

La tabla 5 muestra la *k_{cat}* de TIMs mutantes. Peimbert y cols, reportan TIMs de levadura en las que fue modificada la interfase, las DesTim fueron diseñadas por medio de Rosetta y las EvoTIM fueron modificadas mediante evolución dirigida. MonoTIM y RE-TIM son proteínas que se originaron por mutaciones a la TIM de *T. brucei*, con el fin de hacer una TIM monomérica. Podemos ver en la tabla que las actividades (*k_{cat}*) que se obtuvieron fueron

significativamente menores a las de las TIMs silvestres. Aunque no hemos encontrado aún las condiciones adecuadas para medir los parámetros cinéticos de TIM63, realizamos mediciones de actividad con 150 ng/ml de TIM en la celda de reacción y 4mM de G3P y obtuvimos valores de 1700 umolas/min/mg, estos valores de actividad nos hacen suponer que los valores de *k_{cat}* que se obtendrán podrían caer dentro del rango de las otras TIMs, lo que nos dice que pudimos inferir la secuencia de una TIM ancestral funcional.

Tabla 5. *k_{cat}* de TIMs diseñadas comparadas con TIMs silvestres.

Proteína	<i>k_{cat}</i> (s⁻¹)	referencia
Wild type levadura	6.4 x 10 ³	(Peimbert, Domínguez-Ramírez et al. 2008)
Wild type humano	3.7 x 10 ⁵	(Schliebs, Thanki et al. 1997)
LFYAA (DesTIM)	1.04 ± 0.04	(Peimbert, Domínguez-Ramírez et al. 2008)
LFFFL (DesTIM)	0.69 ± 0.05	(Peimbert, Domínguez-Ramírez et al. 2008)
PLATA (EvoTIM)	4.16 x 10 ³ ± 2.2 x 10 ²	(Peimbert, Domínguez-Ramírez et al. 2008)
AFAAS (EvoTIM)	7.02 x 10 ³ ± 2.1 x 10 ²	(Peimbert, Domínguez-Ramírez et al. 2008)
monoTIM	3.1 x 10 ²	(Schliebs, Thanki et al. 1997)
RE-TIM	1.3 x 10 ²	(Schliebs, Thanki et al. 1997)

Por otro lado, los espectros de fluorescencia y de dicroísmo circular se observan similares a los de otras TIM silvestres. Señal de que TIM63 está plegada y tiene estructura terciaria comparable con la de otros barriles.

Estos resultados deben leerse con precaución. La TIM63 que obtuvimos tiene aún la etiqueta de histidina, queda la duda de si ésta podría estar interfiriendo con la actividad de la enzima, por lo tanto es necesario repetir las mediciones con una TIM63 sin etiqueta, además de estandarizar mejor el protocolo de purificación.

Es difícil dar un significado evolutivo a los resultados obtenidos, si bien se podría decir que la perfección catalítica de TIM (como se definió en la introducción) se obtuvo después de la divergencia de animales y hongos, hace 1.37 Gyr, no es fácil distinguir entre un resultado negativo (baja actividad) y una mala reconstrucción, por la cantidad de supuestos que se han realizado.

Para aclarar esto, es necesario también realizar las mediciones para las demás TIM ancestrales, de tal manera que se pueda tener una comparación que sugiera cómo fueron cambiando las propiedades de la enzima.

La actividad de la TIM es importante, sin embargo también hay otras propiedades que es necesario explorar. La estructura secundaria y terciaria de la TIM63 es característica de los barriles TIM y los espectros son comparables con los de otras TIM, sin embargo falta estudiar la estructura cuaternaria (el estado de oligomerización) de la proteína al igual que su resistencia a la dilución. De esta manera tendremos información tanto de la estructura como de la función.

La desnaturalización térmica mostró que la proteína tenía una $T_{m,app}$ de 64.6°C, casi 10 °C más que la TbTIM y EhTIM, y cercana a yTIM. TIM63 tiene una $T_{m,app}$ en el mismo rango que las TIM diseñadas que se estabilizaron en el estudio de Peimbert et al. (Tabla 5). El aumento en la $T_{m,app}$ de TIM63, podría deberse a su naturaleza ancestral. El aumento en la T_m , es una característica que se ha observado en otras reconstrucciones ancestrales.(EF-TU, Tiorredoxinas) y se ha relacionado con un ambiente ancestral termofílico.

Tabla 5. T_m par diferentes TIM

Proteína	T_m (°C)	Referencia
TIM63	64.6	-
EhTIM	40 (diss) 55 (U)	(Tellez, 2008)
ScTIM	65.8	(Peimbert, 2008)
TbTIM	53.6	(Guzmán-Luna, 2006)
LFFYA	69.8	(Peimbert, 2008)
LFFFL	69.9	(Peimbert, 2008)

Se incluyen en la tabla TIMs silvestres y TIMs diseñadas (para estabilizar la interfase). Diss: disociación y U: desplegamiento.

El aumento en la estabilidad térmica parece ser una característica de las reconstrucciones ancestrales. Esto muestra un acercamiento en el que se llevan a cabo múltiples cambios en una proteína, en un sólo paso, y se obtienen enzimas más estables. Resultados similares se obtienen al hacer proteínas consenso (se hace un alineamiento y se coloca en cada sitio de la secuencia consenso, el residuo que se encuentra más veces en esa posición dentro del alineamiento). No se ha dado una explicación convincente para este fenómeno, sin embargo, se podría pensar que las secuencias consenso se parecen más a las proteínas ancestrales

que a las extantes, ya que se toman los aminoácidos que tienen en común un mayor número de secuencias en los alineamientos, aminoácidos que posiblemente estaban presentes en el ancestro común. (Lehmann, Pasamontes et al. 2000; Lehmann and Wyss 2001)

Como conclusión, en esta tesis mostramos que es posible inferir la secuencia de TIM ancestrales que se pueden expresar experimentalmente de manera soluble. Iniciamos la caracterización de TIM63, perteneciente al ancestro de animales y hongos. Esta proteína es activa, tiene estructura terciaria de barril (β/α)₈ y tiene una T_m aparente de 64.6 °C.

Es necesario continuar con la caracterización de TIM63, al igual que de TIM55 y TIM52, para poder compararlas y conocer los cambios que fue sufriendo la TIM de los eucariontes. Por otro lado, para entender la evolución de la TIM se debe estudiar el pasado, pero también el presente. El árbol de los eucariontes en la figura 9, muestra en óvalos verdes se señalan los organismos de los cuales se ha estudiado la TIM. Se puede ver que hay regiones que todavía no han sido estudiadas. Por esta razón, proponemos estudiar las tres proteínas extantes de los organismos señalados en rojo (*C. Elegans*, *Z. Mays* y *B. Morii*), trabajo que será llevado a cabo posteriormente en el laboratorio. De igual manera, un estudio similar será realizado para el grupo de procariontes y eventualmente con arqueas, de tal manera que se podrá contar la historia evolutiva de la trisfosfato isomerasa, y en el camino, conocer a fondo las propiedades fisicoquímicas de esta enzima.

Bibliografía

- Albery, W. J. and J. R. Knowles (1976). "Evolution of enzyme function and the development of catalytic efficiency." Biochemistry **15**: 5631-40.
- Alvarez, M., J. P. Zeelen, et al. (1998). "Triose-phosphate isomerase (TIM) of the psychrophilic bacterium *Vibrio marinus*. Kinetic and structural properties." The Journal of biological chemistry **273**: 2199-206.
- Beaucamp, N., a. Hofmann, et al. (1997). "Dissection of the gene of the bifunctional PGK-TIM fusion protein from the hyperthermophilic bacterium *Thermotoga maritima*: design and characterization of the separate triosephosphate isomerase." Protein science : a publication of the Protein Society **6**: 2159-65.
- Benítez-Cardoza, C. G., a. Rojo-Domínguez, et al. (2001). "Temperature-induced denaturation and renaturation of triosephosphate isomerase from *Saccharomyces cerevisiae*: evidence of dimerization coupled to refolding of the thermally unfolded protein." Biochemistry **40**: 9049-58.
- Benner, S. a., S. O. Sassi, et al. (2007). "Molecular paleoscience: systems biology from the past." Advances in enzymology and related areas of molecular biology **75**: 1-132, xi.
- Chávez-Cárdenas, M. E., D. A. Fernández-Velasco, et al. (2002). "Unfolding of triosephosphate isomerase from *Trypanosoma brucei*: identification of intermediates and insight into the denaturation pathway using tryptophan mutants." Archives of biochemistry and biophysics **399**: 117-29.
- Chávez-Cárdenas, M. E., G. Pérez-Hernández, et al. (2005). "Reversible equilibrium unfolding of triosephosphate isomerase from *Trypanosoma cruzi* in guanidinium hydrochloride involves stable dimeric and monomeric intermediates." Biochemistry **44**: 10883-92.
- Ciccarelli, F. D., T. Doerks, et al. (2006). "Toward automatic reconstruction of a highly resolved tree of life." Science **311**: 1283.
- Cole, M. F. and E. a. Gaucher (2010). "Exploiting Models of Molecular Evolution to Efficiently Direct Protein Engineering." Journal of molecular evolution.
- Creighton, T. E. (1997). Proteins, W. H. Freeman and Company.
- Field, S. F. and M. V. Matz (2010). "Retracing Evolution of Red Fluorescence in GFP-Like Proteins from Faviina Corals Research article." Molecular Biology **27**: 225-233.
- Gaucher, E. a., S. Govindarajan, et al. (2008). "Palaeotemperature trend for Precambrian life inferred from resurrected proteins." Nature **451**: 704-7.
- Gaucher, E. a., J. T. Kratzer, et al. (2010). "Deep phylogeny--how a tree can help characterize early life on Earth." Cold Spring Harbor perspectives in biology **2**: a002238.
- Gokhale, R. S., S. S. Ray, et al. (1999). "Unfolding of *Plasmodium falciparum* triosephosphate isomerase in urea and guanidinium chloride: evidence for a novel disulfide exchange reaction in a covalently cross-linked mutant." Biochemistry **38**: 423-31.
- Guzmán-Luna, V. (2006). Participación de las cisteínas en la estabilidad y estructura de la triosafosfato isomerasa de *Trypanosoma Brucei*.
- Jermann, T. M. O., J.G.; Stackhouse, J.; Benner, S.A. (1995). "Reconstructing the evolutionary history of the artiodactyl ribonuclease superfamily." Nature **374**: 57-59.
- Knobeloch, D., a. Schmidt, et al. (2010). "A coleopteran triosephosphate isomerase: X-ray structure and phylogenetic impact of insect sequences." Insect molecular biology **19**: 35-48.
- Knowles, J. (1991). "Enzyme catalysis: not different, just better." Nature **350**: 121-4.
- Konno, A., A. Kitagawa, et al. (2011). "Tracing Protein Evolution through Ancestral Structures of Fish Galectin." Structure (London, England : 1993) **19**: 711-21.

- Lambeir, a. M., J. Backmann, et al. (2000). "The ionization of a buried glutamic acid is thermodynamically linked to the stability of *Leishmania mexicana* triose phosphate isomerase." European journal of biochemistry / FEBS **267**: 2516-24.
- Le, S. Q. and O. Gascuel (2008). "An improved general amino acid replacement matrix." Molecular biology and evolution **25**: 1307-20.
- Lehmann, M., L. Pasamontes, et al. (2000). "The consensus concept for thermostability engineering of proteins." Nature **1543**: 408-415.
- Lehmann, M. and M. Wyss (2001). "Engineering proteins for thermostability : the use of sequence alignments versus rational design and directed evolution." Current Opinion in Biotechnology: 371-375.
- Lolis, E., T. Alber, et al. (1990). "Structure of yeast triosephosphate isomerase at 1.9-A resolution." Biochemistry **29**: 6609-18.
- López-Velázquez, G., D. Molina-Ortiz, et al. (2004). "An unusual triosephosphate isomerase from the early divergent eukaryote *Giardia lamblia*." Proteins **55**: 824-34.
- Lutzoni, F., F. Kauff, et al. (2004). "Assembling the fungal tree of life: progress, classification, and evolution of subcellular traits." American Journal of Botany **91**: 1446-1480.
- Mainfroid, V., S. C. Mande, et al. (1996). "Stabilization of human triosephosphate isomerase by improvement of the stability of individual alpha-helices in dimeric as well as monomeric forms of the protein." Biochemistry **35**: 4110-7.
- Mainfroid, V., P. Terpstra, et al. (1996). "Three hTIM mutants that provide new insights on why TIM is a dimer." Journal of molecular biology **257**: 441-56.
- Morgan, C. J., D. K. Wilkins, et al. (2000). "A compact monomeric intermediate identified by NMR in the denaturation of dimeric triose phosphate isomerase." Journal of molecular biology **300**: 11-6.
- Nagano, N. (2002). "One Fold with Many Functions: The Evolutionary Relationships between TIM Barrel Families Based on their Sequences, Structures and Functions." Journal of Molecular Biology **321**: 741-765.
- Nájera, H., M. Costas, et al. (2003). "Thermodynamic characterization of yeast triosephosphate isomerase refolding: insights into the interplay between function and stability as reasons for the oligomeric nature of the enzyme." The Biochemical journal **370**: 785-92.
- Nei, M. K., Sudhir (2000). *Molecular Evolution and Phylogenetics*, Oxford University Press.
- Ortlund, E. a., J. T. Bridgham, et al. (2007). "Crystal structure of an ancient protein: evolution by conformational epistasis." Science (New York, N.Y.) **317**: 1544-8.
- Page, R. D. and E. Holmes (1998). *Molecular Evolution. A Phylogenetic Approach*, Blackwell Science.
- Peimbert, M., L. Domínguez-Ramírez, et al. (2008). "Hydrophobic repacking of the dimer interface of triosephosphate isomerase by in silico design and directed evolution." Biochemistry **47**: 5556-64.
- Peimbert, M. e. a. (2008). *Triosephosphate Isomerase en Advances in Protein Physical Chemistry*, Transworld Research Network.
- Perez-Jimenez, R., A. Inglés-Prieto, et al. (2011). "Single-molecule paleoenzymology probes the chemistry of resurrected enzymes." Nature Structural & Molecular Biology: 1-6.
- Pompliano, D. L., a. Peyman, et al. (1990). "Stabilization of a reaction intermediate as a catalytic device: definition of the functional role of the flexible loop in triosephosphate isomerase." Biochemistry **29**: 3186-94.
- Reardon, D. F., Gregory (1995). "The structure and evolution of a/b barrel proteins." The FA-SEB Journal: 497-503.

- Richter, M., M. Bosnali, et al. (2010). "Computational and experimental evidence for the evolution of a (beta alpha)8-barrel protein from an ancestral quarter-barrel stabilised by disulfide bonds." Journal of molecular biology **398**: 763-73.
- Rietveld, A. W. and S. T. Ferreira (1998). "Kinetics and energetics of subunit dissociation/unfolding of TIM: the importance of oligomerization for conformational persistence and chemical stability of proteins." Biochemistry **37**: 933-7.
- Schliebs, W., N. Thanki, et al. (1997). "A double mutation at the tip of the dimer interface loop of triosephosphate isomerase generates active monomers with reduced stability." Biochemistry **36**: 9655-62.
- Sharma, A. S., Stephen G. (1999). Introduction to Fluorescence Spectroscopy, Wiley-Interscience.
- Shi , Y. L., JH; Zhang, HJ; Ding, Y (2008). "Equilibrium unfolding mechanism of chicken muscle triose phosphate isomerase." Protein Pept Lett. **15**: 365-70.
- Tellez, L. a., L. M. Blancas-Mejia, et al. (2008). "Thermal unfolding of triosephosphate isomerase from *Entamoeba histolytica*: dimer dissociation leads to extensive unfolding." Biochemistry **47**: 11665-73.
- Thomson, J. M., E. a. Gaucher, et al. (2005). "Resurrecting ancestral alcohol dehydrogenases from yeast." Nature genetics **37**: 630-5.
- Thornton, J. W. (2004). "Resurrecting ancient genes: experimental analysis of extinct molecules." Nature reviews. Genetics **5**: 366-75.
- Ugalde, J. a., B. S. W. Chang, et al. (2004). "Evolution of coral pigments recreated." Science (New York, N.Y.) **305**: 1433.
- Vázquez-Pérez, A. R. and D. A. Fernández-Velasco (2007). "Pressure and denaturants in the unfolding of triosephosphate isomerase: the monomeric intermediates of the enzymes from *Saccharomyces cerevisiae* and *Entamoeba histolytica*." Biochemistry **46**: 8624-33.
- Vega, C. L., Esben; Linden, Anni; Wilmanns, Matthias (2003). "Evolutionary markers in the (β / α)8-barrel fold." Current Opinion in Chemical Biology **7**: 694-701.
- Wallace, L. a. M., Robert (2002). "Highly divergent dihydrofolate reductases conserve complex folding mechanisms." Journal of molecular biology **315**: 193-211.
- Waterhouse, A. M., J. B. Procter, et al. (2009). "Jalview Version 2--a multiple sequence alignment editor and analysis workbench." Bioinformatics (Oxford, England) **25**: 1189-91.
- Williams, P. D., D. D. Pollock, et al. (2006). "Assessing the accuracy of ancestral protein reconstruction methods." PLoS computational biology **2**: e69.
- Yang, Z., S. Kumar, et al. (1995). "A new method of inference of ancestral nucleotide and amino acid sequences." Genetics **141**: 1641-50.
- Zuckerandl, E. and L. Pauling (1965). "Molecules as documents of evolutionary history." Journal of theoretical biology **8**: 357-66.
- Murzin, A.G., Brenner, S.E., et al.(1995). "SCOP: a structural classification of proteins database for the investigation of sequences and structures." Journal of molecular biology **247**, 536-540

ANEXO 1. Números de acceso y claves de las secuencias utilizadas para las reconstrucciones ancestrales. Como número de acceso se muestra la clave GI de la base de datos del NCBI y después de los dos puntos, la región que corresponde al cDNA de la TIM. La secuencia de *Escherichia coli* fue retirada de NCBI, pero aún es posible acceder con el número de acceso de esta tabla.

Número de acceso	Organismo	Dominio	Clave
226529872:39-788	<i>Homo sapiens</i>	eucarionte	HomoSapie
93117368:555724-556470	<i>Saccharomyces cerevisiae</i>	eucarionte	saccharom
212771:118-864	<i>Gallus gallus</i>	eucarionte	chicken
226958348:25-924	<i>Mus musculus</i>	eucarionte	musmuscul
28572003:67-1113	<i>Drosophila melanogaster</i>	eucarionte	drosophil
63054406:939945-940694	<i>Schizosaccharomyces pombe</i>	eucarionte	schozosac
68472495:64914-65660	<i>Candida albicans</i>	eucarionte	candidaal
212645598:10-753	<i>Caenorhabditis elegans</i>	eucarionte	caenorhab
187281707:1-747	<i>Bombyx mori</i>	eucarionte	bombyxmor
169856983:1-756	<i>Coprinopsis cinerea okayama</i>	eucarionte	coprinops
156690:1-744	<i>Culex tarsalis</i>	eucarionte	culextars
67541459:1-750	<i>Aspergillus nidulans</i>	eucarionte	aspergill
109133020:45-791	<i>Macaca mulatta</i>	eucarionte	macacamul
115392090:1-750	<i>Pan troglodytes</i>	eucarionte	pantroglo
24475833:78-824	<i>Danio rerio</i>	eucarionte	Danioreri
47075137:1-747	<i>Ashbya gossypii</i>	eucarionte	ashbyagos
75832099:39-788	<i>Bos taurus</i>	eucarionte	bostaurus
73997308:90-839	<i>Canis familiaris</i>	eucarionte	canisfamis
50289458:1-747	<i>Candida glabrata</i>	eucarionte	candidagl
202952834:597317-598063	<i>Debaryomyces hansenii</i>	eucarionte	Debaryomy
66821310:1-774	<i>Dictyostelium discoideum</i>	eucarionte	dyctioste
50313011:1729736-1730482	<i>Kluyveromyces lactis</i>	eucarionte	kluuverom
80971509:42-792	<i>Sus scrofa</i>	eucarionte	susscrofa
117935063:25-774	<i>Rattus norvegicus</i>	eucarionte	rattusnor
4965073:247294-248133	<i>Yarrowia lipolytica</i>	eucarionte	yarrowial
164423940:154-900	<i>Neurospora crassa</i>	eucarionte	neurospor
40805817:100-849	<i>Paracoccidioides brasiliensis</i>	eucarionte	paracocci
46849382:1-631	<i>Oryzias latipes</i>	eucarionte	oryziasla
197097647:41-790	<i>Pongo abelii</i>	eucarionte	pongoabel
148236350:40-768	<i>Xenopus laevis</i>	eucarionte	xenopusla
258597794:318-1064	<i>Plasmodium falciparum</i>	eucarionte	plasmodiu
471325:590-1345	<i>Leishmania mexicana</i>	eucarionte	lmexicana
164451110:136381-137136	<i>Trypanosoma cruzi</i>	eucarionte	tripanoso
256071062:1-762	<i>Schistosoma mansoni</i>	eucarionte	schistos1
183233902:24-809	<i>Entamoeba histolytica</i>	eucarionte	entamoeba
237688746:1-774	<i>Giardia lamblia</i>	eucarionte	giardiala
226471579:32-790	<i>Schistosoma japonicum</i>	eucarionte	schistos2
10644116:22-774	<i>Taenia solium</i>	eucarionte	taeniasol
115480366:54-968	<i>Oryza sativa</i>	eucarionte	oryzasativa
340520:1-762	<i>Coptis japonica</i>	eucarionte	coptisjap

Número de acceso	Organismo	Dominio	Clave
226495390:260-1030	<i>Zea mays</i>	eucarionte	zmays
602589:1-765	<i>P.hybrida</i>	eucarionte	phybrida
609261:64-960	<i>Secale cereale</i>	eucarionte	scereale
1785947:61-822	<i>Hordeum vulgare</i>	eucarionte	hordeumvu
254000649:191092-191796	<i>Helicobacter pylori</i>	procarionte	Helicob1
47118316:5326376-5327131	<i>Bradyrhizobium japonicum</i>	procarionte	Bradyrhi
47118328:6013696-6014460	<i>Mesorhizobium loti</i>	procarionte	Mesorhiz
254667570:297132-297905	<i>Neisseria meningitidis</i>	procarionte	Neisseri
156617157:752410-753201	<i>Haemophilus influenzae</i>	procarionte	Haemoph1
33149228:598103-598873	<i>Haemophilus ducreyi</i>	procarionte	Haemoph2
12057206:369332-370156	<i>Chlamydia trachomatis</i>	procarionte	Chlamytr
41584206:300711-301436	<i>Bdellovibrio bacteriovorus</i>	procarionte	Bdellovi
33862273:894457-895188	<i>Prochlorococcus marinus</i>	procarionte	Prochlorm
57636010:448998-449759	<i>Staphylococcus epidermidis</i>	procarionte	Staphyep
83316099:2483944-2484699	<i>Listeria monocytogenes</i>	procarionte	Listemon
16414852:214432-215187	<i>Listeria innocua</i>	procarionte	Listeinn
38199912:281324-282106	<i>Corynebacterium diphtheriae</i>	procarionte	Corynedi
17986243:443069-443839	<i>gravis</i>	procarionte	Corynedi
32263428:1273743-1274450	<i>Brucella melitensis</i>	procarionte	Brucelme
81167692:1284223-1285014	<i>Helicobacter hepaticus</i>	procarionte	Helicob2
37508091:1114986-1115708	<i>Synechococcus elongatus</i>	procarionte	Synechel
261370870:2915526-2916287	<i>Gloeobacter violaceus</i>	procarionte	Gloeobav
6626248:237559-238302	<i>Fibrobacter succinogenes</i>	procarionte	Fibrobas
46911657:241320-242108	<i>Aquifex aeolicus</i>	procarionte	Aquifex
20095250:2016853-2017608	<i>Photobacterium profundum</i>	procarionte	Photobap
148503909:1616874-1617659	<i>Fusobacterium nucleatum</i>	procarionte	Fusobanu
29251571:723587-724399	<i>Mycobacterium tuberculosis</i>	procarionte	Mycobtub
10176182:3659870-3660625	<i>Chlamydia muridarum Nigg</i>	procarionte	Chlamymu
162809319:200974-201798	<i>Bacillus halodurans</i>	procarionte	Bacilhal
140843962:1818516-1819295	<i>Onion yellows phytoplasma</i>	procarionte	Onionyel
49240382:875019-875780	<i>Corynebacterium glutamicum</i>	procarionte	Corynegl
159139455:1604904-1605674	<i>Staphylococcus aureus</i>	procarionte	Corynegl
33236960:1215655-1216419	<i>subsp. aureus cepa</i>	procarionte	Staphyau
	<i>MRSA252</i>	procarionte	Staphyau
	<i>Agrobacterium tumefaciens</i>	procarionte	Agrobacte
	<i>Chlamydophila pneumoniae</i>	procarionte	Chlamydp

Número de acceso	Organismo	Dominio	Clave
42632302:2491393-2492151	<i>Oceanobacillus iheyensis</i>	procarionte	Oceanobih
167964044:313011-313760	<i>Xylella fastidiosa M12</i>	procarionte	Xylellafa
229264291:4861262-4862017	<i>Bacillus anthracis str. A0248</i>	procarionte	Bacilant Blochflo
33518905:669260-670048	<i>Blochmannia floridanus</i>	procarionte	
116619145:c8151196-8150447	<i>Solibacter usitatus Ellin6076</i>	procarionte	Solibaus
225791207:721488-722237	<i>Acidobacterium capsulatum</i>	procarionte	Acidobca
42516522:1269926-1270378	<i>Treponema denticola</i>	procarionte	Treponem1
189025236:580487-581236	<i>Treponema pallidum</i>	procarionte	Treponem2
66043271:4976938-4977693	<i>Pseudomonas syringae</i>	procarionte	Pseudomo2
110645304:5332746-5333501	<i>Pseudomonas aeruginosa</i>	procarionte	Pseudomo3
206588805:313633-314379	<i>Ralstonia solanacearum</i>	procarionte	Ralstoni2
21240774:3166135-3166890	<i>Xanthomonas axonopodis</i>	procarionte	Xanthomo2
188989396:1878463-1879263	<i>Xanthomonas campestris</i>	procarionte	xanthomo3
33594723:221899-222645	<i>Bordetella parapertussis</i>	procarionte	Bordetel1
16120353:96170-96937	<i>Yersinia pestis</i>	procarionte	Yersinia1
28896774:253572-254342	<i>Vibrio parahaemolyticus</i>	procarionte	Vibriopar
117956319:1327349-1328119	<i>Vibrio vulnificus</i>	procarionte	Vibriovul
147673035:2391102-2391899	<i>Vibrio cholerae</i>	procarionte	Vibriocho
29345410:5102705-5103463	<i>Bacteroides thetaiotaomicron</i>	procarionte	Bacteroi2
30248031:1922996-1923754	<i>Nitrosomonas europaea</i>	procarionte	Nitrosom2
15805042:1346447-1347181	<i>Deinococcus radiodurans</i>	procarionte	Deinococ1
24378532:674360-675118	<i>Streptococcus mutans</i>	procarionte	Streptoc1
94991497:495522-496280	<i>Streptococcus pyogenes</i>	procarionte	Streptoc2
116515308:1423585-1424343	<i>Streptococcus pneumoniae</i>	procarionte	Streptoc5
15671982:1148515-1149273	<i>Lactococcus lactis subsp lactis</i>	procarionte	Lactococl
15893298:823748-824494	<i>Clostridium acetobutylicum</i>	procarionte	clostrid3
28209834:369920-370687	<i>Clostridium tetani</i>	procarionte	clostrid5
18308982:1541381-1542127	<i>Clostridium perfringens</i>	procarionte	clostrid6
42518084:779566-780321	<i>Lactobacillus johnsonii</i>	procarionte	lactobac5
254555169:693053-693811	<i>Lactobacillus plantarum</i>	procarionte	lactobac8
189423082:1721400-1722155	<i>Geobacter lovleyi</i>	procarionte	geobactel
42519920:83976-84704	<i>Wolbachia endosymbiont of Drosophila melanogaster</i>	procarionte	wolbachi1
24371600:1244412-1245194	<i>Shewanella oneidensis</i>	procarionte	shewanel1
20806542:1707322-1708068	<i>Thermoanaerobacter tengcongensis</i>	procarionte	thermoan3
16124256:2088540-2089301	<i>Caulobacter crescentus</i>	procarionte	caulobact
189438863:518425-519224	<i>Bifidobacterium longum</i>	procarionte	Bifidoba1
57233530:683187-683942	<i>Dehalococcoides ethenogenes</i>	procarionte	dehaloco1
32490749:324239-325009	<i>Wigglesworthia glossinidia</i>	procarionte	wiggleswo
254286927:4017217-4017984	<i>Escherichia coli</i>	procarionte	ecoli
218547440:3948347-3949114	<i>Escherichia fergusonii</i>	procarionte	eferguson

Número de acceso	Organismo	Dominio	Clave
110804074:3654885-3655652	<i>Shigella flexneri</i>	procarionte	shigellaf
16763390:4290265-4291032	<i>Salmonella typhimurium</i>	procarionte	salmonelt
29140506:3640059-3640826	<i>Salmonella enterica</i>	procarionte	salmonele
37524032:338493-339260	<i>Photorhabdus luminescens</i>	procarionte	photorhab
15601865:1505418-1506203	<i>Pasteurella multocida</i>	procarionte	pasteurel
148545259:5125580-5126335	<i>Pseudomonas putida</i>	procarionte	pseudomo6
212217688:530991-531806	<i>Coxiella burnetii</i>	procarionte	coxiellab
34495455:974488-975234	<i>Chromobacterium violaceum</i>	procarionte	chromobac
39649689:124989-125825	<i>Rhodopseudomonas palustris</i>	procarionte	rhodopseu
34556458:71183-71881	<i>Wolinella succinogenes</i>	procarionte	wolinella
157163852:1523241-1523924	<i>Campylobacter concisus</i>	procarionte	campylob1
46562128:1756281-1757036	<i>Desulfovibrio vulgaris</i>	procarionte	desulfovi
46198308:564035-564787	<i>Thermus thermophilus</i>	procarionte	thermusth
16329170:2423621-2424349	<i>Synechocystis</i>	procarionte	synechocy
186680550:5410307-5411038	<i>Nostoc punctiforme</i>	procarionte	Nostocpun
162960844:7562063-7562839	<i>Streptomyces avermitilis</i>	procarionte	streptom1
32141095:259065-259841	<i>Streptomyces coelicolor</i>	procarionte	Streptom2
119025018:1041321-1042115	<i>Bifidobacterium adolescentis</i>	procarionte	bifidoba2
218249165:50442-51203	<i>Borrelia burgdorferi</i>	procarionte	borreliab
167777849:2177053-2177802	<i>Leptospira biflexa</i>	procarionte	leptospir
116329799:1449242-1449994	<i>Leptospira borgpetersenii</i>	procarionte	Leptospi1
45655914:2533875-2534627	<i>Leptospira interrogans</i>	procarionte	Leptospi2
21672841:1356164-1356901	<i>Chlorobium tepidum</i>	procarionte	chlorobi1
119355857:2147117-2147866	<i>Chlorobium phaeobacteroides</i>	procarionte	chlorobi2
34539880:677401-678156	<i>Porphyromonas gingivalis</i>	procarionte	Porphyrom
29374661:1899386-1900141	<i>Enterococcus faecalis</i>	procarionte	Enterococ
15668172:1505740-1506399	<i>Methanocaldococcus jannaschii</i>	Arquea	Metjanna
21226102:1513886-1514554	<i>Methanosarcina mazei</i>	Arquea	Metmazei
20088899:5683577-5684245	<i>Methanosarcina acetivorans</i>	Arquea	Metacet
11497621:1163419-1164090	<i>Archaeoglobus fulgidus</i>	Arquea	Argloful
18976372:1771227-1771913	<i>Pyrococcus furiosus</i>	Arquea	Pyrfurio
15896971:2358713-2359396	<i>Sulfolobus solfataricus</i>	Arquea	Sulfsof
16081186:334841-335491	<i>Thermoplasma acidophilum</i>	Arquea	Theracid
20093440:1666501-1667181	<i>Methanopyrus kandleri</i>	Arquea	Metkand
18311643:879522-880205	<i>Pyrobaculum aerophilum</i>	Arquea	Pyroaero
14518450:1712660-1713346	<i>Pyrococcus abyssi</i>	Arquea	Pyroabys
150401930:1655748-1656401	<i>Methanococcus maripaludis</i>	Arquea	Metmarip
13540831:1351870-1352511	<i>Thermoplasma volcanium</i>	Arquea	Thevolcan
14589963:1638174-1638869	<i>Pyrococcus horikoshii</i>	Arquea	Pyrohori
118430835:970493-971164	<i>Aeropyrum pernix</i>	Arquea	Aeroprnx
15789340:786530-787174	<i>Halobacterium</i>	Arquea	Halobac
24473558:2045425-2046114	<i>Sulfolobus tokodaii</i>	Arquea	Sulftoko
15678031:951977-952666	<i>Methanothermobacter thermautotrophicus</i>	Arquea	Metther

Anexo 2. Alineamientos múltiples de secuencias de TIM.

Eucariontes

	1	10	20	30	40	50																																														
HomoSapie	M	A	P	S	R	K	F	F	V	G	G	N	W	K	M	N	G	R	-	K	Q	S	L	G	E	L	I	G	T	L	N	A	A	K	V	P	A	-	D	-	-	T	E	V	V	C	A	P	P	T		
Macacumul	M	A	P	S	R	K	F	F	V	G	G	N	W	K	M	N	G	R	Q	-	K	Q	N	L	G	E	L	I	G	T	L	N	A	K	V	P	A	-	D	-	-	T	E	V	V	R	A	P	P	T		
bostaurus	M	A	P	S	R	K	F	F	V	G	G	N	W	K	M	N	G	R	-	K	K	N	L	G	E	L	I	N	T	L	N	A	A	K	V	P	A	-	D	-	-	T	E	V	V	C	A	P	P	T		
canisfami	M	A	P	S	R	K	F	F	V	G	G	N	W	K	M	N	G	R	-	K	K	N	L	G	E	L	I	T	T	L	N	A	A	K	V	P	A	-	D	-	-	T	E	V	V	C	A	P	P	T		
Danioreri	M	-	S	S	R	K	F	F	V	G	G	N	W	K	M	N	G	D	-	K	E	S	L	G	E	L	I	M	T	L	N	A	S	L	N	D	-	K	-	-	T	D	V	V	C	G	A	P	S			
pantroglo	M	A	P	S	R	K	F	F	V	G	G	N	W	K	M	N	G	R	-	K	Q	S	L	G	E	L	I	G	T	L	N	A	A	K	V	P	A	-	D	-	-	T	E	V	V	C	A	P	P	T		
Pongoabel	M	A	P	S	R	K	F	F	V	G	G	N	W	K	M	N	G	R	-	K	Q	S	L	G	E	L	I	G	T	L	N	A	A	K	V	P	A	-	D	-	-	T	E	V	V	C	A	P	P	T		
chicken	M	A	P	-	R	K	F	F	V	G	G	N	W	K	M	N	G	D	-	K	K	S	L	G	E	L	I	H	T	L	N	G	A	K	L	S	A	-	D	-	-	T	E	V	V	C	G	A	P	S		
Oryziasla	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	?	D	-	K	K	S	L	G	E	L	I	Q	T	L	N	A	G	K	V	D	P	-	N	-	-	V	E	V	V	C	G	A	P	S
Rattusnor	M	A	P	S	R	K	F	F	V	G	G	N	W	K	M	N	G	R	-	K	K	C	L	G	E	L	I	C	T	L	N	A	A	K	V	P	A	-	D	-	-	T	E	V	V	C	A	P	P	T		
musmuscul	M	A	P	T	R	K	F	F	V	G	G	N	W	K	M	N	G	R	-	K	K	C	L	G	E	L	I	C	T	L	N	A	A	N	V	P	A	-	G	-	-	T	E	V	V	C	A	P	P	T		
Xenopusla	M	S	P	-	R	K	F	F	V	G	G	N	W	K	M	N	G	D	-	K	K	S	L	G	E	L	I	N	T	L	N	S	G	K	M	N	A	-	D	-	-	T	E	V	V	C	G	A	P	A		
Susscrofa	M	A	P	A	R	K	F	F	V	G	G	N	W	K	M	N	G	R	-	K	K	N	L	G	E	L	I	N	T	L	N	A	A	K	L	P	A	-	D	-	-	T	E	V	V	C	A	P	P	T		
Bombymor	M	G	-	-	R	K	F	V	V	G	G	N	W	K	M	N	G	D	-	K	N	Q	I	N	E	I	V	N	N	L	K	K	G	P	L	D	P	-	N	-	-	V	E	V	I	V	G	V	P	A		
drosophil	M	S	-	-	R	K	F	C	V	G	G	N	W	K	M	N	G	D	-	Q	K	S	I	A	E	I	A	K	T	L	S	S	A	A	L	D	P	-	N	-	-	T	E	V	V	I	G	C	P	A		
Culextars	M	G	-	-	R	K	F	C	V	G	G	N	W	K	M	N	G	D	-	K	A	S	I	A	D	L	C	K	V	L	T	T	G	P	L	N	A	-	D	-	-	T	E	V	V	V	G	C	P	A		
Taeniasol	M	T	-	-	R	K	L	F	V	G	G	N	W	K	M	N	G	S	-	Y	S	H	I	N	T	F	F	D	T	L	Q	K	A	D	T	D	P	-	N	-	-	A	D	I	V	I	G	V	P	A		
Schistos1	M	S	G	S	R	K	F	F	V	G	G	N	W	K	M	N	G	S	-	R	D	D	N	D	K	L	L	K	L	S	E	A	H	F	D	D	-	N	-	-	T	E	V	L	I	A	P	P	S			
Schistos2	M	S	G	S	R	K	F	F	V	G	G	N	W	K	M	N	G	S	-	R	A	D	N	K	V	L	I	H	S	S	E	A	H	F	C	G	-	D	-	-	T	E	I	L	I	A	A	P	F			
Candidaal	M	A	-	-	R	Q	F	F	V	G	G	N	F	K	M	N	G	T	-	K	Q	Q	I	T	S	I	D	N	L	N	K	A	D	L	P	K	-	D	-	-	V	E	V	V	I	C	P	P	A			
Debaryomy	M	A	-	-	R	Q	F	F	V	G	G	N	F	K	M	N	G	T	-	R	E	S	V	S	K	I	V	D	G	L	N	K	A	E	L	P	S	-	N	-	-	V	E	V	V	I	A	P	P	A		
Coprinops	M	T	-	-	R	S	F	F	V	G	G	N	W	K	L	N	P	T	S	-	I	S	A	A	K	A	L	V	E	A	N	K	A	D	L	D	P	-	S	-	-	T	E	V	V	V	A	P	P	A		
Caenorhab	M	T	-	-	R	K	F	F	V	G	G	N	W	K	M	N	G	D	-	Y	A	S	V	D	G	I	V	T	F	L	N	A	S	A	D	N	S	-	S	-	-	V	D	V	V	V	A	P	P	A		
Plasmodiu	M	A	-	-	R	K	Y	F	V	A	A	N	W	K	C	N	G	T	-	L	E	S	I	K	S	L	T	N	S	F	N	N	D	F	D	P	S	K	-	-	L	D	V	V	F	P	V	S				
Dictyoste	M	T	G	T	R	T	F	F	V	G	G	N	H	K	M	N	G	S	-	K	S	M	L	E	S	L	S	K	G	M	N	E	S	V	E	N	K	E	N	-	-	V	D	I	F	I	A	P	S	Y		
Neurospor	M	A	-	-	R	K	F	F	V	G	G	N	F	K	M	N	G	T	-	R	K	D	L	K	A	I	V	D	N	L	N	N	A	Q	L	D	P	-	N	-	-	A	E	V	V	I	A	P	P	A		
Aspergill	M	P	-	-	R	K	F	F	V	G	G	N	F	K	M	N	G	N	-	A	E	S	T	S	I	K	N	L	N	S	A	N	L	D	K	S	-	S	-	-	V	E	V	V	S	P	P	A				
Yarrowial	L	T	Q	S	R	T	F	F	V	G	G	N	F	K	M	N	G	S	-	L	E	S	I	K	A	I	V	E	R	L	N	A	S	E	L	D	P	-	K	-	-	T	E	V	V	I	S	P	P	F		
Paracocci	M	P	-	-	R	K	F	F	V	G	G	N	F	K	M	N	G	T	-	A	K	S	I	T	H	I	T	N	L	N	S	A	K	L	D	P	-	S	-	-	T	E	I	V	I	A	P	P	A			
Kluyverom	M	A	-	-	R	T	F	F	I	G	G	N	F	K	M	N	G	S	-	K	A	S	I	K	E	I	V	D	R	L	N	G	A	S	I	P	S	-	N	-	-	V	E	V	V	I	A	P	P	A		
Schizosac	M	A	-	-	R	K	F	F	V	G	G	N	F	K	M	N	G	S	-	L	E	S	M	K	T	I	E	G	N	T	T	K	L	N	V	G	D	-	-	V	E	T	V	I	F	P	Q	N				
Candidagl	M	A	-	-	R	T	F	F	V	G	G	N	F	K	L	N	G	T	-	K	K	S	I	K	E	I	V	D	R	L	N	T	A	S	L	P	E	-	N	-	-	V	E	V	V	I	C	P	P	A		
Saccharom	M	A	-	-	R	T	F	F	V	G	G	N	F	K	L	N	G	S	-	K	Q	S	I	K	E	I	V	E	R	L	N	T	A	S	I	P	E	-	N	-	-	V	E	V	V	I	C	P	P	A		
Ashbyagos	M	A	-	-	R	T	F	F	V	G	G	N	F	K	L	N	G	S	-	K	Q	S	I	K	E	I	V	E	R	L	N	T	A	D	L	A	D	-	N	-	-	V	E	V	V	I	C	P	P	A		
Giardiala	M	P	A	-	R	R	P	F	I	G	G	N	F	K	C	N	G	S	-	L	D	F	I	K	S	H	V	A	A	I	A	A	H	K	I	P	D	-	S	-	-	V	D	V	V	I	A	P	S	A		
Entamoeba	M	S	A	-	R	K	F	V	V	G	G	N	W	K	C	N	G	T	-	L	A	S	I	E	T	L	T	K	G	V	A	A	S	V	D	A	E	L	A	K	K	V	E	V	I	V	G	V	P	F		
lmexicana	M	S	A	K	P	Q	P	I	A	A	A	N	W	K	C	N	G	T	-	T	A	S	I	E	K	L	V	Q	V	F	N	E	H	T	I	S	H	-	D	-	-	V	Q	C	V	V	A	P	T	F		
Trypanoso	M	A	S	K	P	Q	P	I	A	A	A	N	W	K	C	N	G	S	-	E	S	L	V	P	L	I	E	T	L	N	A	A	T	F	D	H	-	D	-	-	V	Q	C	V	V	A	P	T	F			
Zeamays	M	G	-	-	R	K	F	F	V	G	G	N	W	K	C	N	G	T	-	T	D	Q	V	E	K	I	V	K	T	L	N	E	G	Q	V	P	P	S	D	-	S	-	V	V	E	V	V	S	P	P	Y	
Oryzasati	M	A	G	S	G	K	F	F	V	G	G	N	W	K	C	N	G	T	-	K	D	S	V	S	K	L	V	T	E	L	N	A	A	T	L	E	P	-	D	-	-	V	D	V	V	V	A	P	P	F		
Scereale	M	A	G	S	G	K	F	F	V	G	G	N	W	K	C	N	G	T	-	K	E	S	I	S	K	L	V	S	D	L	N	A	A	T	L	E	S	-	D	-	-	V	D	V	V	V	A	P	P	F		
Coptisjap	M	G	-	-	R	K	F	F	V	G	G	N	W	K	C	N	G	T	-	S	E	E	V	K	K	I	V	T	L	L	N	E	A	E	V	P	S	E	D	-	-	V	V	E	V	V	S	P	P	Y		
Phybrida	M	G	-	-	R	K	F	F	V	G	G	N	W	K	C	N	G	T	-	A	E	E	V	K	K	I	L	A	T	L	N	A	A	D	V	P	S	Q	D	-	-	V	V	E	V	V	S	P	P	Y		
Hordeumvu	M	G	-	-	R	K	F	V	G	G	N	W	K	C	N	G	T	-	V	E	Q	V	E	A	I	V	Q	T	L	N	A	G	Q	I	V	S	P	D	-	-	V	V	E	V	V	S	P	P	Y			
Streptoc1	M	S	-	-	R	K	P	I	I	A	G	N	W	K	M	N	K	T	-	A	A	E	A	R	E	F	I	D	A	V	K	N	N	I	P	S	N															

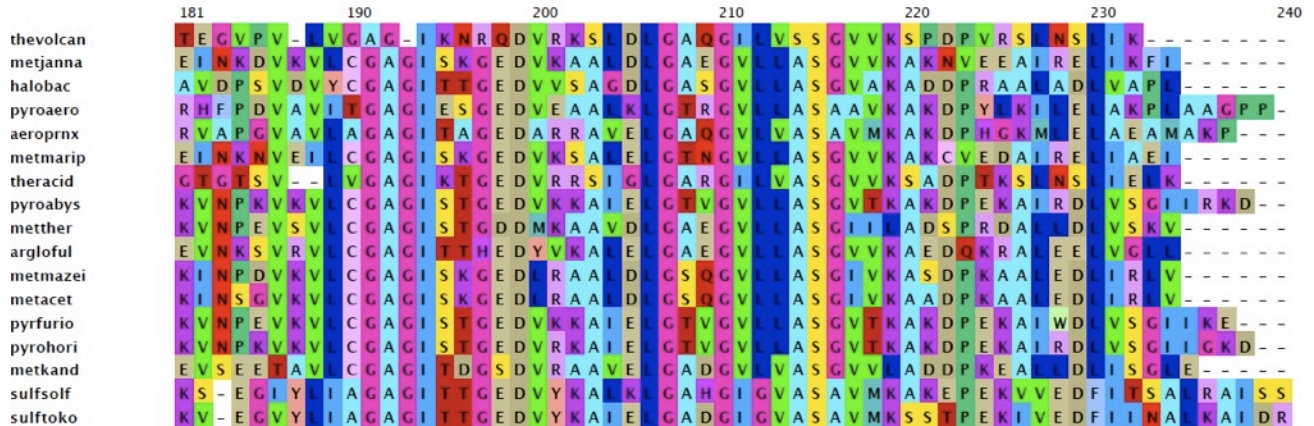
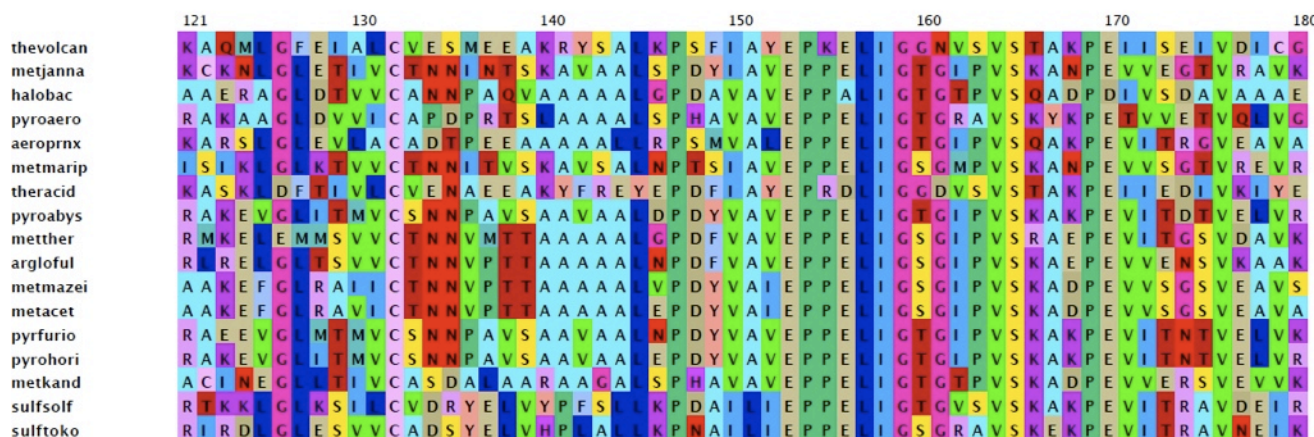
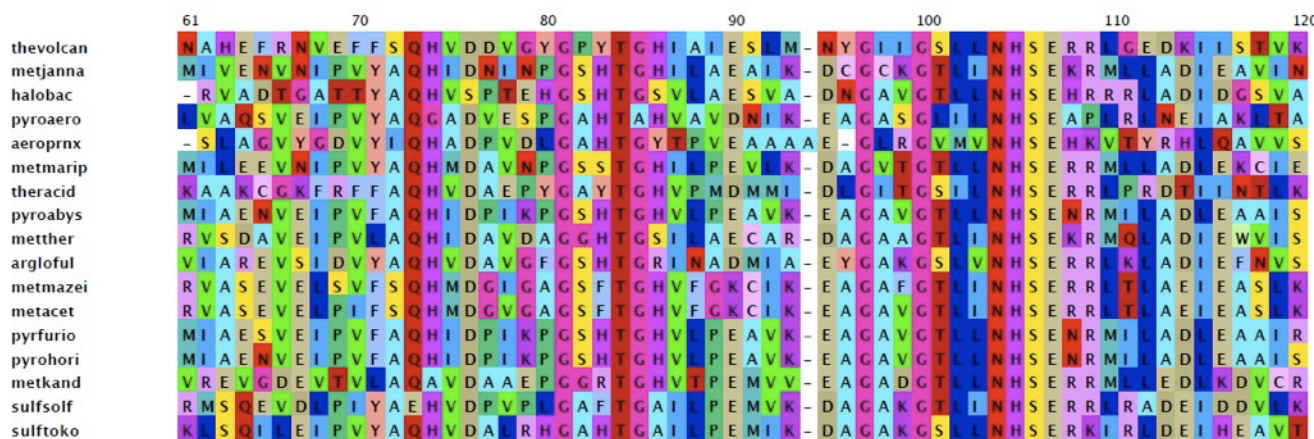
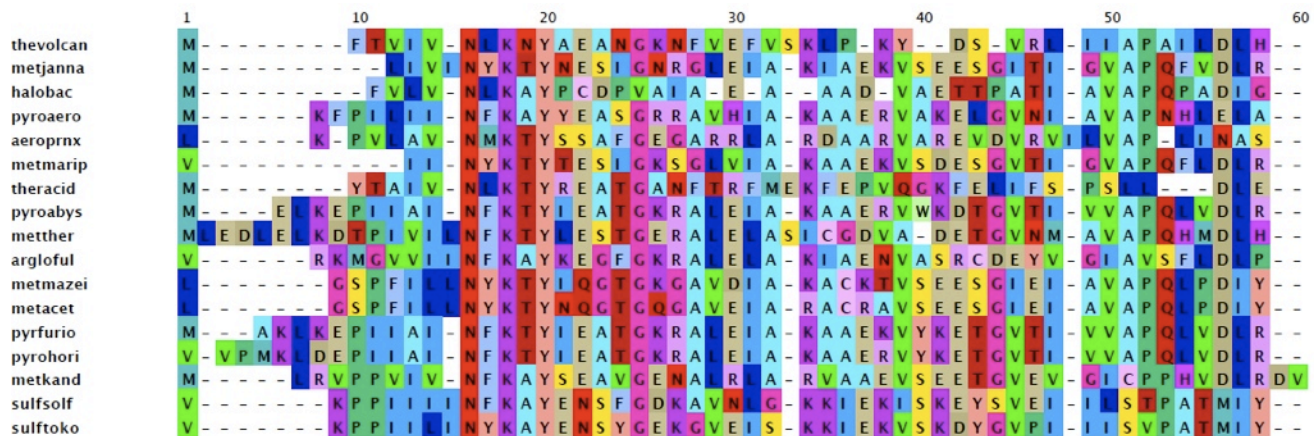
	51	60	70	80	90	100
HomoSapie	A Y I D - F A R Q K - - - L D - P K - I A - - - V A A Q N C Y K V T N G A F T G E I S P G M I K D C					
Macacamul	A Y I D - F A Q Q K - - - L D - P K - T A - - - V A A Q N C Y K V T N G D F T G E I S P G M I K D C					
bostaurus	A Y I D - F A R Q K - - - L D - P K - I A - - - V A A Q N C Y K V A N G A F T G E I S P G M I K D C					
canisfami	A Y I D - F A R Q K - - - L D - A K - I A - - - V A A Q N C Y K V T N G A F T G E I S P G M I K D C					
Danioreri	I Y L D - Y A R S K - - - L D - Q R - I G - - - V A A Q N C Y K V P K G A F T G E I S P A M I K D C					
pantro glo	A Y I D - F A R Q K - - - L D - P K - I A - - - V A A Q N C Y K V T N G A F T G E I S P G M I K D C					
Pongoabel	A Y I D - F A R Q K - - - L D - P K - I A - - - V A A Q N C Y K V T N G A F T G E I S P G M I K D Y					
chicken	I Y L D - F A R Q K - - - L D - A K - I G - - - V A A Q N C Y K V P K G A F T G E I S P A M I K D I					
Oryziasla	I Y L D - F V R S K - - - L D - P K - F G - - - V A A Q N C Y K V P K G A F T G E I S P A M I K D C					
Rattusnor	A Y I D - F A R Q K - - - L D - P K - I A - - - V A A Q N C Y K V T N G A F T G E I S P G M I K D L					
musmuscul	A Y I D - F A R Q K - - - L D - P K - I A - - - V A A Q N C Y K V T N G A F T G E I S P G M I K D L					
Xenopusla	I Y L D - F A R Q K - - - L D - A K - I A - - - I S A Q N C Y K V A K G A F T G E I S P A M I K D C					
Susscrofa	A Y I D - F A R Q K - - - L D - P K - I A - - - V A A Q N C Y K V A N G A F T G E I G P G M I K D L					
Bombyxmor	I Y L S - Y V K T I - - - I - P D N V E - - - V A A Q N C W K S P K G A F T G E I S P A M I K D V					
drosophil	I Y L M - Y A R N L - - - L - P C E I G - - - I A G Q N A Y K V A K G A F T G E I S P A M L K D I					
Culextars	P Y L T - I A R S Q - - - L - P D S V C - - - V A A Q N C Y K V P K G A F T G E I S P A M L K D L					
Taeniasol	C Y L K - Y A Q D K - - - A - P K G I K - - - I A A E N C Y K V G S G A F T G E I S T E M I K D C					
Schistos1	V F L H - E I R K S - - - L K - K E I H - - - V A A Q N C Y K V S K G A F T G E I S P A M I K D I					
Schistos2	V Y L N - E V R Q S - - - L A - K E I H - - - V A A Q N C Y K V P K G A F T G E I S P S M I K D V					
Candidaal	I Y L G - L A V E Q N - - - K Q - P T V A - - - I G A Q N V F D K S C G A F T G E I T C A S Q I L D V					
Debaryomy	P Y I A - L A V N E N - - - K Q - P T I E - - - V S A Q N C F D K A S G A Y T G E I S P E Q I K D L					
Coprinops	I Y L I - P I Q E I - - - A G - K A V K - - - V A A Q N A Y F K E S G A F T G E I S P K Q I S D A					
Caenorhab	P Y L A - Y A K S K - - - L K - A G V L - - - V A A Q N C Y K V P K G A F T G E I S P A M I K D L					
Plasmodiu	V H Y D - H T R K L - - - L Q - S K F S - - - T G I Q N V S K F G N G S Y T G E V S A E I A K D L					
Dictyoste	P Y L E - F L S N K - - - L D - N K K F K - - - V C S Q N C Y S V A K G A F T G E I S A N M L V D L					
Neurospor	I Y L D - F V K Q N - - - L Q K P N - V E - - - V A A Q N V F N K P N G A F T G E I S A T Q L L D L					
Aspergill	I Y L L - Q A R E V - - - A N - K E - I G - - - V A A Q N V F D K P N G A F T G E I S V Q Q L R E A					
Yarrowial	P Y L L - L A K E S - - - L K K P T - V S - - - V A G Q N S F D K G D G A F T G E I S V A Q L K D V					
Paracocci	I Y L V - L A R Q - - - L A N G Q - V A - - - V S A Q N V F D K P N G A F T G E I S V E Q L R D E					
Kluyverom	I Y L D - H A V A L - - - N K R - K E V K - - - I S A Q N A Y S K A S G A Y T G E N S V E Q I K D V					
Schizosac	M Y L I - T T R Q Q - - - V - K - K D I G - - - V G A Q N V F D K K N G A Y T G E N S A Q S L I D A					
Candidagl	T Y L D - Y T V S L - - - V S K - K Q V T - - - V G G Q N T Y T K A S G A Y T G E N S V D Q L K D V					
Saccharom	T Y L D - Y S V S L - - - V K K - P Q V T - - - V G A Q N A Y L K A S G A F T G E N S V D Q I K D V					
Ashbyagos	T Y L D - H A V S L - - - V S H - P Q V T - - - V G A Q N A Y T K A S G A Y T G E N S V D Q I K D V					
Giardiala	V H L S - T A I A A - - - N T S - K Q L R - - - I A A Q N V Y L E G N G A W T G E T S V E M L Q D M					
Entamoeba	I Y I P - K V Q Q I L A G E A N G A N - I L - - - V S A E N A W T K S - G A Y T G E V H V G M L V D C					
Lmexicana	V H I P - L V Q A K L - - - R N - P K - Y V - - - I S A E N A I A K - S G A F T G E V S M P I L K D I					
Trypanoso	L H I P - M T K A R L - - - T N - P K - F Q - - - I A A Q N A I T R - S G A F T G E V S L Q I L K D Y					
Zeamays	V F L P - V V K S Q - - - L R - Q E - F H - - - V A A Q N C W V K K G G A F T G E V S A E M L V N L					
Oryzasati	I Y I D - Q V K N S - - - L T - D R - I E - - - V S A Q N V W I G K G G A Y T G E I S A E Q L V D I					
Scereale	I Y I D - Q V K S S - - - L T - D R - I E - - - V S A Q N T W I G K G G A F T G E I S A E Q L V D I					
Coptisjap	V F L P - F V K N L - - - L R - A D - F H - - - V A A Q N C W V K K G G A F T G E V S A E M L V N L					
Phybrida	V F L P - L V K N E - - - L R - P D - F H - - - V A A Q N C W V K K G G A F T G E V S A E M L V N L					
Hordeumvu	V F L P - I V K A K - - - L R - P E - I Q - - - V A A Q N C W V K K G G A F T G E V S A E M L A N L					
Streptoc1	L F L E - G M K K G - - - V K - G T E L Q - - - V A A Q N C Y W E D F G A F T G E T S P A A L A A L					
Enterococ	L F L A - P L A W N - - - L K - D S E V K - - - I A A Q N C Y W E N A G A F T G E N S P A A I A D L					
Clostrid1	V C L P - G V V E A - - - A K - G S N I K - - - V A A Q N M H W E E K G A F T G E V S G P M L A E L					

	101	110	120	130	140	150																																													
HomoSapie	G	A	T	W	V	V	L	G	H	S	E	R	R	H	V	F	G	E	S	D	E	L	I	G	Q	K	V	A	H	A	L	A	E	G	L	G	V	I	A	C	I	G	E	K	L	D	E	R	E	A	
Macacamul	G	A	T	W	V	V	L	G	H	S	E	R	R	H	V	F	E	E	S	D	E	L	I	G	H	K	V	A	H	A	L	S	E	G	L	G	V	I	A	C	I	G	E	K	L	D	E	R	E	A	
bostaurus	G	A	T	W	V	V	L	G	H	S	E	R	R	H	V	F	G	E	S	D	E	L	I	G	Q	K	V	A	H	A	L	A	E	G	L	G	V	I	A	C	I	G	E	K	L	D	E	R	E	A	
canisfami	G	A	T	W	V	V	L	G	H	S	E	R	R	H	V	F	G	E	S	D	E	L	I	G	Q	K	V	A	H	A	L	A	E	G	L	G	V	I	A	C	I	G	E	K	L	D	E	R	E	A	
Danioreri	G	I	D	W	V	I	L	G	H	S	E	R	R	H	V	F	G	E	S	D	E	L	I	G	Q	K	V	A	H	C	L	E	S	D	L	G	V	I	A	C	I	G	E	K	L	E	E	R	E	A	
panthrolo	G	A	T	W	V	V	L	G	H	S	E	R	R	H	V	F	G	E	S	D	E	L	I	G	Q	K	V	A	H	A	L	A	E	G	L	G	V	I	A	C	I	G	E	K	L	D	E	R	E	A	
Pongoabel	G	A	T	W	V	V	L	G	H	S	E	R	R	H	V	F	G	E	S	D	E	L	I	G	Q	K	V	A	H	A	L	A	E	G	L	G	V	I	A	C	I	G	E	K	L	D	E	R	E	A	
chicken	G	A	A	W	V	I	L	G	H	S	E	R	R	H	V	F	G	E	S	D	E	L	I	G	Q	K	V	A	H	A	L	A	E	G	L	G	V	I	A	C	I	G	E	K	L	D	E	R	E	A	
Oryziasla	G	V	N	W	V	I	L	G	H	S	E	R	R	H	V	F	G	E	S	D	E	L	I	G	Q	K	T	A	H	A	L	E	S	G	L	G	V	I	A	C	I	G	E	K	L	D	E	R	E	G	
Rattusnor	G	A	T	W	V	V	L	G	H	S	E	R	R	H	I	F	G	E	S	D	E	L	I	G	Q	K	V	N	H	A	L	S	E	G	L	G	V	I	A	C	I	G	E	K	L	D	E	R	E	A	
musmuscul	G	A	T	W	V	V	L	G	H	S	E	R	R	H	V	F	G	E	S	D	E	L	I	G	Q	K	V	S	H	A	L	A	E	G	L	G	V	I	A	C	I	G	E	K	L	D	E	R	E	A	
Xenopusla	G	A	T	W	V	I	L	G	H	S	E	R	R	H	V	F	G	E	C	D	E	L	I	G	Q	K	V	A	H	A	L	S	E	G	L	G	V	I	A	C	I	G	E	K	L	D	Q	R	E	A	
Susscrofa	G	A	T	W	V	V	L	G	H	S	E	R	R	H	V	F	G	E	S	D	E	L	I	G	Q	K	V	A	H	A	L	A	E	G	L	G	V	I	A	C	I	G	E	K	L	D	E	R	E	A	
Bombyxmor	G	V	N	W	V	I	L	G	H	S	E	R	R	T	I	F	G	E	K	D	E	L	V	A	E	K	V	A	H	A	L	E	S	G	L	K	V	I	A	C	I	G	E	T	L	E	E	R	E	S	
drosophil	G	A	D	W	V	I	L	G	H	S	E	R	R	A	I	F	G	E	S	D	A	L	I	A	E	K	A	E	H	A	L	A	E	G	L	K	V	I	A	C	I	G	E	T	L	E	E	R	E	A	
Culextars	N	I	G	W	V	I	L	G	H	S	E	R	R	A	I	F	G	E	S	D	E	L	I	A	D	K	V	A	H	A	L	A	E	G	L	K	V	I	A	C	I	G	E	T	L	Q	E	R	E	A	
Taeniasol	G	C	E	W	V	I	L	G	H	S	E	R	R	H	I	F	G	E	S	N	E	L	I	G	E	K	V	K	H	A	L	D	S	G	L	N	V	I	P	C	I	G	E	L	L	S	E	R	E	A	
Schistos1	G	C	D	W	V	I	L	G	H	S	E	R	R	N	I	F	G	E	S	D	E	L	I	A	E	K	V	Q	H	A	L	A	E	G	L	S	V	I	A	C	I	G	E	T	L	S	E	R	E	S	
Schistos2	G	C	D	W	V	I	L	G	H	S	E	R	R	S	I	F	N	E	S	D	E	L	V	A	E	K	V	Q	H	A	L	A	E	G	L	S	V	V	A	C	I	G	E	T	L	S	E	R	E	S	
Candidaom	G	A	S	W	T	L	T	G	H	S	E	R	R	T	I	K	E	S	D	E	F	I	A	E	K	T	K	F	A	L	D	T	G	V	K	V	I	L	C	I	G	E	T	L	E	E	R	K	G		
Debaryomy	G	A	T	W	T	L	T	G	H	S	E	R	R	T	I	K	E	S	D	D	F	I	A	S	K	T	K	F	A	L	D	Q	G	L	S	V	I	L	C	I	G	E	T	L	E	E	R	K	A		
Coprinops	G	I	P	Y	V	I	L	G	H	S	E	R	R	T	L	F	H	E	T	S	E	V	V	A	L	K	T	R	A	A	L	D	N	G	L	K	V	I	L	C	I	G	E	T	L	K	E	R	E	E	
Caenorhab	G	I	E	W	V	I	L	G	H	S	E	R	R	H	V	F	G	E	S	D	A	L	I	A	E	K	T	V	H	A	L	A	E	A	G	I	K	V	V	F	C	I	G	E	K	L	E	E	R	E	A
Plasmodiu	N	I	E	Y	V	I	I	G	H	F	E	R	R	K	Y	F	H	E	T	D	E	D	V	R	E	K	L	Q	A	S	L	K	N	N	L	K	A	V	V	C	F	G	E	S	L	E	Q	R	E	Q	
Dictyoste	G	I	P	Y	V	I	L	G	H	S	E	R	R	N	V	F	S	E	S	E	S	E	L	I	T	K	K	T	K	Y	A	I	S	L	G	L	T	V	I	L	C	I	G	E	L	L	A	D	R	K	S
Neurospor	G	V	K	W	V	I	L	G	H	S	E	R	R	N	E	L	G	E	S	D	E	F	I	A	S	K	T	K	Y	A	L	D	N	G	I	S	V	I	W	C	C	G	E	S	K	D	T	R	Q	A	
Aspergill	N	I	D	W	T	I	L	G	H	S	E	R	R	V	I	L	K	E	T	D	E	F	I	A	R	K	T	K	A	A	I	E	G	L	Q	V	I	F	C	I	G	E	T	L	E	E	R	E	A		
Yarrowial	G	A	K	W	V	I	L	G	H	S	E	R	R	T	I	N	K	E	S	S	E	W	I	A	D	K	T	K	Y	A	L	D	N	G	L	D	V	I	L	C	I	G	E	T	L	D	E	K	K	A	
Paracocci	K	I	T	W	T	L	A	G	H	S	E	R	R	V	L	L	R	E	D	D	E	F	V	A	R	K	T	K	A	A	L	N	G	L	N	V	I	L	C	I	G	E	S	L	E	E	R	E	A		
Kluyverom	G	A	E	W	V	I	L	G	H	S	E	R	R	T	Y	F	N	E	T	D	E	I	V	A	E	K	T	K	F	A	L	D	S	G	V	K	V	I	L	C	I	G	E	T	L	E	E	K	Q	Q	
Schizosac	G	I	T	Y	T	L	T	G	H	S	E	R	R	T	I	F	K	E	S	D	E	F	V	A	D	K	T	K	F	A	L	E	Q	G	L	T	V	V	A	C	I	G	E	T	L	A	E	R	E	A	
Candidagl	G	A	K	W	V	I	L	G	H	S	E	R	R	T	Y	F	H	E	D	D	K	I	V	A	E	K	T	K	F	A	L	D	Q	G	L	G	V	I	L	C	I	G	E	T	L	E	E	K	K	A	
Saccharom	G	A	K	W	V	I	L	G	H	S	E	R	R	S	Y	F	H	E	D	D	K	F	I	A	D	K	T	K	F	A	L	G	Q	G	V	G	V	I	L	C	I	G	E	T	L	E	E	K	K	A	
Ashbyagos	G	A	K	W	V	I	L	G	H	S	E	R	R	T	Y	F	N	E	D	D	E	Q	V	A	E	K	T	A	F	A	L	E	R	G	V	S	V	I	L	C	I	G	E	T	L	D	E	K	K	A	
Giardiala	G	I	K	H	V	I	V	G	H	S	E	R	R	R	I	M	G	E	T	D	E	Q	S	A	K	K	A	K	R	A	L	E	K	G	M	T	V	I	F	C	V	G	E	T	L	D	E	R	K	A	
Entamoeba	Q	V	P	Y	V	I	L	G	H	S	E	R	R	Q	I	F	H	E	S	N	E	Q	V	A	E	K	V	K	V	A	I	D	A	G	L	K	V	I	A	C	I	G	E	T	E	A	Q	R	I	A	
Lmexicana	G	V	H	W	V	I	L	G	H	S	E	R	R	T	Y	Y	G	E	T	D	E	I	V	A	Q	K	V	S	E	A	C	K	Q	G	F	M	V	I	A	C	I	G	E	T	L	Q	Q	R	E	A	
Trypanoso	G	I	K	W	V	V	L	G	H	S	E	R	R	L	Y	Y	G	E	T	N	E	I	V	A	E	K	V	A	Q	A	C	A	A	G	F	H	V	I	V	C	V	G	E	T	N	E	E	R	E	A	
Zeamays	G	V	P	W	V	I	L	G	H	S	E	R	R	A	L	L	G	E	S	N	E	F	V	G	D	K	V	A	Y	A	L	S	Q	G	L	K	V	I	A	C	V	G	E	T	L	E	Q	R	E	A	
Oryzasati	G	C	Q	W	V	I	L	G	H	S	E	R	R	H	V	I	G	E	D	D	Q	F	I	G	K	K	A	A	Y	A	L	S	Q	N	V	K	V	I	A	C	I	G	E	L	L	E	E	R	E	A	
Scereale	G	C	Q	W	V	I	L	G	H	S	E	R	R	H	V	I	G	E	D	D	E	F	I	G	K	K	A	A	Y	A	L	S	Q	N	L	K	V	M	A	C	I	G	E	L	L	E	E	R	E	A	
Coptisjap	G	I	P	W	V	I	L	G	H	S	E	R	R	A	L	L	N	E	S	N	E	F	V	G	D	K	T	A	Y	A	L	S	Q	G	L	K	V	I	A	C	V	G	E	T	L	E	Q	R	E	A	
Phybrida	S	I	P	W	V	I	L	G	H	S	E	R	R	A	L	L	G	E	S	N	E	F	V	G	D	K	V	A	Y	A	L	S	Q	G	L	K	V	I	A	C	V	G	E	T	L	E	E	R	E	S	
Hydrdeumvu	G	V	P	W	V	I	L	G	H	S	E	R	R	S	L	G	E	S	S	E	F	V	G	E	K	V	A	Y	A	L	A	Q	G	L	K	V	I	A	C	V	G	E	T	L	E	Q	R	E	A		
Streptoc1	G	V	D	Y	V	I	I	G	H	S	E	R	R	D	Y	F	H	E	T	D	Q	E	I	N	K	K	A	H	A	I	F	K	H	K	M	T	P	I	L	C	C	G	E	S	L	E	T	Y	E	A	
Enterococ	G	V	D	Y	V	I	I	G	H	S	E	R	R	E	Y	F	H	E	T	D	E	D	I	N	K	K	A	K	A	I	F	A	N	G	M	T	P	I	F	C	C	G	E	T	L	E	T	Y	E	A	
Clostrid1	G	V	D	Y	V	I	I	G	H	S																																									

	151	160	170	180	190	200
HomoSapie	G I T E K V V F E Q	- T K V I A D N -	V - - K D - -	W S K V V L A Y E P V W A I	G T G K T A T P Q Q	
Macacamul	G I T E K V V F Q Q	- T K V I A D N -	V - - K D - -	W S K A V L A Y E P M W A I	G T G K T A T P Q Q	
bostraurus	G I T E K V V F E Q	- T K V I A D N -	V - - K D - -	W S K V V L A Y E P V W A I	G T G K T A T P Q Q	
canisfami	G I T E K V V F E Q	- T K V I A D N -	V - - K D - -	W S K V V L A Y E P V W A I	G T G K T A T P Q Q	
Danioreri	G T T E D V V F E Q	- T K V I A D N -	V - - K D - -	W T R V V L A Y E P V W A I	G T G K T A S P E Q	
pantroгло	G I T E K V V F E Q	- T K V I A D N -	V - - K D - -	W S K V V L A Y E P V W A I	G T G K T A T P Q Q	
Pongoabel	G I T E K V V F E Q	- T K V I A D N -	V - - K D - -	W S K V V L A Y E P V W A I	G T G K T A T P Q Q	
chicken	G I T E K V V F E Q	- T K A I A D N -	V - - K D - -	W S K V V L A Y E P V W A I	G T G K T A T P Q Q	
Oryziasla	G I T E K V V F S Q	- T K F I A D N -	V - - K D - -	W S K V V L A Y E P V W A I	G T G K T A S P Q Q	
Rattusnor	G I T E K V V F E Q	- T K A I A D N -	V - - K D - -	W C K V V L A Y E P V W A I	G T G K T A T P Q Q	
musmuscul	G I T E K V V F E Q	- T K V I A D N -	V - - K D - -	W S K V V L A Y E P V W A I	G T G K T A T P Q Q	
Xenopusla	G I T E K V V F E Q	- T K A I A D N -	V - - K D - -	W S K V V L A Y E P V W A I	G T G K T A T P E Q	
Susscrofa	G I T E K V V F E Q	- T K V I A D N -	V - - K D - -	W N K V V L A Y E P V W A I	G T G K T A T P Q Q	
Bombyxmor	G K T E E V V F R Q	- L K A I V S A -	I - - G D K -	W E N I V L A Y E P V W A I	G T G K T A T P Q Q	
drosophil	G K T N E V V A R Q	- M C A Y A Q K -	I - - K D - -	W K N V V V A Y E P V W A I	G T G Q T A T P D Q	
Culextars	G Q T E A V C F R Q	- T K A I A D K -	V - - K D - -	W S N V V I A Y E P V W A I	G T G K T A S P E Q	
Taeniasol	G K T N D V C F A Q	- M D A I A K N -	V P S K E A -	W D K V V I A Y E P V W A I	G T G K T A T P A Q	
Schistos1	N K T E E V C V R Q	- L K A I A N K -	I K S A D E -	W K R V V V A Y E P V W A I	G T G K V A T P Q Q	
Schistos2	G K T E E V C V R Q	- L N A I A N K -	I K S I D E -	W K R V V V A Y E P V W A I	G T G K V A S P G Q	
Candidaal	G V T L D V C A R Q	- L D A V S K I -	V - - S D - -	W S N I V V A Y E P V W A I	G T G L A A T P E D	
Debaryomy	N V T L D V C A R Q	- L D A V A R V -	V - - S D - -	W S K I V V A Y E P V W A I	G T G L A A T P E D	
Coprinops	G R T A A V C E E Q	- L S A V V K Q -	I - - K E E D -	W S N I V I A Y E P V W A I	G T G K V A T T S Q	
Caenorhab	G H T K D V N F R Q	- L Q A I V D K G -	V - - S - - -	W E N I V I A Y E P V W A I	G T G K T A S G E Q	
Plasmodiu	N K T I E V I T K Q	- V K A F V D -	I - - D N - -	F D N V I L A Y E P L W A I	G T G K T A T P E Q	
Dictyoste	N N T E H I L S E Q	- L K D F A S F -	T - - P E E -	W S K I V I A Y E P V W A I	G T G A V A T P Q E	
Neurospor	G E T I K F V E N Q	- L A A L A K E -	I - - N D - -	W K N V V I A Y E P I W A I	G T G L V A T K E Q	
Aspergill	N K T I D V V T R Q	- L N A A A K E -	L S - K E Q -	W A K V V I A Y E P V W A I	G T G K V A T T E Q	
Yarrowial	G K T L D V V R S Q	- L D P V I A K -	I - - K D - -	W S N V V I A Y E P V W A I	G T G L A A T A E D	
Paracocci	G K T I D V V T R Q	- L D A V A E E -	V S P A E -	W N K V V I A Y E P I W A I	G T G K V A T T E Q	
Kluyverom	N I T L Q V V Q R Q	- L Q A V L E K -	V - - Q D - -	W T N V V V A Y E P V W A I	G T G L A A T A E D	
Schizosac	N E T I N V V V R Q	- L N A I A D K -	V - - Q N - -	W S K I V I A Y E P V W A I	G T G K T A T P E Q	
Candidagl	G V T L K V V E R Q	- L D A V I A E -	V - - K D - -	W T N V V I A Y E P V W A I	G T G L A A T A E D	
Saccharom	G K T L D V V E R Q	- L N A V L E E -	V - - K D - -	W T N V V V A Y E P V W A I	G T G L A A T P E D	
Ashbyagos	G V T L D V V K R Q	- L T P V L E K -	V - - K D - -	W T N V V I A Y E P V W A I	G T G L A A T A E D	
Giardia	N R T M E V N I A Q	- L E A L G K E L G E S	K M L - W K E V V I	A Y E P V W S I	G T G V V A T P E Q	
Entamoeba	N Q T E E V V A A Q	- L K A I N N A -	I S - K E A -	W K N I I L A Y E P V W A I	G T G K T A T P D Q	
Lmexicana	N Q T A K V V L S Q	- T S A I A A K -	L T - K D -	A W N Q V V L A Y E P V W A I	G T G K V A T P E Q	
Trypanoso	G R T A A V V L T Q	- L A A V A Q K -	L S - K E A -	W A H V V I A Y E P V W A I	G T G K V A T P Q Q	
Zeamays	G S T M D V V A A Q	- T K A I A E K -	I - - K D - -	W S N V V V A Y E P V W A I	G T G K V A T P A Q	
Oryzasati	G K T F D V C F K Q	- M K A F A D S -	I - - T N - -	W A D V V I A Y E P V W A I	G T G K V A T P E Q	
Scereale	G K T F D V C F K Q	- M K A F A D N -	I - - T D - -	W T N V V I A Y E P V W A I	G T G K V A S P E Q	
Coptisjap	G S T I S V V A A Q	- T K A I A E K -	V - - S D - -	W T N I V V A Y E P V W A I	G T G K V A S P A Q	
Phybrida	G S T M D V V A A Q	- T K A I A D R -	V - - K D - -	W T N V V V A Y E P V W A I	G T G K V A S P A Q	
Hordeumvu	G S T M E V V A E Q	- T K A I A G K -	I - - K D - -	W S N G V V A Y E P V W A I	G T G K V A T P A Q	
Streptoc1	G K T A E W I E G Q I	- T A D L K G L S A E Q	- - - V S S M V I	A Y E P I W A I	G T G K S A D A N I	
Enterococ	G K T A E W I E G Q I	- T K G L V G L S N E Q	- - - V A S M V I	A Y E P I W A I	G T G K S A D A N I	
Clostrid1	G V T A E L V R Y Q V	- K I A L L G L S A E Q	- - - - V K E A V I	A Y E P I W A I	G T G K T A T N E Q	

	201	210	220	230	240	250																																													
HomoSapie	A	Q	E	V	H	E	K	L	R	G	W	L	K	S	N	V	S	D	A	V	A	Q	S	T	R	I	I	Y	G	G	S	V	T	G	A	T	C	K	E	L	A	S	Q	P	D	V	D	G	F	L	
Macacamu	A	Q	E	V	H	E	K	F	R	G	W	L	E	S	N	V	S	E	A	L	A	Q	S	T	G	I	I	Y	G	G	S	V	T	G	A	T	Y	K	E	L	A	S	Q	P	D	V	D	G	F	L	
bostaurus	A	Q	E	V	H	E	K	L	R	G	W	L	K	S	N	V	S	D	A	V	A	Q	S	A	R	I	I	Y	G	G	S	V	T	G	A	T	C	K	E	L	A	S	Q	P	D	V	D	G	F	L	
canisfami	A	Q	E	V	H	E	K	L	R	G	W	L	K	S	N	V	S	D	A	V	A	Q	S	T	R	I	I	Y	G	G	S	V	T	G	A	T	C	K	E	L	A	S	Q	P	D	V	D	G	F	L	
Danioreri	A	Q	E	V	H	E	K	L	R	E	W	L	R	A	N	V	S	D	A	V	A	D	S	V	R	I	I	Y	G	G	S	V	T	G	G	N	C	K	E	L	A	A	Q	A	D	V	D	G	F	L	
pantrogl	A	Q	E	V	H	E	K	L	R	G	W	L	K	S	N	V	S	D	A	V	A	Q	S	T	R	I	I	Y	G	G	S	V	T	G	A	T	C	K	E	L	A	S	Q	P	D	V	D	G	F	L	
Pongoabel	A	Q	E	V	H	E	K	L	R	G	W	L	K	S	N	V	S	D	A	V	A	Q	S	T	R	I	I	Y	G	G	S	V	T	G	A	T	C	K	E	L	A	S	Q	P	D	V	D	G	F	L	
chicken	A	Q	E	V	H	E	K	L	R	G	W	L	K	S	H	V	S	D	A	V	A	Q	S	T	R	I	I	Y	G	G	S	V	T	G	G	N	C	K	E	L	A	S	Q	H	D	V	D	G	F	L	
Oryziasla	A	Q	E	V	H	E	K	L	R	E	W	L	K	S	N	V	S	E	A	V	A	N	S	V	R	I	I	Y	G	G	S	V	T	G	G	T	C	K	E	L	A	S	Q	K	D	V	-	-	-	-	
Rattusnor	A	Q	E	V	H	E	K	L	R	G	W	L	K	C	N	V	S	E	G	V	A	Q	C	T	R	I	I	Y	G	G	S	V	T	G	A	T	C	K	E	L	A	S	Q	P	D	V	D	G	F	L	
musmuscul	A	Q	E	V	H	E	K	L	R	G	W	L	K	S	N	V	S	N	D	G	V	A	Q	S	T	R	I	I	Y	G	G	S	V	T	G	A	T	C	K	E	L	A	S	Q	P	D	V	D	G	F	L
Xenopusla	A	Q	E	V	H	K	L	R	E	W	V	K	T	N	V	S	E	G	V	A	Q	S	V	R	I	I	Y	G	G	S	V	T	G	G	T	C	R	E	L	A	G	P	D	I	D	G	F	L			
Susscrofa	A	Q	E	V	H	E	K	L	R	G	W	L	K	T	H	V	P	E	A	V	A	H	S	T	R	I	I	Y	G	G	S	V	T	G	A	T	C	K	E	L	A	S	Q	P	D	V	D	G	F	R	
Bombyxm	A	Q	D	V	H	H	A	L	R	N	W	L	S	A	N	V	S	G	S	V	S	D	A	V	R	I	Q	Y	G	G	S	V	T	A	A	N	A	K	E	L	A	S	C	K	D	I	D	G	F	L	
drosophil	A	Q	E	V	H	A	F	L	R	Q	W	L	S	D	N	I	S	K	E	V	S	A	S	R	I	Q	Y	G	G	S	V	T	A	A	N	A	K	E	L	A	K	K	P	D	I	D	G	F	L		
Culextars	A	Q	E	V	H	A	A	L	R	K	W	F	T	E	N	V	S	A	D	V	S	A	A	R	I	Q	Y	G	G	S	V	T	A	A	N	C	R	E	L	A	A	K	P	D	I	D	G	F	L		
Taeniasol	A	Q	E	V	H	K	V	V	R	D	W	I	R	K	H	V	D	A	G	I	A	D	K	V	R	I	L	Y	G	G	S	V	T	A	S	N	A	K	D	L	G	T	Q	P	D	V	D	G	F	L	
Schistos1	A	Q	E	V	H	N	F	L	R	K	W	F	K	T	N	A	P	N	G	V	D	E	K	I	R	I	I	Y	G	G	S	V	T	A	A	N	C	K	E	L	A	Q	Q	H	D	V	D	G	F	L	
Schistos2	A	Q	E	V	H	N	F	L	R	K	W	F	K	T	N	T	P	S	G	V	D	Q	Q	I	R	I	I	Y	G	G	S	V	T	A	A	N	C	K	E	L	A	Q	Q	H	D	V	D	G	F	L	
Candidaal	A	E	E	T	H	K	G	I	R	A	H	L	A	K	S	I	G	A	E	Q	A	E	K	T	R	I	L	Y	G	G	S	V	N	G	K	N	A	K	D	F	K	D	K	A	N	V	D	G	F	L	
Debaryomy	A	Q	E	T	H	K	A	I	R	E	H	L	S	K	S	I	G	A	D	A	A	E	K	T	R	I	L	Y	G	G	S	V	N	G	K	N	A	P	E	F	K	D	K	A	D	V	D	G	F	L	
Coprinops	A	Q	E	T	H	V	D	V	R	K	Y	L	A	T	A	V	S	P	K	V	A	S	E	T	R	V	I	Y	G	G	S	V	N	A	A	N	S	K	D	L	A	S	Q	Q	D	I	D	G	F	L	
Caenorhab	A	Q	E	V	H	E	W	I	R	A	F	L	K	E	K	V	S	P	A	V	A	D	A	T	R	I	I	Y	G	G	S	V	T	A	D	N	A	A	E	L	G	K	K	P	D	I	D	G	F	L	
Plasmodiu	A	Q	L	V	H	K	E	I	R	K	I	V	K	D	T	C	G	E	K	Q	A	N	Q	I	R	I	L	Y	G	G	S	V	N	T	E	N	C	S	S	L	I	Q	Q	E	D	I	D	G	F	L	
Dictyoste	A	Q	D	T	H	V	F	I	R	K	W	I	S	E	K	V	S	E	D	V	A	K	K	T	S	I	M	Y	G	G	S	V	N	A	D	N	C	H	N	L	S	K	Q	S	D	I	D	G	F	L	
Neurospor	A	Q	E	V	H	A	A	I	R	S	W	L	K	Q	N	V	S	D	K	V	A	E	E	T	R	I	L	Y	G	G	S	V	N	A	K	N	C	K	D	L	A	K	E	Q	D	I	D	G	F	L	
Aspergill	A	Q	E	V	H	S	A	I	R	K	W	L	K	D	A	I	S	A	E	A	E	N	T	R	I	I	Y	G	G	S	V	S	E	K	N	C	K	E	L	A	K	E	A	D	I	D	G	F	L		
Yarrowial	A	Q	Q	I	H	H	E	I	R	A	Y	L	K	D	K	I	G	A	Q	-	A	D	K	V	R	I	I	Y	G	G	S	V	N	G	K	N	S	G	T	F	K	D	K	S	D	V	D	G	F	L	
Paracocci	A	Q	E	V	H	A	S	I	R	K	W	L	N	E	K	I	S	P	E	A	A	E	N	T	R	V	I	Y	G	G	S	V	T	E	S	N	C	R	D	L	A	A	Q	P	D	V	D	G	F	L	
Kluyverom	A	Q	D	I	H	H	S	I	R	E	F	L	A	E	K	L	S	R	D	V	A	D	S	V	R	I	L	Y	G	G	S	A	N	G	K	N	A	V	T	F	K	D	K	A	D	V	D	G	F	L	
Schizosac	A	Q	E	V	H	A	E	I	R	K	W	A	T	N	K	L	G	A	S	V	A	E	G	R	V	I	Y	G	G	S	V	N	G	G	N	C	K	E	F	L	K	F	H	D	I	D	G	F	L		
Candidagl	A	Q	D	I	H	H	S	I	R	S	Y	L	A	E	K	L	G	K	D	T	A	E	K	T	R	I	L	Y	G	G	S	A	N	G	A	N	A	G	T	F	K	D	K	A	D	V	D	G	F	L	
Saccharom	A	Q	D	I	H	A	S	I	R	K	F	L	A	S	K	L	G	D	K	A	A	S	E	L	R	I	L	Y	G	G	S	A	N	G	S	N	A	V	T	F	K	D	K	A	D	V	D	G	F	L	
Ashbyagos	A	Q	D	I	H	A	A	I	R	D	F	L	A	E	R	T	S	R	D	V	A	D	A	V	R	I	L	F	G	G	S	A	N	G	A	N	A	S	F	R	D	K	A	D	V	D	G	F	L		
Giardiala	A	E	E	V	H	V	G	I	R	K	W	F	A	E	K	V	C	A	E	G	A	Q	H	I	R	I	Y	G	G	S	V	A	N	G	S	N	C	E	K	L	G	Q	C	P	N	I	D	G	F	L	
Entamoeba	A	Q	E	V	H	Q	Y	I	R	K	W	M	T	E	N	I	S	K	E	V	A	E	A	T	R	I	Q	Y	G	G	S	V	N	P	A	N	C	N	E	L	A	K	K	A	D	I	D	G	F	L	
Lmexicana	A	Q	E	V	H	L	L	L	R	K	W	V	S	E	N	I	G	T	D	V	A	A	K	L	R	I	L	Y	G	G	S	V	N	A	A	N	A	A	T	L	Y	A	K	P	D	I	N	G	F	L	
Trypanoso	A	Q	E	V	H	E	L	L	R	R	W	V	R	S	K	L	G	A	D	I	A	A	Q	L	R	I	L	Y	G	G	S	V	T	A	K	N	A	R	T	L	Y	Q	M	R	D	I	N	G	F	L	
Zeamays	A	Q	E	V	H	A	S	L	R	D	W	L	K	T	N	A	S	P	E	V	A	E	S	T	R	I	I	Y	G	G	S	V	T	A	A	N	C	K	E	L	A	A	Q	P	D	V	D	G	F	L	
Oryzasati	A	Q	E	V	H	A	A	V	R	D	W	L	K	T	N	V	S	P	E	V	A	S	G	I	R	I	I	Y	G	G	S	V	N	A	A	N	C	A	E	L	A	K	K	E	D	I	D	G	F	L	
Scereale	A	Q	E	V	H	A	A	V	R	D	W	L	K	T	N	V	S	A	D	V	A	S	T	V	R	I	I	Y	G	G	S	V	N	A	A	N	C	A	E	L	A	K	K	E	D	I	D	G	F	L	
Coptisjap	A	Q	E	V	H	F	E	L	R	K	W	I	K	E	N	V	G	A	D	V	A	G	S	V	R	I	I	Y	G	G	S	V	N	G	A	N	S	K	E	L	A	G	Q	P	D	I	D	G	F	L	
Phybrida	A	Q	E	V	H	A	E	L	R	K	W	L	A	A	N	V	S	P	E	V	A	A	S	T	R	I	I	Y	G	G	S	V	N	G	A	N	C	K	E	L	G	Q	P	D	V	D	G	F	L		
Hordeumvu	A	Q	E	V	H	A	N	L	R	D	W	L	K	T	N	V	S	P	E	V	A	E	S	T	R	I	I	Y	G	G	S	V	T	G	A	S	C	K	E	L	A	A	Q	A	D	V	D	G	F	L	
Streptoc1	A	D	D	I	C	G	V	V	R	A	T	V	E	K	L	Y	G	K	E	V	A	Q	A	V	R	I	Q	Y	G	G	S	V	K	P	E	N	V	A	E	Y	M	A	K	E	N	V	D	G	A	L	
Enterococ	A	D	E	I	C	G	V	V	R	S	T	V	E	K	L	Y	G	K	E	V	S	E	A	V	R	I	Q	Y	G	G	S	V	K	P	E	N	I	A	E	Y	M	A	K	E	N	V	D	G	A	L	
Clostrid1																																																			

Arqueas



Procariontes

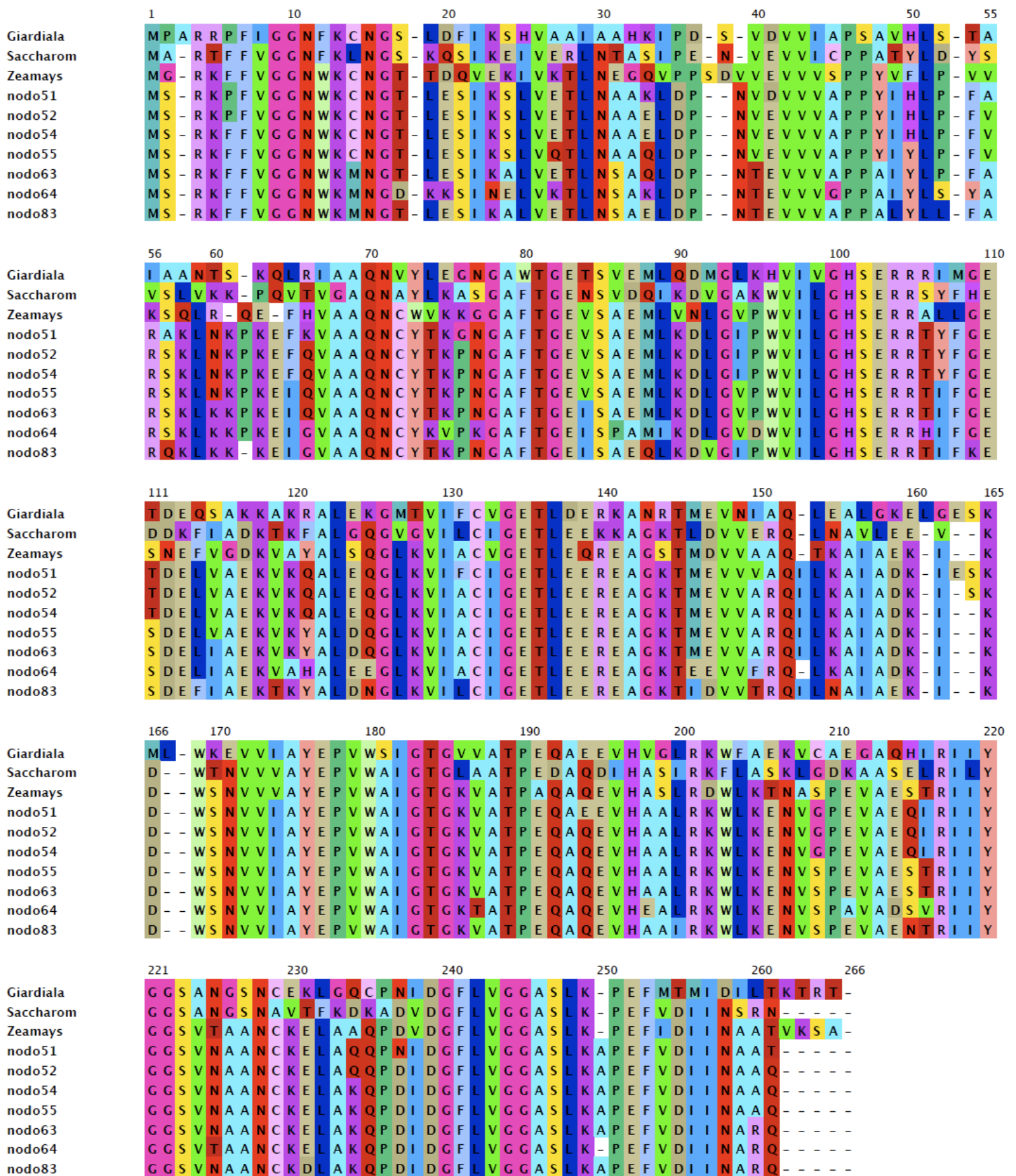
	1	10	20	30	40	50	60	70	80	85																																																																					
Nitrosom2	M	---	G	R	A	G	F	V	A	---	G	---	N	W	K	M	H	G	S	V	R	N	Q	E	---	V	S	C	V	V	C	V	P	---	Y	P	I	A	O	T	O	S	V	---	E	---	G	S	H	V	---	W	G	G																									
Bacteroi1	M	---	R	K	N	I	V	A	---	G	---	N	W	K	M	M	K	T	---	G	E	G	I	A	---	A	K	E	---	N	E	A	---	A	N	E	K	P	N	---	C	D	V	I	C	T	P	---	F	I	H	---	A	S	V	T	P	---	V	D	---	A	A	K	I	---	V	G	A										
Chlorobi1	M	---	R	K	K	I	V	V	---	G	---	N	W	K	M	M	K	T	---	V	A	A	E	---	V	A	A	---	E	G	F	---	T	G	---	C	D	V	G	I	A	P	P	---	F	T	A	---	Y	E	V	G	K	---	T	---	G	S	R	A	---	I	V	A															
Chlorobi2	M	---	---	V	V	---	G	---	N	W	K	M	M	N	I	A	E	S	V	---	A	A	I	---	A	E	---	K	V	---	A	D	---	G	V	O	---	C	E	V	G	I	A	P	P	---	F	P	A	---	Y	E	V	C	H	---	V	E	---	W	S	G	I	---	R	C	A												
Ecoli	M	---	R	H	P	V	M	---	G	---	N	W	K	N	G	S	R	H	M	V	---	H	E	---	V	S	N	---	R	K	E	---	A	G	V	A	---	C	A	V	A	I	A	P	P	---	E	M	---	I	D	M	A	---	K	R	E	A	E	---	E	---	G	S	H	I	---	M	G	A									
Eferguson	M	---	R	H	P	V	M	---	G	---	N	W	K	N	G	S	R	H	M	V	---	H	E	---	V	S	N	---	R	K	E	---	A	G	V	A	---	C	A	V	A	I	A	P	P	---	E	M	---	I	D	M	A	---	K	R	E	A	E	---	E	---	G	S	H	I	---	M	G	A									
Haemoph2	M	---	A	R	R	P	V	M	---	G	---	N	W	K	N	G	S	K	A	F	---	T	O	---	I	A	E	---	K	T	E	---	A	E	I	---	T	E	G	---	C	D	V	A	I	A	P	P	---	V	M	---	S	E	A	E	I	A	---	A	G	---	G	S	T	I	---	R	G	A									
Pasteurel	M	---	A	R	R	P	V	M	---	G	---	N	W	K	N	G	S	K	A	F	---	T	O	---	I	A	E	---	K	T	E	---	A	E	I	---	T	E	G	---	C	D	V	A	I	A	P	P	---	V	M	---	S	E	A	E	I	A	---	A	G	---	K	C	S	C	S	C	G	---	G	K	A	I	---	G	A		
Haemoph1	M	---	A	R	R	P	V	M	---	G	---	N	W	K	N	G	S	K	A	F	---	T	O	---	I	A	E	---	K	T	E	---	A	E	I	---	T	E	G	---	C	D	V	A	I	A	P	P	---	V	M	---	S	E	A	E	I	A	---	A	G	---	S	G	C	G	C	S	C	G	---	G	K	S	V	I	---	G	A
Photorhab	M	---	R	H	P	V	M	---	G	---	N	W	K	N	G	S	T	H	M	V	---	N	E	---	I	---	E	G	---	R	G	E	---	S	S	V	T	---	C	D	V	A	I	A	P	P	---	A	---	Y	S	O	A	K	A	---	A	---	G	S	R	I	---	A	G	A													
Ralstoni2	V	---	S	A	R	P	K	V	V	---	G	---	N	W	K	H	G	S	N	G	---	N	A	E	---	E	K	I	---	K	A	A	---	G	O	T	R	---	A	A	---	A	V	C	A	P	---	F	P	---	A	O	C	C	S	---	A	---	G	S	V	V	---	A	G	A													
Salmonele	M	---	R	H	P	V	M	---	G	---	N	W	K	N	G	S	R	H	M	V	---	H	E	---	V	S	N	---	R	K	E	---	A	G	V	A	---	C	A	V	A	I	A	P	P	---	E	M	---	I	D	M	A	---	K	R	E	A	E	---	E	---	G	S	H	I	---	M	G	A									
Shigella	M	---	R	H	P	V	M	---	G	---	N	W	K	N	G	S	R	H	M	V	---	H	E	---	V	S	N	---	R	K	E	---	A	G	V	A	---	C	A	V	A	I	A	P	P	---	E	M	---	I	D	M	A	---	K	R	E	A	E	---	E	---	G	S	H	I	---	M	G	A									
Salmoneit	M	---	R	H	P	V	M	---	G	---	N	W	K	N	G	S	R	H	M	V	---	H	E	---	V	S	N	---	R	K	E	---	A	G	V	A	---	C	A	V	A	I	A	P	P	---	E	M	---	I	D	M	A	---	K	R	E	A	E	---	E	---	G	S	H	I	---	M	G	A									
Pseudomo3	M	---	R	R	P	V	A	---	G	---	N	W	K	M	H	G	T	S	V	---	S	V	A	E	---	I	K	---	R	O	A	---	F	S	G	---	V	D	V	A	M	P	P	---	C	---	F	I	S	O	V	I	---	G	A	---	G	K	A	I	---	D	V	G	A														
Pseudomo6	M	---	R	R	P	V	A	---	G	---	N	W	K	M	H	G	T	S	V	---	S	V	A	E	---	I	K	---	R	O	A	---	F	S	G	---	V	D	V	A	M	P	P	---	C	---	F	I	S	O	V	I	---	G	A	---	G	K	A	I	---	D	V	G	A														
Pseudomo2	M	---	R	R	P	V	A	---	G	---	N	W	K	M	H	G	T	S	V	---	S	V	A	E	---	I	K	---	R	O	A	---	F	S	G	---	V	D	V	A	M	P	P	---	C	---	F	I	S	O	V	I	---	G	A	---	G	K	A	I	---	D	V	G	A														
Xanthomo2	M	---	R	R	K	I	V	A	---	G	---	N	W	K	H	G	S	R	A	F	---	A	T	---	E	---	V	A	---	K	A	E	---	H	D	V	T	---	C	D	V	A	I	A	P	P	---	V	M	---	S	E	A	E	I	A	---	A	G	---	K	S	V	I	---	G	A												
Xanthomo3	M	---	R	R	K	I	V	A	---	G	---	N	W	K	H	G	S	R	A	F	---	A	T	---	E	---	V	A	---	K	A	E	---	H	D	V	T	---	C	D	V	A	I	A	P	P	---	V	M	---	S	E	A	E	I	A	---	A	G	---	K	S	V	I	---	G	A												
Shewanell	M	---	A	R	R	P	V	A	---	G	---	N	W	K	M	N	G	S	A	A	---	A	Q	E	---	F	---	K	F	A	S	---	K	D	S	---	A	E	V	V	I	C	P	P	---	S	I	---	Y	E	S	V	R	---	O	E	A	N	K	E	A	---	D	---	G	S	V	R	M	---	G	A							
Xylella	M	---	R	H	P	V	M	---	G	---	N	W	K	N	G	S	H	A	F	A	---	A	---	V	A	---	A	G	---	P	---	P	---	F	G	---	V	S	V	I	---	I	---	P	---	S	---	D	---	T	---	R	F	K	---	A	---	G	E	G	---	A	F	A															
Yersinia1	M	---	R	H	P	V	M	---	G	---	N	W	K	N	G	S	T	H	M	V	---	N	E	---	I	---	A	---	G	---	R	K	E	---	A	G	V	A	---	C	A	V	A	I	A	P	P	---	E	M	---	I	D	M	A	---	K	R	E	A	E	---	E	---	G	S	H	I	---	M	G	A							
Bordetel1	M	T	---	A	E	N	---	R	A	R	V	---	G	---	N	W	K	M	H	G	---	A	E	N	---	A	E	---	N	A	A	---	T	A	E	---	R	A	A	A	A	---	C	E	M	G	V	C	V	P	---	F	P	---	A	O	---	A	A	A	---	O	---	G	S	A	I	---	G	A									
Coxiellab	M	---	R	R	P	V	A	---	G	---	N	W	K	M	H	G	S	R	E	S	V	---	G	---	I	---	R	A	---	K	H	G	---	C	E	R	---	E	T	---	A	E	---	A	V	F	P	P	---	F	V	---	F	O	---	C	E	E	A	---	M	---	R	T	I	---	S	W	A										
Streptoc1	M	---	S	R	K	P	I	A	---	G	---	N	W	K	M	M	K	T	---	A	A	E	A	---	R	E	F	---	D	A	---	K	N	---	P	S	N	---	V	D	V	I	---	G	S	P	---	A	---	F	E	G	M	---	K	G	V	K	---	G	T	E	---	V	A	A	---	G	A										
bacillant	M	---	R	K	P	I	V	A	---	G	---	N	W	K	M	M	K	T	---	S	E	A	V	---	S	---	E	---	A	E	---	V	K	---	P	A	S	A	---	V	D	A	V	I	A	P	P	---	V	M	---	S	E	A	E	I	A	---	A	G	---	K	S	V	I	---	G	A											
bacillalb	M	---	R	K	P	I	V	A	---	G	---	N	W	K	M	M	K	T	---	V	G	E	A	---	K	A	---	E	V	---	K	G	A	---	P	S	S	D	---	V	D	V	I	A	P	P	---	V	M	---	S	E	A	E	I	A	---	A	G	---	K	S	V	I	---	G	A												
Porphyrom	M	---	R	K	N	I	V	A	---	G	---	N	W	K	M	M	K	T	---	O	E	G	---	A	---	A	---	A	---	K	A	E	---	D	A	A	---	K	G	T	I	---	N	---	C	D	V	I	G	T	P	---	F	I	H	---	A	S	I	---	A	A	A	---	D	---	T	T	R	I	---	G	V	A					
Streptoc2	M	---	S	R	K	P	I	A	---	G	---	N	W	K	M	M	K	T	---	O	E	A	---	K	A	---	F	---	E	A	V	---	A	S	K	---	P	S	D	---	V	D	V	A	V	A	A	P	---	A	V	D	---	V	T	I	---	E	A	A	---	K	---	D	S	V	---	K	V	A	---	A	G	E	I	---	G	A	
Staphyau	M	---	R	T	P	I	A	---	G	---	N	W	K	M	M	K	T	---	V	E	A	---	K	D	F	---	V	N	A	---	P	---	P	---	D	S	K	---	E	---	V	E	S	V	I	C	A	P	---	A	I	---	D	A	---	T	---	T	A	V	---	K	E	G	---	K	A	G	E	I	---	G	A						
Staphyep	M	---	R	T	P	I	A	---	G	---	N	W	K	M	M	K	T	---	V	E	A	---	K	D	F	---	V	N	A	---	P	---	P	---	D	S	K	---	E	---	V	E	S	V	I	C	A	P	---	A	I	---	D	A	---	T	---	T	A	V	---	K	E	G	---	K	A	G	E	I	---	G	A						
aquifex	M	---	R	R	---	I	A	---	A	---	N	W	K	M	M	K	T	---	V	A	---	K	E	E	---	E	E	---	I	N	T	---	F	K	V	---	E	H	P	E	S	---	R	E	I	---	I	C	P	P	---	F	---	T	S	---	Y	V	A	---	G	K	Q	---	G	T	G	V	K	---	G	A							
listein	M	---	R	K	P	I	V	A	---	G	---	N	W	K	M	M	K	T	---	A	A	K	A	---	G	A	---	F	A	E	---	D	---	K	N	---	V	P	S	D	---	V	E	S	V	A	A	P	---	A	---	F	---	Q	E	---	V	R	T	---	E	---	G	T	N	---	R	V	A	---	A	G	---	G	A				
listemon	M	---	R	K	P	I	V	A	---	G	---	N	W	K	M	M	K	T	---	A	A	K	A	---	G	A	---	F	A	E	---	D	---	K	N	---	V	P	S	D	---	V	E	S	V	A	A	P	---	A	---	F	---	Q	E	---	V	R	T	---	E	---	G	T	N	---	R												

	86	90	100	110	120	130	140	150	160	170			
Nitrosom2	QNV	SRYHHGAY	TEVSADM	ADG7RYVY	VHSERR	TFGE	GNQVVAE	KFRAAERS	IT-PI	CVGET	AARE	SDETE	VIAMQ
Bacteroi2	ENC	ADKASGAY	TEVSADM	ASGAKYVI	VHSERR	AYYGE	TVAI	EKKVKA	ANGT	PIFCJ	GEVLE	EERANK	ONEVVAQ
Chlorob1	QNC	HFPDGAFT	TEVSASMV	REAGCSAVI	VHSERR	CFGET	NAIVNR	VKKHASEG	CVIM	CAGET	EERE	GGVT	GOVVT
Chlorob2	QNC	HYESDGF	TEVSARM	AAAGCSYVI	VHSERR	CFGET	NAIVNR	VKKHASEG	CVIM	CAGET	EERE	GGVT	GOVVT
Ecoli	QNV	DINLSGAF	TEVSAAM	KDIGAYI	VHSERR	YHKS	ESDE	IAKKFVAV	KEGG	T-PV	CI	GETE	AEAGKTE
Eferugson	QNV	DINLSGAF	TEVSAAM	KDIGAYI	VHSERR	YHKS	ESDE	IAKKFVAV	KEGG	T-PV	CI	GETE	AEAGKTE
Haemoph2	QNV	DINTGGAF	TEVSTAM	EFCAKYI	VHSERR	YHKS	ESDE	IAKKFVAV	KEGG	T-PV	CI	GETE	AEAGKTE
Pasteurel	QNV	DVNVGGAF	TEVSTEM	KDFGAKYI	VHSERR	YHKS	ESDE	IAKKFVAV	KEGG	T-PV	CI	GETE	AEAGKTE
Haemoph1	QNV	DINLVGAF	TEVSTEM	KDFGAKYI	VHSERR	YHKS	ESDE	IAKKFVAV	KEGG	T-PV	CI	GETE	AEAGKTE
Photorhab	QDV	DVNLGAF	TEVSAAM	KDIGAYI	VHSERR	YHKS	ESDE	IAKKFVAV	KEGG	T-PV	CI	GETE	AEAGKTE
Raistoni2	QDV	SAEARGAF	TEVSAAM	SEFGCYAI	VHSERR	YHKS	ESDE	IAKKFVAV	KEGG	T-PV	CI	GETE	AEAGKTE
Salmonele	QNV	DINLSGAF	TEVSAAM	KDIGAYI	VHSERR	YHKS	ESDE	IAKKFVAV	KEGG	T-PV	CI	GETE	AEAGKTE
Shigella	QNV	DINLSGAF	TEVSAAM	KDIGAYI	VHSERR	YHKS	ESDE	IAKKFVAV	KEGG	T-PV	CI	GETE	AEAGKTE
Salmonelt	QNV	DINLSGAF	TEVSAAM	KDIGAYI	VHSERR	YHKS	ESDE	IAKKFVAV	KEGG	T-PV	CI	GETE	AEAGKTE
Pseudom3	QNS	AVPEMGA	TEVSAAM	ADG7RYVY	VHSERR	TFGE	GNQVVAE	KFRAAERS	IT-PI	CVGET	AARE	SDETE	VIAMQ
Pseudom6	QNS	AVPEMGA	TEVSAAM	ADG7RYVY	VHSERR	TFGE	GNQVVAE	KFRAAERS	IT-PI	CVGET	AARE	SDETE	VIAMQ
Pseudom2	QDAAI	AAEGAF	TEVSAAM	ADG7RYVY	VHSERR	TFGE	GNQVVAE	KFRAAERS	IT-PI	CVGET	AARE	SDETE	VIAMQ
Xanthom2	QDV	SSNEKAY	TEVSAM	VDYGAY	VHSERR	YHKS	ESDE	IAKKFVAV	KEGG	T-PV	CI	GETE	AEAGKTE
Xanthom3	QDV	SSNEKAY	TEVSAM	VDYGAY	VHSERR	YHKS	ESDE	IAKKFVAV	KEGG	T-PV	CI	GETE	AEAGKTE
Shewane1	QNV	SHDFGAY	TEVSGM	KDCGCRYI	VHSERR	RYHES	TSNVAE	KFRAAERS	IT-PI	CVGET	AARE	SDETE	VIAMQ
Ytiellaf	QNV	SHDFGAY	TEVSGM	KDCGCRYI	VHSERR	RYHES	TSNVAE	KFRAAERS	IT-PI	CVGET	AARE	SDETE	VIAMQ
Xersinia1	QNV	DVNLGAF	TEVSAAM	KDIGAYI	VHSERR	YHKS	ESDE	IAKKFVAV	KEGG	T-PV	CI	GETE	AEAGKTE
Bordetel	QDV	SAHAKGAY	TEVAAPM	AEFGCRWV	VHSERR	YHKS	ESDE	IAKKFVAV	KEGG	T-PV	CI	GETE	AEAGKTE
Coxiellaf	QDV	SEFERGAY	TEVSAAM	RDFGCRYI	VHSERR	YHKS	ESDE	IAKKFVAV	KEGG	T-PV	CI	GETE	AEAGKTE
Streptoc1	QNC	YWEFDGAF	TEVSAAM	ADG7RYVY	VHSERR	TFGE	GNQVVAE	KFRAAERS	IT-PI	CVGET	AARE	SDETE	VIAMQ
bacilant	QNM	HFEKNGAF	TEVSAAM	ADG7RYVY	VHSERR	TFGE	GNQVVAE	KFRAAERS	IT-PI	CVGET	AARE	SDETE	VIAMQ
bacilhal	QNM	HFEKNGAF	TEVSAAM	ADG7RYVY	VHSERR	TFGE	GNQVVAE	KFRAAERS	IT-PI	CVGET	AARE	SDETE	VIAMQ
Porphyrom	ENC	ADNLSGAY	TEVSAAM	ASGARYVI	VHSERR	AYYHE	TSPMEK	VKAASNG	IT-PI	CVGET	AARE	SDETE	VIAMQ
Streptoc2	QNC	YFENTGAF	TEVSAAM	ADG7RYVY	VHSERR	TFGE	GNQVVAE	KFRAAERS	IT-PI	CVGET	AARE	SDETE	VIAMQ
Staphyau	QNC	YFEDNGAF	TEVSAAM	ADG7RYVY	VHSERR	TFGE	GNQVVAE	KFRAAERS	IT-PI	CVGET	AARE	SDETE	VIAMQ
Staphyep	QNA	YFEDNGAF	TEVSAAM	ADG7RYVY	VHSERR	TFGE	GNQVVAE	KFRAAERS	IT-PI	CVGET	AARE	SDETE	VIAMQ
aquifex	QNC	HYEKRGAF	TEVSAAM	ADG7RYVY	VHSERR	TFGE	GNQVVAE	KFRAAERS	IT-PI	CVGET	AARE	SDETE	VIAMQ
listein	QNC	YFEDNGAF	TEVSAAM	ADG7RYVY	VHSERR	TFGE	GNQVVAE	KFRAAERS	IT-PI	CVGET	AARE	SDETE	VIAMQ
listemon	QNC	YFEDNGAF	TEVSAAM	ADG7RYVY	VHSERR	TFGE	GNQVVAE	KFRAAERS	IT-PI	CVGET	AARE	SDETE	VIAMQ
Streptoc5	QNC	YFEDNGAF	TEVSAAM	ADG7RYVY	VHSERR	TFGE	GNQVVAE	KFRAAERS	IT-PI	CVGET	AARE	SDETE	VIAMQ
Vibrioch	QNT	DHNSGAY	TEVSAAM	ADG7RYVY	VHSERR	TFGE	GNQVVAE	KFRAAERS	IT-PI	CVGET	AARE	SDETE	VIAMQ
Vibriopar	QNT	DHNSGAY	TEVSAAM	ADG7RYVY	VHSERR	TFGE	GNQVVAE	KFRAAERS	IT-PI	CVGET	AARE	SDETE	VIAMQ
Vibriovl	QNT	DHNSGAY	TEVSAAM	ADG7RYVY	VHSERR	TFGE	GNQVVAE	KFRAAERS	IT-PI	CVGET	AARE	SDETE	VIAMQ
Enterococ	QNC	YWEFDGAF	TEVSAAM	ADG7RYVY	VHSERR	TFGE	GNQVVAE	KFRAAERS	IT-PI	CVGET	AARE	SDETE	VIAMQ
bradyrh	QDC	HAKASGAY	TEVSAAM	ADG7RYVY	VHSERR	TFGE	GNQVVAE	KFRAAERS	IT-PI	CVGET	AARE	SDETE	VIAMQ
Agrobacte	QDC	HAKASGAY	TEVSAAM	ADG7RYVY	VHSERR	TFGE	GNQVVAE	KFRAAERS	IT-PI	CVGET	AARE	SDETE	VIAMQ
Rhodospa	QDC	HAKASGAY	TEVSAAM	ADG7RYVY	VHSERR	TFGE	GNQVVAE	KFRAAERS	IT-PI	CVGET	AARE	SDETE	VIAMQ
photobap	QNS	AVPEMGA	TEVSAAM	ADG7RYVY	VHSERR	TFGE	GNQVVAE	KFRAAERS	IT-PI	CVGET	AARE	SDETE	VIAMQ
neisseri	QDV	SARFNDGAY	TEVSAAM	ADG7RYVY	VHSERR	TFGE	GNQVVAE	KFRAAERS	IT-PI	CVGET	AARE	SDETE	VIAMQ
Caulobact	QDC	HAKASGAY	TEVSAAM	ADG7RYVY	VHSERR	TFGE	GNQVVAE	KFRAAERS	IT-PI	CVGET	AARE	SDETE	VIAMQ
deinococ	QDV	SAHAKGAY	TEVSAAM	ADG7RYVY	VHSERR	TFGE	GNQVVAE	KFRAAERS	IT-PI	CVGET	AARE	SDETE	VIAMQ
Lactobac5	ENA	YFEDNGAF	TEVSAAM	ADG7RYVY	VHSERR	TFGE	GNQVVAE	KFRAAERS	IT-PI	CVGET	AARE	SDETE	VIAMQ
Lactobac8	ENA	YFEDNGAF	TEVSAAM	ADG7RYVY	VHSERR	TFGE	GNQVVAE	KFRAAERS	IT-PI	CVGET	AARE	SDETE	VIAMQ
Borreliab	QNM	SYMESGAR	TEVSAAM	ADG7RYVY	VHSERR	TFGE	GNQVVAE	KFRAAERS	IT-PI	CVGET	AARE	SDETE	VIAMQ
bifidoba1	QSV	SVTGGAF	TEVSAAM	ADG7RYVY	VHSERR	TFGE	GNQVVAE	KFRAAERS	IT-PI	CVGET	AARE	SDETE	VIAMQ
bifidoba2	QAV	SVTGGAF	TEVSAAM	ADG7RYVY	VHSERR	TFGE	GNQVVAE	KFRAAERS	IT-PI	CVGET	AARE	SDETE	VIAMQ
Dehaloco1	QNV	FYEKNGAY	TEVSAAM	ADG7RYVY	VHSERR	TFGE	GNQVVAE	KFRAAERS	IT-PI	CVGET	AARE	SDETE	VIAMQ
desulfovi	QDV	YPAEAGAY	TEVSAAM	ADG7RYVY	VHSERR	TFGE	GNQVVAE	KFRAAERS	IT-PI	CVGET	AARE	SDETE	VIAMQ
oceanobih	QNM	HFEKNGAF	TEVSAAM	ADG7RYVY	VHSERR	TFGE	GNQVVAE	KFRAAERS	IT-PI	CVGET	AARE	SDETE	VIAMQ
prochlor	QNV	HWEDDGAY	TEVSAAM	ADG7RYVY	VHSERR	TFGE	GNQVVAE	KFRAAERS	IT-PI	CVGET	AARE	SDETE	VIAMQ
synchel	QNI	HWEKNGAY	TEVSAAM	ADG7RYVY	VHSERR	TFGE	GNQVVAE	KFRAAERS	IT-PI	CVGET	AARE	SDETE	VIAMQ
Nostocpun	QNI	HWEKNGAY	TEVSAAM	ADG7RYVY	VHSERR	TFGE	GNQVVAE	KFRAAERS	IT-PI	CVGET	AARE	SDETE	VIAMQ
Synechocy	QNV	HWEKNGAY	TEVSAAM	ADG7RYVY	VHSERR	TFGE	GNQVVAE	KFRAAERS	IT-PI	CVGET	AARE	SDETE	VIAMQ
Gloeobact	QNV	HWEKNGAY	TEVSAAM	ADG7RYVY	VHSERR	TFGE	GNQVVAE	KFRAAERS	IT-PI	CVGET	AARE	SDETE	VIAMQ
Campylob1	QNF	YPCVNGAF	TEVSAAM	ADG7RYVY	VHSERR	TFGE	GNQVVAE	KFRAAERS	IT-PI	CVGET	AARE	SDETE	VIAMQ
Wolinella	QNA	YPAEAGAY	TEVSAAM	ADG7RYVY	VHSERR	TFGE	GNQVVAE	KFRAAERS	IT-PI	CVGET	AARE	SDETE	VIAMQ
helicob2	QNA	YPAEAGAY	TEVSAAM	ADG7RYVY	VHSERR	TFGE	GNQVVAE	KFRAAERS	IT-PI	CVGET	AARE	SDETE	VIAMQ
Helicob1	QNA	YPRDNGAF	TEVSAAM	ADG7RYVY	VHSERR	TFGE	GNQVVAE	KFRAAERS	IT-PI	CVGET	AARE	SDETE	VIAMQ
Leptospi1	QNV	YPSGAAFT	TEVSAAM	ADG7RYVY	VHSERR	TFGE	GNQVVAE	KFRAAERS	IT-PI	CVGET	AARE	SDETE	VIAMQ
Leptospi2	QNC	YHSGAAFT	TEVSAAM	ADG7RYVY	VHSERR	TFGE	GNQVVAE	KFRAAERS	IT-PI	CVGET	AARE	SDETE	VIAMQ
Leptospi3	QNA	YHSGAAFT	TEVSAAM	ADG7RYVY	VHSERR	TFGE	GNQVVAE	KFRAAERS	IT-PI	CVGET	AARE	SDETE	VIAMQ
thermstir	QDV	SAHAKGAY	TEVSAAM	ADG7RYVY	VHSERR	TFGE	GNQVVAE	KFRAAERS	IT-PI	CVGET	AARE	SDETE	VIAMQ
Lactococ	ENS	YFEDNGAF	TEVSAAM	ADG7RYVY	VHSERR	TFGE	GNQVVAE	KFRAAERS	IT-PI	CVGET	AARE	SDETE	VIAMQ
fusobanu	ENV	YPKIEGAY	TEVSAAM	ADG7RYVY	VHSERR	TFGE	GNQVVAE	KFRAAERS	IT-PI	CVGET	AARE	SDETE	VIAMQ
Streptom1	QDV	SAHAKGAY	TEVSAAM	ADG7RYVY	VHSERR	TFGE	GNQVVAE	KFRAAERS	IT-PI	CVGET	AARE	SDETE	VIAMQ
Streptom2	QDI	SAHAKGAY	TEVSAAM	ADG7RYVY	VHSERR	TFGE	GNQVVAE	KFRAAERS	IT-PI	CVGET	AARE	SDETE	VIAMQ
corynedi	QDI	SAHAKGAY	TEVSAAM	ADG7RYVY	VHSERR	TFGE	GNQVVAE	KFRAAERS	IT-PI	CVGET	AARE	SDETE	VIAMQ
corynel	QDV	SAHAKGAY	TEVSAAM	ADG7RYVY	VHSERR	TFGE	GNQVVAE	KFRAAERS	IT-PI	CVGET	AARE	SDETE	VIAMQ
Mycobctub	QDI	SAHAKGAY	TEVSAAM	ADG7RYVY	VHSERR	TFGE	GNQVVAE	KFRAAERS	IT-PI	CVGET	AARE	SDETE	VIAMQ
Acidobca3	QNI	HAKASGAY	TEVSAAM	ADG7RYVY	VHSERR	TFGE	GNQVVAE	KFRAAERS	IT-PI	CVGET	AARE	SDETE	VIAMQ
Clostrid3	QNA	HFEKNGAF	TEVSAAM	ADG7RYVY	VHSERR	TFGE	GNQVVAE	KFRAAERS	IT-PI	CVGET	AARE	SDETE	VIAMQ
Clostrid5	QNM	YFEDNGAF	TEVSAAM	ADG7RYVY	VHSERR	TFGE	GNQVVAE	KFRAAERS	IT-PI	CVGET	AARE	SDETE	VIAMQ
Clostrid6	QNM	HFEKNGAF	TEVSAAM	ADG7RYVY	VHSERR	TFGE	GNQVVAE	KFRAAERS	IT-PI	CVGET	AARE	SDETE	VIAMQ
Geobactel	QDC	YWEKNGAF	TEVSAAM	ADG7RYVY	VHSERR	TFGE	GNQVVAE	KFRAAERS	IT-PI	CVGET	AARE	SDETE	VIAMQ
Solibaus	QNV	HAKASGAY	TEVSAAM	ADG7RYVY	VHSERR	TFGE	GNQVVAE	KFRAAERS	IT-PI	CVGET	AARE	SDETE	VIAMQ
Thermoan3	QNV	FWEKNGAY	TEVSAAM	ADG7RYVY	VHSERR	TFGE	GNQVVAE	KFRAAERS	IT-PI	CVGET	AARE	SDETE	VIAMQ
Chlamyru	QNT	HEVDSGAF	TEVSAAM	ADG7RYVY	VHSERR	TFGE	GNQVVAE	KFRAAERS	IT-PI	CVGET	AARE	SDETE	VIAMQ
Chlamytr	QTT	HEADSGAF	TEVSAAM	ADG7RYVY	VHSERR	TFGE	GNQVVAE	KFRAAERS	IT-PI	CVGET	AARE	SDETE	VIAMQ
Treponem1	QNM	GEEKNGAY	TEVSAAM	ADG7RYVY	VHSERR	TFGE	GNQVVAE	KFRAAERS	IT-PI	CVGET	AARE	SDETE	VIAMQ
Treponem2	QDV	SSAEWGAH	TEVSAAM	ADG7RYVY	VHSERR	TFGE	GNQVVAE	KFRAAERS	IT-PI	CVGET	AARE	SDETE	VIAMQ
Wiggleswo	QNV	DVNLGAF	TEVSAAM	ADG7RYVY	VHSERR	TFGE	GNQVVAE	KFRAAERS	IT-PI	CVGET	AARE	SDETE	VIAMQ
Wobachi1	QNC	HKKFGS	TEVSAAM	ADG7RYVY	VHSERR	TFGE	GNQVVAE	KFRAAERS	IT-PI	CVGET	AARE	SDETE	VIAMQ
Chromobac	QDV	SARFNDGAY	TEVSAAM	ADG7RYVY	VHSERR	TFGE	GNQVVAE	KFRAAERS	IT-PI	CVGET	AARE	SDETE	VIAMQ
Mesorhiz	QNM	HAKASGAY	TEVSAAM	ADG7RYVY	VHSERR	TFGE	GNQVVAE	KFRAAERS	IT-PI	CVGET	AARE	SDETE	VIAMQ
brucelme	QNM	HAKASGAY	TEVSAAM	ADG7RYVY	VHSERR	TFGE	GNQVVAE	KFRAAERS	IT-PI	CVGET	AARE	SDETE	VIAMQ
Bdellovi	QNC	YFEDNGAF	TEVSAAM	ADG7RYVY	VHSERR	TFGE	GNQVVAE	KFRAAERS	IT-PI	CVGET	AARE	SDETE	VIAMQ
fibrobac	QDI	HEADSGAF	TEVSAAM	ADG7RYVY	VHSERR	TFGE	GNQVVAE	KFRAAERS	IT-PI	CVGET	AARE	SDETE	VIAMQ
Onionyle	QNM	FYEKNGAY	TEVSAAM	ADG7RYVY	VHSERR	TFGE	GNQVVAE	KFRAAERS	IT-PI	CVGET	AARE	SDETE	VIAMQ
Blochflo	QNV	DHNSGAY	TEVSAAM	ADG7RYVY	VHSERR	TFGE	GNQVVAE	KFRAAERS	IT-PI	CVGET	AARE	SDETE	VIAMQ
Chlamydp	QNV	HFEKNGAF	TEVSAAM	ADG7RYVY	VHSERR	TFGE	GNQVVAE	KFRAAERS	IT-PI	CVGET	AARE	SDETE	VIAMQ

	171	180	190	200	210	220	230	240	250	255																																																																								
Nitrosom2	D	A	V	I	D	A	G	I	E	A	--	G	S	V	I	A	E	P	V	W	A	I	G	T	G	K	T	A	P	E	A	Q	D	V	H	K	F	I	R	S	R	I	A	V	H	S	--	G	G	I	A	E	N	I	G	--	I	Y	G	G	S	V	K	A	D	N	A	R	E	F	T	M										
Bacterol2	M	E	S	--	V	F	S	S	A	E	D	--	F	S	K	I	I	A	E	P	V	W	A	I	G	T	G	K	T	A	S	E	F	E	A	Q	E	I	H	A	F	I	R	S	I	V	A	D	K	Y	--	G	K	E	I	A	D	M	T	S	--	I	Y	G	G	S	C	K	P	S	N	A	K	E	F	S	N					
Chlorob1	V	R	E	G	--	L	G	I	D	D	--	I	S	N	I	V	I	A	E	P	V	W	A	I	G	T	G	K	T	A	S	E	F	E	A	Q	E	V	H	I	I	R	N	I	V	S	D	F	--	G	E	E	A	A	G	R	R	--	I	Y	G	G	S	K	P	S	N	A	K	E	F	S	M									
Chlorob2	V	V	E	G	--	I	D	V	D	--	I	S	K	V	I	A	E	P	V	W	A	I	G	T	G	K	T	A	K	E	G	A	--	E	V	H	A	I	R	A	K	V	E	E	Y	--	G	K	A	A	D	H	R	--	I	Y	G	G	S	K	P	S	N	A	A	E	F	A	M													
Ecoli	I	D	A	V	--	K	T	G	G	A	A	--	F	E	G	A	V	I	A	E	P	V	W	A	I	G	T	G	K	S	A	P	A	G	A	V	H	K	F	I	R	D	H	I	A	K	V	D	--	A	N	--	I	A	E	Q	V	I	--	I	Y	G	G	S	V	N	A	S	A	A	E	F	A	Q								
Eferguson	I	D	A	V	--	K	T	G	G	A	A	--	F	E	G	A	V	I	A	E	P	V	W	A	I	G	T	G	K	S	A	P	A	G	A	V	H	K	F	I	R	D	H	I	A	K	V	D	--	A	N	--	I	A	E	Q	V	I	--	I	Y	G	G	S	V	N	A	S	A	A	E	F	A	Q								
Haemoph2	I	D	A	V	--	I	N	S	G	V	E	A	--	F	N	G	A	V	I	A	E	P	V	W	A	I	G	T	G	K	S	A	P	A	G	A	V	H	K	F	I	R	D	H	I	A	K	V	D	--	A	N	--	I	A	E	Q	V	I	--	I	Y	G	G	S	V	N	D	T	N	A	A	E	F	S	Q						
Pasteurel	I	D	A	V	--	N	T	G	V	E	A	--	F	N	G	A	V	I	A	E	P	V	W	A	I	G	T	G	K	S	A	P	A	G	A	V	H	K	F	I	R	D	H	I	A	K	V	D	--	A	N	--	I	A	E	Q	V	I	--	I	Y	G	G	S	V	N	D	N	A	A	E	F	T	Q								
Haemoph1	I	D	A	V	--	I	N	A	G	V	E	A	--	F	N	G	A	V	I	A	E	P	V	W	A	I	G	T	G	K	S	A	P	A	G	A	V	H	K	F	I	R	D	H	I	A	K	V	D	--	A	N	--	I	A	E	Q	V	I	--	I	Y	G	G	S	V	N	D	N	A	A	E	F	T	Q							
Phototrab	I	D	A	V	--	N	T	G	V	E	A	--	F	N	G	A	V	I	A	E	P	V	W	A	I	G	T	G	K	S	A	P	A	G	A	V	H	K	F	I	R	D	H	I	A	K	V	D	--	A	N	--	I	A	E	Q	V	I	--	I	Y	G	G	S	V	N	D	N	A	A	E	F	S	Q								
Ralstoni2	I	D	A	V	--	G	A	V	E	A	S	V	E	--	G	G	I	V	V	A	E	P	V	W	A	I	G	T	G	K	S	A	P	A	G	A	V	H	K	F	I	R	D	H	I	A	K	V	D	--	A	N	--	I	A	E	Q	V	I	--	I	Y	G	G	S	V	K	P	D	N	A	A	E	F	S	Q						
Salmonele	I	D	T	V	--	K	T	G	G	A	A	--	F	E	G	A	V	I	A	E	P	V	W	A	I	G	T	G	K	S	A	P	A	G	A	V	H	K	F	I	R	D	H	I	A	K	V	D	--	A	N	--	I	A	E	Q	V	I	--	I	Y	G	G	S	V	N	A	S	A	A	E	F	A	Q								
Shigella	I	D	A	V	--	K	T	G	G	A	A	--	F	E	G	A	V	I	A	E	P	V	W	A	I	G	T	G	K	S	A	P	A	G	A	V	H	K	F	I	R	D	H	I	A	K	V	D	--	A	N	--	I	A	E	Q	V	I	--	I	Y	G	G	S	V	N	A	S	A	A	E	F	A	Q								
Salmonelt	I	D	A	V	--	K	T	G	G	A	A	--	F	E	G	A	V	I	A	E	P	V	W	A	I	G	T	G	K	S	A	P	A	G	A	V	H	K	F	I	R	D	H	I	A	K	V	D	--	A	N	--	I	A	E	Q	V	I	--	I	Y	G	G	S	V	N	A	S	A	A	E	F	A	Q								
Pseudomo3	I	D	A	V	--	G	S	V	I	D	E	G	V	G	A	--	F	A	R	V	V	A	E	P	V	W	A	I	G	T	G	K	T	A	P	E	A	Q	E	V	H	A	I	R	S	I	V	A	D	K	Y	--	A	E	E	N	--	A	E	--	V	A	K	G	V	R	--	I	Y	G	G	S	V	K	A	S	A	A	E	F	G	M
Pseudomo6	I	D	A	V	--	I	E	A	F	G	V	K	A	--	F	A	D	A	V	I	A	E	P	V	W	A	I	G	T	G	L	T	A	P	E	A	Q	D	V	H	A	I	R	S	I	V	A	D	K	Y	--	A	E	E	N	--	A	E	--	V	A	K	G	V	R	--	I	Y	G	G	S	V	K	A	A	N	A	A	E	F	G	M
Pseudomo2	I	D	A	V	--	I	A	E	F	G	V	K	A	--	F	A	D	A	V	I	A	E	P	V	W	A	I	G	T	G	L	T	A	P	E	A	Q	D	V	H	A	I	R	S	I	V	A	D	K	Y	--	A	E	E	N	--	A	E	--	V	A	K	G	V	R	--	I	Y	G	G	S	V	K	A	A	N	A	A	E	F	S	M
Xanthomo2	I	D	A	V	--	D	P	V	A	G	S	A	G	--	F	A	G	A	V	I	A	E	P	V	W	A	I	G	T	G	R	T	A	P	E	A	Q	A	V	H	K	F	I	R	D	H	I	A	K	V	D	--	A	R	--	I	A	D	S	--	P	I	Y	G	G	S	V	K	P	D	N	A	A	E	F	A	Q					
Xanthomo3	I	D	A	V	--	E	P	V	S	V	G	S	T	G	--	F	A	R	V	V	A	E	P	V	W	A	I	G	T	G	R	T	A	P	E	A	Q	A	V	H	K	F	I	R	D	H	I	A	K	V	D	--	A	R	--	I	A	D	S	--	P	I	Y	G	G	S	V	K	P	D	N	A	A	E	F	S	Q					
Shewanel	I	D	A	V	--	I	D	V	K	N	G	T	M	A	--	F	D	N	A	I	A	E	P	V	W	A	I	G	T	G	K	S	A	P	E	A	Q	A	V	H	K	F	I	R	D	H	I	A	K	V	D	--	A	R	--	I	A	D	S	--	P	I	Y	G	G	S	V	T	P	S	N	A	A	E	F	A	Q					
Xylella	I	D	A	V	--	A	P	V	A	V	P	G	G	--	V	A	R	G	I	A	E	P	V	W	A	I	G	T	G	R	H	A	D	S	Q	V	A	M	H	A	F	I	R	D	H	I	A	K	V	D	--	A	R	--	I	A	D	S	--	P	I	Y	G	G	G	I	K	P	C	N	A	A	E	F	S	Q						
Yersinia1	I	D	A	V	--	N	T	G	V	K	A	--	F	E	G	A	V	I	A	E	P	V	W	A	I	G	T	G	K	S	A	P	A	G	A	V	H	K	F	I	R	D	H	I	A	K	V	D	--	A	N	--	I	A	E	Q	V	I	--	I	Y	G	G	S	V	N	D	K	N	A	A	E	F	T	Q							
Bordetel1	I	D	A	V	--	E	P	V	A	L	--	G	R	D	A	--	V	R	M	V	A	E	P	V	W	A	I	G	T	G	R	T	A	P	E	A	Q	E	V	H	A	I	R	S	I	V	A	D	K	Y	--	A	S	V	R	--	V	I	Y	G	G	S	V	K	G	A	N	A	A	E	F	A	Q									
Coxiellab	I	D	A	V	--	A	V	V	E	G	M	N	D	N	A	S	--	E	G	M	V	A	E	P	V	W	A	I	G	T	G	K	N	A	P	S	Q	A	E	V	H	A	I	R	S	I	V	A	D	K	Y	--	A	T	--	A	E	S	T	R	--	I	Y	G	G	S	V	K	P	D	N	A	A	A	E	F	M					
Streptoc1	I	T	A	D	--	K	G	L	S	A	E	D	--	V	S	M	V	I	A	E	P	V	W	A	I	G	T	G	K	S	A	D	A	I	A	D	D	I	C	G	V	V	A	T	V	E	K	Y	--	G	K	E	V	A	A	V	R	--	I	Y	G	G	S	V	K	P	E	N	A	E	Y	M	A	K								
bacilant	I	T	K	A	--	A	G	L	T	E	E	Q	--	V	K	A	T	V	I	A	E	P	V	W	A	I	G	T	G	K	S	S	S	A	D	A	N	E	V	C	A	H	I	R	K	V	V	A	E	V	--	S	P	A	A	A	E	A	V	R	--	I	Y	G	G	S	V	K	P	E	N	I	K	E	Y	M	A					
bacifal	I	T	K	A	--	E	G	T	D	E	Q	--	V	K	V	T	V	I	A	E	P	V	W	A	I	G	T	G	K	S	S	A	D	A	N	E	V	C	S	I	R	K	V	V	E	K	Y	--	S	G	E	A	A	D	A	V	R	--	I	Y	G	G	S	V	K	P	N	A	E	Y	M	A										
Porphyrom	I	T	K	A	--	E	F	T	D	Q	T	--	F	A	K	I	A	E	P	V	W	A	I	G	T	G	K	T	A	A	D	A	Q	E	M	H	A	I	R	S	I	V	A	D	K	Y	--	G	K	E	V	A	C	S	--	I	Y	G	G	S	V	C	N	A	A	N	A	K	E	F	S	R										
Streptoc2	I	T	K	A	--	V	S	A	A	A	G	S	A	E	D	--	V	A	S	V	I	A	E	P	V	W	A	I	G	T	G	K	S	A	D	D	A	N	M	C	A	V	R	D	V	V	A	A	D	F	--	G	G	E	V	A	D	V	R	--	V	Q	Y	G	G	S	V	K	P	E	N	I	K	E	Y	M	A					
Staphyau	I	T	K	A	--	V	K	K	A	V	A	G	S	E	D	--	K	S	V	V	I	A	E	P	V	W	A	I	G	T	G	K	S	S	E	D	A	N	E	M	C	A	V	R	O	T	I	A	D	S	--	S	K	E	V	S	E	A	T	R	--	I	Y	G	G	S	V	K	P	N	I	K	E	Y	M	A						
Staphyep	I	T	K	A	--	V	E	G	S	D	D	--	K	E	V	V	I	A	E	P	V	W	A	I	G	T	G	K	S	S	E	D	A	N	E	M	C	A	V	R	O	T	I	A	D	S	--	S	K	E	V	A	D	A	T	R	--	I	Y	G	G	S	V	K	P	N	I	K	E	Y	M	A										
aquifex	I	T	K	A	--																																																																													

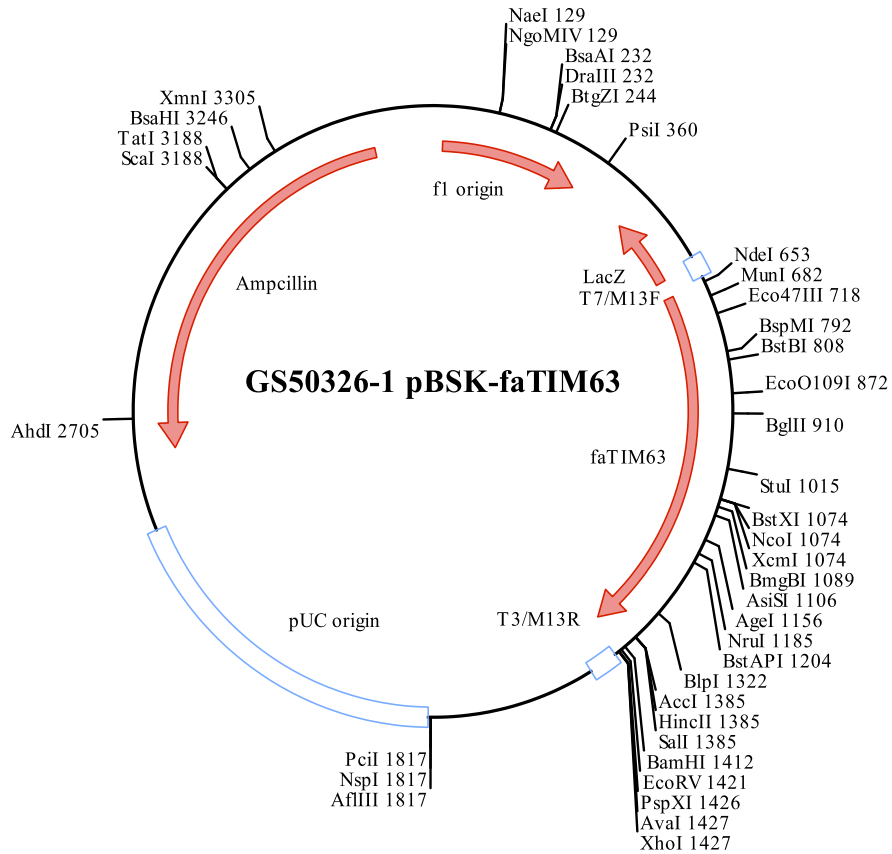
	256	260	270	280	290	300	302
Nitrosom2	PDI	DGGI	GGAS	VAAEFISIC	AAAGN		
Bacterio2	PDV	DGGI	GGAA	KVSDFKGI	DAFNA		
Chlorobi1	PN	DGGI	GGAS	NAADFVAI	VKAAV		
Chlorobi2	PDI	DGGI	GGAS	NADDFMAI	VEAAG		
Ecoli	PDI	DGA	VGGAS	KADAFAVI	VKAAEA	KKQA	
Eerguson	PDI	DGA	VGGAS	KADAFAVI	VKAAEA	KKQA	
Haemoph2	PDI	DGA	VGGAS	KATAFAEI	VKAAEK	AKA	
Pasteurel	PDI	DGA	VGGAS	KAPAFAVI	VKAAAK	AKN	
Haemoph1	PDI	DGA	VGGAS	KAPAFAVI	VKAAAA	AKN	
Photorhab	PDI	DGA	VGGAS	KANAFAVI	VKAAAA	KKQA	
Ralstoni2	ADI	DGGI	GGAS	KAEDFAI	GRA		
Salmonel	PDI	DGA	VGGAS	KADAFAVI	VKAAEA	KKQA	
Shigellaf	PDI	DGA	VGGAS	KADAFAVI	VKAAEA	KKQA	
Salmonelt	PDI	DGA	VGGAS	KADAFAVI	VKAAEA	KKQA	
Pseudomo3	PDI	DGGI	GGAS	NADEF	GAI	CRAAGS	
Pseudomo6	PDI	DGGI	GGAS	NADEF	GAI	CRAAGN	
Pseudomo2	PDI	DGGI	GGAS	NADEF	GAI	CRAAGN	
Xanthomo2	PDV	DGGI	VGGAS	VAEDFAI	ARAAAAC		
Xanthomo3	PDV	DGGI	VGGAS	VAEDFAI	ARAAAAC		
Shewanell	PDV	DGGI	VGGAS	NSTEFLS	CIAMSA		
Xyellaf	QDV	DGGI	GGAS	VADDF	AIARAT	V	
Yersinia1	PDI	DGA	VGGAS	KADAFAVI	VKAAA	KKAKA	
Bordetel	PDI	DGGI	VGGAS	VAEER	RIAAA		
Coxiellab	PN	DGA	VGGAS	EAEQF	XIGQC	NOSF	
Streptoc1	ENV	DGA	VGGAS	GADSF	ALDFV	K	
bacilant	S	D	DGA	VGGAS	EPASF	GAGAV	K
bacihal	S	D	DGA	VGGAS	DPDSF	QTEAV	K
Porphyrom	ADV	DGGI	GGAS	SVDKF	PIIEAF		
Streptoc2	P	D	DGA	VGGAS	EADSF	ALDF	N
Staphyau	T	D	DGA	VGGAS	KVEDF	VDLEGA	K
Staphyep	S	D	DGA	VGGAS	KVEDF	VDLEGA	K
aquifex	E	E	I	NGI	VGTAS	DPESF	AKIVYSF
listein	S	D	DGA	VGGAS	EPASF	ALTEAV	K
listemon	S	D	DGA	VGGAS	EPASF	ALTEAV	K
Streptoc5	P	D	DGA	VGGAS	EAESE	ALDFV	K
Vibrioch	P	D	DGA	VGGAS	DAKFA	AAIAKAAA	EAKKA
Vibriopar	P	D	DGA	VGGAS	DAKFA	AAIAKAAA	EAKKA
Vibriouv	P	D	DGA	VGGAS	DAKFA	AAIAKAAA	EAKKA
Enterococ	ENV	DGA	VGGAS	EADSF	ALDAV	K	
bradyrh	K	N	V	NGA	VGGAS	KAADF	AIKGC
Agrobacte	ENV	DGA	VGGAS	KAADF	SIYAYE	DTA	
Rhodopseu	P	H	V	NGA	VGGAS	KASDF	AIAGCP
photobap	P	D	DGA	VGGAS	DAKFA	AAIAKAAA	EAKKA
neisseri	P	Y	V	DGA	VGGAS	SYDSF	TAISAAGNA
Caulobact	P	E	V	DGA	VGGAS	KAKDF	AIIGAL
deinococ1	P	N	V	NGA	VGGAS	KVPDV	GMDAR
Lactobac5	P	D	DGA	VGGAS	DPESF	ALVNYGD	
Lactobac8	D	D	DGA	VGGAS	QPDSF	EIVNYDNN	
Borrelliab	PN	D	DGA	VGGAS	KAESF	STINNVI	
bifidoba1	P	D	V	DGF	I	GGAA	DVEEAKIARAIKSTKSRN
bifidoba2	P	D	V	DGF	I	GGAA	KVDETRICKIETTA
Dehaloco1	S	D	DGA	VGGAS	KAESF	VSICRGAADV	RVK
desulfovi	D	N	V	DGA	VGGAS	QAVSF	RRIIA
oceanobih	S	D	DGA	VGGAS	DPESF	QVEAGTK	
prochlor	S	D	DGA	VGGAS	DPESF	ARIANYQTS	
synechel	P	E	I	DGA	VGGAS	EAESE	FARIVNYGS
Nostocpun	P	E	I	DGA	VGGAS	EPESF	FARIVNFHIV
Synechocy	P	E	I	DGA	VGGAS	EPESF	FARIVNFQF
Gloeobact	S	D	DGA	VGGAS	EAKSF	FARIVNFEV	
Campylob1	N	H	C	GGV	I	GSAA	KVENFIKIKG
Wolinella	P	E	V	DGA	VGGAS	KVEDFCMI	QKI
helicob2	P	Y	V	NGV	VGGAS	NIQSF	TDIIRASK
Helicob1	D	S	V	DGL	I	GSAS	WENFKIISF
Leptospi1	K	D	D	DGA	VGGAS	QKI	DYAGF
Leptospi2	K	D	D	DGA	VGGAS	QKISS	FAGF
Leptospir	P	N	I	DGGI	VGGAS	QKID	FGLK
thermush	P	N	V	DGGI	VGGAS	ELESF	AIIRIAG
Lactococ1	E	N	I	DGA	VGGAS	EADSF	ALEMYK
fusobanu	E	D	I	DGGI	VGGAS	KADSF	FELIAGN
Streptom1	P	D	I	DGA	VGGAS	DADEF	VKIARFRDQ
Streptom2	P	D	I	DGA	VGGAS	DSDEF	VKIARFRDQ
corynedi	P	D	V	DGGI	VGGAS	DGEAF	AKAANAATVVD
corynegl	P	D	V	DGGI	VGGAS	DGEAF	AKAANAASVA
Mycobtub	D	D	V	DGGI	VGGAS	DGEHF	ATAAIAAGGP
Acidobca	P	E	I	DGA	VGGAS	VAESF	AKIVXF
Clostrid3	P	E	I	DGA	VGGAS	KAADF	AAIVNF
Clostrid5	S	D	I	DGA	VGGAS	KANDF	SAINVYK
Clostrid6	S	D	I	DGA	VGGAS	VADDF	ADIVNY
Geobactel	P	D	I	DGA	VGGAS	KAESF	AGIVRFKKG
Soliba	P	E	I	DGA	VGGAS	DPVSF	ASIVNF
Thermoan3	S	D	I	DGA	VGGAS	VAEDF	AKIVNY
Chlamy	P	D	V	NGI	VGGAS	SVESF	AIIOQI
Chlamytr	P	D	V	NGI	VGGAS	SSENF	SIIQI
Treponem1	H	N	I	DGGI	VGGAS	KTEF	PIAFSE
Treponem2	E	H	I	DGGI	VGGAS	EASAF	VPIARSV
Wiggleswo	K	D	V	NGV	I	GRSS	TNIEEFYI
Wolbach1	S	N	S	GV	I	GSAS	DFDHFYKI
Chromobac	ENV	D	G	A	VGGAS	DADSF	RVICGAAGKMI
Mesorhiz	P	N	I	DGF	I	GRSAW	GADGYIDI
brucelme	P	H	I	DGF	I	GRSAW	VEGYDI
Bdellovi	P	H	V	NGF	VGGAS	EAKSF	SEIASV
fibrobas	K	D	I	DGGI	VGGAS	KANTF	KELIDAAEAK
Onionyel	P	A	I	DGI	VGGAS	QTEDF	FFAGIASKQA
Blochflo	K	D	I	DGV	VGGAS	DIRN	IKIINSSN
Chlamydp	S	D	V	DGI	VGGAS	EGOS	FFEVAKNFHV

Anexo 3. Alineamiento de las secuencias ancestrales de TIM de eucariontes inferidas por máxima verosimilitud, comparada con secuencias de TIM extantes.



Anexo 4 . Mapas de plásmidos de TIMs ancestrales. A) mapa de TIM63, se muestra el esquema donde se ven los sitios de corte para diferentes enzimas. Se puede ver que NdeI se encuentra del lado 5' mientras que BamHI y XhoI se encuentran del lado 3'. El mapa de TIM52 es análogo. B) Mapa de TIM55. Se puede ver que a diferencia del mapa en A, XhoI se encuentra del lado 5', junto a NdeI y del lado 3' sólo se encuentra BamHI.

A)



B)

