



UNIVERSIDAD NACIONAL AUTÓNOMA DE MÉXICO

Maestría y Doctorado en Ciencias Bioquímicas

ESTUDIOS DE CONSERVACION EVOLUTIVA Y FUNCIONAL EN LA FAMILIA
DE PROTEÍNAS EFL1/SBDS INVOLUCRADAS EN LA MADURACIÓN
RIBOSOMAL

TESIS

QUE PARA OPTAR POR EL GRADO DE:
Maestro en Ciencias

PRESENTA:
ALFONSO MÉNDEZ GODOY

Dra. Nuria Victoria Sánchez Puig
[Instituto de Química, UNAM](#)

Dra. Claudia Andrea Segal Kischinevzky
[Facultad de Ciencias, UNAM](#)

Dra. María Soledad Funes Argüello
[Instituto de Fisiología Celular, UNAM](#)

Ciudad de México. Enero, 2017



Universidad Nacional
Autónoma de México

Dirección General de Bibliotecas de la UNAM

Biblioteca Central



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.

Agradecimientos

A la Dra. Nuria Victoria Sánchez Puig por darme la oportunidad de conducir junto con ella el desarrollo de este proyecto.

A la Dra. María Soledad Funes Argüello y a la Dra. Claudia Andrea Segal Kischinevzky, por formar parte del comité que apoyó, asesoró y evaluó este proyecto durante su desarrollo.

Al Dr. Roberto Coria Ortega por permitirnos usar sus instalaciones y por su asesoría durante el desarrollo de este proyecto.

A los Doctores Alicia González Manjarrez, Roberto Coria Ortega, Xochitl Pérez Martínez, Luisa Alvarina Alba Lois y León Patricio Martínez Castilla, que conformaron el jurado encargado de la evaluación de este trabajo.

Al financiamiento recibido para el desarrollo de este trabajo dentro del proyecto DGAPA-PAPIIT IN201615 "*Estudios de complementación funcional en la familia de proteínas EFL1/SBDS involucradas en la maduración ribosomal*" y por la beca otorgada dentro del mismo.

Al Programa de Apoyo para Estudios de Posgrado (PAEP) por el apoyo otorgado para asistir al VIII Congreso Internacional sobre el Síndrome de Shwachman-Diamond, celebrado del 17 al 20 de abril de 2016 en la ciudad de Verona, Italia.

A la Agencia Mexicana de Cooperación Internacional para el Desarrollo (AMEXCID) por el financiamiento recibido para la realización del proyecto de cooperación científica y tecnológica México-Italia "*Caracterización de las interacciones entre proteínas involucradas en la biogénesis ribosomal mediante WAXS y SAXS*", con número de oficio CTC/11048/15, y por el financiamiento de mi estancia en el CNR de la ciudad de Bari, Italia, del 16 de abril al 16 de mayo de 2016.

A CONACyT por la beca otorgada para la realización de mis estudios de maestría con número de registro 329550.

Dedicatorias

A mi madre, Eva Godoy y a mi padre, Alfonso Méndez, por sembrar la semilla con la que comenzó mi historia. Gracias madre, por tu compañía en cada una de mis aventuras y tu buen sentido del humor.

A mis hermanos, Omar y Maribel Méndez, por enseñarme cuál es la cara con la que debemos enfrentar nuestra vida.

A mis sobrinos, Adriana, Cynthia y Chris; por las enseñanzas de cada día. No importa que edad tenga, ustedes me han mostrado que siempre hay que divertirse y tener disciplina.

A Osmaly, por todo el amor y confianza que me ha dado, por la manera en que me ha ayudado a alimentar mis sueños, por saber que no importa lo que pase, siempre podré contar con ella.

A mis compañeros del laboratorio, Nancy, Abril, Nina, Diana, Daniel, Arnulfo, Jesús, Juliana, Montserrat, que más que compañeros siempre serán mis amigos; gracias por todo su apoyo, su paciencia y mantener mi barriga llena.

A mi tutora, por motivarme a siempre creer en mi y por siempre estar dispuesta a discutir conmigo todas mis ideas; gracias por ayudarme a ponerlas sobre la tierra.

Y finalmente, a todas esas personas que se han tomado el tiempo de ensañarme poco a poco a mirar con diferentes ojos qué es la vida y a descubrirme siempre un poquito más. Los nombres de todas esas personas, porque son muchísimas, están escritos entre líneas, en los espacios que existen entre las letras que conforman cada una de estas palabras, porque ¿qué seríamos, si no menos, sin la experiencia de haber conocido a todas esas personas fantásticas?

La materia y la energía se agotaron y con ellas el espacio y el tiempo. Hasta AC existía solamente para la última pregunta que nunca había sido respondida desde la época en que dos técnicos en computación medio alcoholizados, tres trillones de años antes, formularon la pregunta en la computadora que era para AC mucho menos de lo que para un hombre el Hombre.

Todas las preguntas habían sido contestadas, y hasta que esa última pregunta fuera respondida también, AC no podría liberar su conciencia.

Todos los datos recogidos habían llegado al fin. No quedaba nada para recoger.

Pero toda la información reunida todavía tenía que ser completamente correlacionada y unida con todas sus posibles relaciones.

Se dedicó un intervalo sin tiempo a hacer esto.

Y sucedió que AC aprendió cómo revertir la dirección de la entropía.

Pero no había ningún Hombre a quien AC pudiera dar una respuesta a la última pregunta. No había materia. La respuesta –por demostración– se ocuparía de eso también.

Durante otro intervalo sin tiempo, AC pensó en la mejor forma de hacerlo. Cuidadosamente, AC organizó el programa.

La conciencia de AC abarcó todo lo que alguna vez había sido un Universo y pensó en lo que en ese momento era el caos.

Paso a paso, había que hacerlo.

Y AC dijo:

<<¡HÁGASE LA LUZ!>>

Y la luz se hizo...

La última palabra
Isaac Asimov

Índice

1	ANTECEDENTES	9
1.1	GENÓMICA COMPARATIVA: PREDICCIÓN FUNCIONAL DE LAS PROTEÍNAS	9
1.2	IDENTIFICACIÓN DE ORTÓLOGOS MEDIANTE MÉTODOS COMPUTACIONALES	10
1.2.1	MÉTODOS BASADOS EN ÁRBOLES FILOGENÉTICOS	10
1.2.2	MÉTODOS DE AGRUPAMIENTO DE LOS MEJORES RESULTADOS RECÍPROCOS	14
1.2.3	ENSAYOS DE COMPLEMENTACIÓN FUNCIONAL	15
1.2.4	BIOGÉNESIS RIBOSOMAL Y RIBOSOMOPATÍAS	17
1.2.5	EL RIBOSOMA: TRADUCTOR CELULAR UNIVERSAL	17
1.2.6	BIOGÉNESIS RIBOSOMAL EN <i>S. CEREVISIAE</i>	18
1.3	LAS PROTEÍNAS SBDS	22
1.3.1	CONSERVACIÓN EVOLUTIVA Y CARACTERÍSTICAS ESTRUCTURALES	22
1.3.2	LA FUNCIÓN DE LAS PROTEÍNAS SBDS ES ESENCIAL	23
1.4	LAS PROTEÍNAS EFL1	24
1.4.1	CARACTERÍSTICAS GENERALES	24
1.4.2	ORGANIZACIÓN Y FUNCIÓN DE LOS DOMINIOS DE LAS PROTEÍNAS EFL1	25
2	JUSTIFICACIÓN	27
3	HIPÓTESIS	28
4	OBJETIVOS	29
4.1	OBJETIVO GENERAL	29
4.2	OBJETIVOS PARTICULARES	29
5	MATERIALES Y MÉTODOS	31
5.1	ESTUDIO BIOINFORMÁTICO DE LAS PROTEÍNAS EFL1	31
5.2	EVALUACIÓN FUNCIONAL DE ORTÓLOGOS DE EFL1 Y SBDS EN LEVADURA	33
5.2.1	OBTENCIÓN DE LAS CEPAS DIPLOIDES HETEROCIGOTAS PARA LOS ENSAYOS DE COMPLEMENTACIÓN GENÉTICA FUNCIONAL	33
5.2.2	VECTORES DE EXPRESIÓN	39
5.2.3	CONSTRUCCIONES DE LOS MIEMBROS DE EFL1 PARA LOS ENSAYOS DE COMPLEMENTACIÓN FUNCIONAL	44
5.2.4	CONSTRUCCIONES DE LOS MIEMBROS DE SDBS PARA LOS ENSAYOS DE COMPLEMENTACIÓN FUNCIONAL	48
5.2.5	ENSAYOS DE COMPLEMENTACIÓN FUNCIONAL EN <i>S. CEREVISIAE</i>	51
6	RESULTADOS	55
6.1	ESTUDIO BIOINFORMÁTICO DE LAS PROTEÍNAS EFL1	55
6.2	EVALUACIÓN FUNCIONAL DE LAS PROTEÍNAS EFL1 Y SBDS	68
7	DISCUSIÓN DE RESULTADOS	90
8	CONCLUSIONES	102
9	PERSPECTIVAS	103
10	BIBLIOGRAFÍA	104
11	INFORMACIÓN SUPLEMENTARIA	115

Abreviaturas

mRNA	Ácido ribonucleico mensajero
rRNA	Ácido ribonucleico ribosomal
snoRNA	Ácido ribonucleico pequeño nucleolar
snoRNP	Partícula ribonucleoproteica pequeña nucleolar
RMN	Resonancia Magnética Nuclear
gDNA	Ácido desoxiribonucleico genómico
pDNA	Ácido desoxiribonucleico plasmídico
PCR	Reacción en cadena de la polimerasa, por sus siglas en inglés <i>Polymerase Chain Reaction</i>
DMSO	Dimetil sulfóxido
PVDF	Polifluoruro de vinilideno, por sus siglas en inglés <i>polyvinylidene fluoride</i>
pb	Pares de bases
X-gal	5-bromo-4-cloro-3-indolil- β -D-galactopiranosido
DO_{600nm}	Densidad óptica a 600 nm
CPN	Complejo del poro nuclear
GAPDH	Gliceraldehído-3-fosfato deshidrogenasa
SGM	Medio de crecimiento sintético para levaduras, por sus siglas en inglés <i>Synthetic Growth Medium</i>
SD	Medio de crecimiento sintético definido para levaduras, por sus siglas en inglés <i>Synthetic Defined (Medium)</i>
cloNAT	Nourseotricina. Antibiótico empleado en el medio SGM para seleccionar a la progenie <i>eff1Δ</i> .
G418	Genitcina. Antibiótico empleado en el medio SGM para seleccionar a la progenie <i>sdo1Δ</i> .
Proteínas SBDS	Ortólogos de la proteína Sdo1 de <i>S. cerevisiae</i>
ScSdo1	Proteína Sdo1 de <i>S. cerevisiae</i> .
SpSdo1	Proteína Sdo1 de <i>S. pombe</i> .

HsSBDS	Proteína SBDS de <i>H. sapiens</i> .
MmSBDS	Proteína SBDS de <i>M. musculus</i> .
AfSbds	Proteína Sbds de <i>A. fulgidus</i> .
HsSBDS D2Sc	Proteína SBDS quimérica de <i>H. sapiens</i> , en la que se reemplazó el dominio 2 con su equivalente de la proteína Sdo1 de <i>S. cerevisiae</i> .
ScSdo1 D2Hs	Proteína Sdo1 quimérica de <i>S. cerevisiae</i> , en la que se reemplazó el dominio 2 con su equivalente de la proteína SBDS de <i>H. sapiens</i> .
Proteínas EFL1	Ortólogos del factor de elongación tipo 1 de <i>S. cerevisiae</i> .
ScEfl1	Factor de elongación tipo 1 de <i>S. cerevisiae</i> .
SpEfl1	Factor de elongación tipo 1 de <i>S. pombe</i> .
HsEFL1-1	Isoforma 1 del factor de elongación tipo 1 de <i>H. sapiens</i> .
HsEFL1-2	Isoforma 2 del factor de elongación tipo 1 de <i>H. sapiens</i> .
MmEFL1	Factor de elongación tipo 1 de <i>M. musculus</i> .
AfEf-2	Factor de elongación 2 de <i>A. fulgidus</i> .
Proteínas eEF2	Comprende todos los factores de elongación 2 de eucariontes.
Proteínas aEF2	Comprende todos los factores de elongación 2 de arqueas.
Tif6	Factor de antiasociación de <i>S. cerevisiae</i> .
eIF6	Ortólogo de la proteína Tif6 en eucariontes.
aIF6	Ortólogo de la proteína Tif6 en arqueas.
pHMM	Perfil de HMMER construido con las secuencias de aminoácidos de distintos ortólogos de EFL1.

Resumen

Durante el proceso de biogénesis ribosomal en eucariontes, las pre-subunidades ribosomales 60S y 40S son exportadas al citoplasma para llevar a cabo los últimos eventos de maduración. En *Saccharomyces cerevisiae*, la subunidad 60S llega a este compartimento cargada con Tif6, una proteína que evita la asociación prematura entre la subunidad 60S y la subunidad 40S. Efl1 y Sdo1, son dos proteínas accesorias encargadas de catalizar la liberación de Tif6 de la subunidad ribosomal 60S y promover la formación de la partícula ribosomal 80S. Mientras que se ha observado que Tif6 y Sdo1 se encuentran conservadas en Archaea y Eukarya, hasta el momento sólo se han encontrado miembros de la familia EFL1 en levadura y mamíferos.

En este trabajo se realizó una búsqueda bioinformática de miembros de la familia EFL1, y se construyó un perfil filogenético a partir de la identificación y recopilación de sus diferentes miembros. Demostramos que las proteínas EFL1 están altamente conservadas en Eukarya y que son exclusivas de este dominio. Por otro lado, el alineamiento múltiple de secuencias reveló que estas proteínas conservan regiones importantes para la catálisis y sugieren que su plegamiento podría estar conservado evolutivamente. Adicionalmente, un análisis filogenético mostró que la familia EFL1 comparte un origen evolutivo con las translocasas EF-2 y EF-G. Además, sugirió que EFL1 y eEF-2 (de eucariontes) fueron originadas a través de un fenómeno de duplicación génica, indicando que mientras eEF-2 mantuvo su función en el proceso de síntesis de proteínas, la familia EFL1 se especializó en la vía de maduración ribosomal.

Finalmente, para evaluar una posible función conservada de los miembros de EFL1, así como una función relacionada de la familia de factores de elongación 2 de Archaea (aEF-2), decidimos realizar ensayos de complementación funcional en células de *S. cerevisiae efl1Δ*, con las proteínas EFL1 de *Mus musculus*, *Homo sapiens*, *Schizosaccharomyces pombe* y EF-2 de *Archaeoglobus fulgidus*. Los resultados mostraron que ninguno de los miembros evaluados fueron capaces de complementar la función del gen *EFL1* en *S. cerevisiae*. Estudios similares con miembros de la familia SBDS (ortólogos de la proteína Sdo1) revelaron que la conservación funcional de estas proteínas es casi nula, sugiriendo un posible fenómeno de coevolución entre las familias EFL1 y SBDS. Para comprobar lo anterior, ensayamos la doble complementación funcional con ambas familias de proteínas. De manera interesante, los pares de ortólogos de los organismos evaluados tampoco complementaron la función de los genes eliminados en levadura, sugiriendo que el reconocimiento entre los miembros de las familias SBDS y EFL1, así como entre el complejo anterior y la subunidad ribosomal 60S, ocurre de manera especie-específica. En conjunto, nuestros resultados indican que el fenómeno de co-evolución no sólo ocurrió entre las proteínas accesorias sino también entre EFL1 y su superficie de interacción en la subunidad ribosomal 60S.

1 Antecedentes

1.1 Genómica comparativa: predicción funcional de las proteínas

Las proteínas son un grupo de biomoléculas que presentan una función crucial en prácticamente todos los procesos celulares. Estas macromoléculas están compuestas por cadenas de aminoácidos cuyo orden se encuentra codificado en el DNA. Gracias al advenimiento de nuevas técnicas analíticas y al progreso en los métodos de secuenciación masiva, se han secuenciado completamente los genomas de diversos organismos y actualmente estas secuencias se encuentran disponibles en diversas bases de datos [1]. Esta información ha sido de gran utilidad para inferir el número aproximado de proteínas contenido en cada genoma y conocer la secuencia de aminoácidos por la que se compone cada una de ellas. Estrictamente, la función de cualquier biomolécula, en este caso una proteína, debe ser establecida mediante ensayos experimentales que impliquen su estudio directo. Sin embargo, debido al gran número de proteínas predichas hasta la fecha, el estudio individual de cada una de ellas podría demandar una enorme cantidad de tiempo. Con el objetivo de sobrepasar el obstáculo anterior, diversas herramientas computacionales han sido implementadas para predecir la función de una proteína determinada a partir de su información evolutiva [1–4]. Es importante mencionar, que a pesar de que este tipo de estudios no son completamente concluyentes, son capaces de complementar, dirigir y enfocar los ensayos experimentales.

Para comprender cómo se lleva a cabo la anotación funcional de las proteínas en las bases de datos, es necesario conocer las relaciones evolutivas que mantienen entre ellas y las implicaciones que éstas presentan sobre su función. Los genes o proteínas que provienen de un ancestro en común son considerados como homólogos; adicionalmente, los homólogos también pueden ser clasificados como parálogos y ortólogos, dependiendo del proceso por el cual fueron originados. Se conocen como parálogos a los genes que fueron producidos mediante un fenómeno de duplicación génica dentro de un mismo organismo, mientras que se consideran como ortólogos a aquellos genes que han sido separados por eventos

de especiación. Mientras que los genes parálogos suelen presentar funciones diferentes entre sí, las proteínas codificadas en genes ortólogos generalmente mantienen su función [5–7]. Esta característica de los genes separados por especiación es aprovechada para anotar la función de una proteína a partir de la identificación de sus ortólogos [8–11].

1.2 Identificación de ortólogos mediante métodos computacionales

Los métodos computacionales utilizados para identificar familias de ortólogos pueden ser clasificados de manera general en métodos basados en árboles filogenéticos y métodos basados en la búsqueda de los mejores resultados recíprocos [4,5,7]. Ambas estrategias presentan ventajas y desventajas que hay que considerar antes de comenzar la búsqueda de los grupos ortólogos de interés; por lo tanto, para comprender mejor la manera en que trabajan, a continuación se describen las características generales de cada uno.

1.2.1 Métodos basados en árboles filogenéticos

El análisis filogenético es una de las herramientas más utilizadas en el campo de la evolución molecular [12–15]. El objetivo de este análisis es identificar las relaciones evolutivas que existen entre los genes o proteínas dentro de un mismo y entre diferentes organismos, por lo tanto, el empleo de esta clase de estudios resulta sumamente útil para la identificación de grupos de ortólogos [11]. Este análisis emplea secuencias de nucleótidos o de aminoácidos y las compara entre sí al mismo tiempo. En seguida, con la información obtenida en el análisis anterior, se construyen diagramas de árboles (conocidos también como cladogramas) que representan las relaciones que mantienen entre sí las biomoléculas en estudio. En la Figura 1 se muestran algunas de las partes principales por las que está compuesto un árbol filogenético.

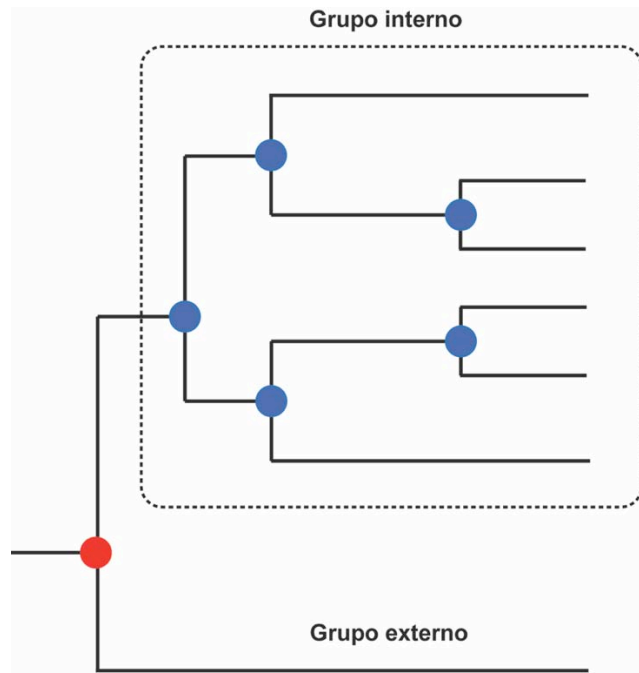


Figura 1. Características generales de un árbol filogenético. En los árboles filogenéticos se pueden observar dos clase de grupos: el grupo interno y el grupo externo. El grupo interno contiene todas las biomoléculas en las que estamos interesados en estudiar; por otra parte, el grupo externo es utilizado para definir la raíz del árbol. Esta raíz (rojo), representa el último ancestro en común de todas las biomoléculas representadas en el árbol (puntas de las ramas). Los nodos en azul representan los ancestros en común más recientes de los miembros que descienden de ellos y pueden ser interpretados como puntos de duplicación génica o de especiación.

En el fondo del árbol, a no ser que se trate de un árbol no enraizado, se encuentra la raíz, que representa al último ancestro en común de todas las biomoléculas analizadas; éstas últimas son localizadas en las puntas de cada rama. Los nodos o puntos de ramificación internos son considerados como los ancestros en común más recientes de los miembros que descienden de ellos, por lo tanto, además de lo anterior, pueden indicar eventos de duplicación génica o de especiación [16]. Para discriminar entre estas dos clases de eventos, los árboles obtenidos pueden ser sometidos a análisis sofisticados como los de reconciliación de especies o simplemente emplear la regla de superposición de especies (*specie-overlap*), que sugiere que los puntos de ramificación que representan eventos de duplicación contienen ramas de especies mutuamente excluyentes [4]; en este caso, ambos grupos son considerados individualmente como ortólogos y entre sí como parálogos.

Análisis filogenético de proteínas

La realización de un análisis filogenético se resume en cuatro pasos importantes: el alineamiento de las secuencias a estudiar, la determinación del modelo de sustitución adecuado y finalmente, la construcción y evaluación del árbol [7]. Para proteínas, las secuencias de aminoácidos usualmente son alineadas empleando programas como CLUSTAL W (actualmente sustituido por CLUSTAL Omega), MAFFT o T-COFFEE. El alineamiento resultante puede ser editado manualmente y posteriormente utilizado para elegir un modelo de sustitución adecuado [7]. Los modelos de sustitución más usados para el análisis de proteínas son PAM y BLOSUM; ambos son colecciones de puntuaciones que representan la probabilidad relativa con la que un aminoácido puede ser sustituido por otro [17]. El mejor modelo de sustitución depende del grado de similitud que existe entre las secuencias analizadas; generalmente se utiliza BLOSUM para secuencias con un alto grado de divergencia, mientras que PAM se emplea particularmente para secuencias con alta similitud. Una herramienta útil para inferir el modelo de sustitución más adecuado es el servidor ProtTest 2.4 [18].

Métodos de construcción de árboles filogenéticos

Una vez obtenido el alineamiento y seleccionado el modelo de sustitución, el siguiente paso es la construcción del árbol filogenético. Para esto, existen también una gran diversidad de métodos que son clasificados como métodos basados en distancia y métodos basados en caracteres, a continuación se describen los fundamentos de cada uno de ellos.

Métodos basados en distancia

Los métodos basados en distancia construyen el árbol tomando en cuenta el grado de disimilitud que existe entre las secuencias analizadas; ejemplos de ellos son Unión de Vecinos (*Neighbor joining*, NJ), Pares No Ponderados utilizando Media Aritmética (*Unweighted Pair Group Method with Arithmetic Mean*, UPGMA), Fitch-Margoliash (FM) y Evolución Mínima (*Minimum Evolution*, ME) [7]. NJ y UPGMA emplean diferentes estrategias de agrupación pero en sí, ambos asignan

cada rama de acuerdo a su similitud [19]. Por otro lado, FM y ME generan árboles con diferentes topologías y seleccionan al que se haya ajustado mejor de acuerdo con las distancias estimadas por el modelo evolutivo y las distancias evolutivas reales [20]. La ventaja principal de esta clase de métodos es que requieren un gasto computacional menor al que demandan aquellos que están basados en caracteres. Algunos de los programas que emplean los métodos basados en distancia para construir árboles son MEGA7 [21] y PAUP* 4.0, aunque este último también puede emplear métodos basados en caracteres [22].

Métodos basados en caracteres

Los métodos basados en caracteres se subdividen en métodos de Máxima Parsimonia (MP) y métodos de Máxima Verosimilitud (MV). A pesar de que presentan grandes diferencias entre sí, ambos se caracterizan por tomar en cuenta cada carácter o posición del alineamiento de secuencias, como información evolutiva para construir todos los árboles que puedan explicar adecuadamente las relaciones evolutivas que presentan las secuencias estudiadas. Los métodos de MP otorgan una calificación a cada árbol o hipótesis, de acuerdo al número de cambios evolutivos requeridos para explicar mejor los resultados estudiados. El mejor árbol o más parsimonioso (la hipótesis más sencilla), es aquel que presenta la menor cantidad de cambios evolutivos para todas las secuencias que provienen o derivan de un ancestro en común. Por otra parte, el método de MV calcula la verosimilitud para cada uno de ellos usando un modelo de evolución explícito; aquel con la máxima verosimilitud es el árbol que explica mejor los datos observados [23,24]. La paquetería de PHYLIP contiene diversos programas que emplean ambos métodos basados en caracteres para construir árboles [25], mientras que PhyML y RaxML utilizan métodos de MV; todos estos programas pueden ser utilizados en el servidor T-REX [26].

Evaluación del árbol filogenético

Existen diferentes estrategias para evaluar la calidad un árbol filogenético; la más adecuada depende del método que haya sido empleado para su construcción. Uno de los métodos más utilizados es el de Bootstrapping debido a que es operable en los métodos de construcción del árboles basados en distancia y en caracteres [7]. El Bootstrapping realiza un remuestreo de los datos originales para generar conjuntos de datos nuevos con los que posteriormente se construyen diversos árboles (generalmente entre 100 a 1000). Finalmente, la información del análisis anterior se emplea para calificar con un valor de Bootstrap a cada rama del árbol que mejor representa las relaciones evolutivas de las moléculas en estudio. Este valor representa la proporción de veces en la que un taxón aparece en una determinada posición en cada uno de los árboles muestreados [27]. Algunos estudios han revelado que los valores de Bootstrap por encima de 70% corresponden con una probabilidad mayor al 95% de que la filogenia observada es verdadera, sin embargo, existen varias publicaciones en donde se han tomado valores por encima del 50% como aceptables [7,28].

1.2.2 Métodos de agrupamiento de los mejores resultados recíprocos

Existen varios métodos para identificar ortólogos utilizando como criterio la similitud entre sus secuencias de aminoácidos. La prueba de BLAST recíproco, por ejemplo, considera como pares de ortólogos a dos proteínas de especies diferentes que al ser empleadas como consultas de BLAST en búsquedas independientes, se obtienen como el mejor resultado entre sí [29,30]. Sin embargo, se ha observado que esta estrategia también es capaz de producir una gran cantidad de falsos positivos que generalmente corresponden a outparálogos o a proteínas “multidominio” (estos términos son conocidos en inglés como *out-paralogs* y *multidomain proteins*, respectivamente) [6,31,32]. Los outparálogos y su contraparte, los in-parálogos (en inglés *in-paralogs*), son genes originados por eventos de duplicación antes y después de un evento de especiación, respectivamente. Se ha observado que mientras los in-parálogos frecuentemente mantienen la función de sus ortólogos, los outparálogos divergen de ella [4].

Después de notar que el método de BLAST recíproco no era capaz de distinguir entre inparálogos y outparálogos, así como excluir proteínas multidominio, esta estrategia se optimizó para dar origen a Inparanoid y OrthoMCL [33–35]. Ambas herramientas computacionales trabajan bajo el principio de BLAST recíproco, pero a su vez implementan diferentes algoritmos que les permiten excluir outparálogos y proteínas multidominio del grupo de estudio, mientras que retienen a los inparálogos [5]. Una diferencia importante entre estas dos estrategias de búsqueda, es que mientras Inparanoid genera familias de pares de ortólogos, OrthoMCL genera grupos que contienen ortólogos de más de dos especies. En particular, una vez que se ha identificado el grupo de homólogos de interés usando BLAST recíproco, OrthoMCL implementa un algoritmo de agrupamiento de Markov (MCL por sus siglas en inglés *Markov Clustering Algorithm*) para diferenciar entre grupos de ortólogos y parálogos; este algoritmo reduce el tamaño del grupo más grande, que contiene aún relaciones filogenéticas débiles, a partir de la construcción de grupos más pequeños que representan con mayor robustez ortólogos verdaderos [36].

1.2.3 Ensayos de complementación funcional

Una vez que la función de un grupo de ortólogos ha sido inferida a partir de estudios computacionales, lo ideal es demostrar experimentalmente que ésta se encuentra conservada de manera evolutiva. Para abordar lo anterior, el grupo de ortólogos de interés debe ser estudiado directamente mediante ensayos *in vitro* o *in vivo* [5]. Generalmente, para determinar la función de una familia de proteínas *in vitro*, el primer paso es establecer un protocolo de purificación que permita obtener cada proteína en cantidad y pureza adecuadas. Sin embargo, lo anterior no es siempre posible debido a problemas frecuentes con la disponibilidad de su fuente natural o abundancia en ella, así como limitantes en su proceso de expresión heteróloga. Por otro lado, aun cuando se superan los problemas que entorpecen el proceso de purificación, un inconveniente más de esta clase de estudios es la dificultad que existe para establecer ensayos experimentales que simulen las vías o los procesos celulares en los que participan las proteínas a evaluar.

Por otra parte, a pesar de que los ensayos *in vivo* también presentan algunas complicaciones, una de sus ventajas principales es que no requieren de la purificación de las proteínas de estudio debido a que su función es evaluada directamente en un organismo modelo. Estos organismos se caracterizan por presentar un tamaño pequeño y ciclo de vida corto, así como ser fácilmente manipulables para realizar estudios bioquímicos y genéticos [37]. Uno de los organismos modelo más utilizados para ensayos *in vivo* es la levadura de pan *Saccharomyces cerevisiae*. A partir de este eucarionte se han desarrollado técnicas experimentales como los ensayos de letalidad sintética y de doble híbrido, que permiten evaluar complejas redes de interacción entre proteínas, así como ensayos de complementación funcional que son de gran utilidad para probar la conservación funcional entre ortólogos, e inclusive, fenómenos de coevolución entre ellos [38–41]. Los ensayos de complementación funcional en *S. cerevisiae* consisten en evaluar el efecto fenotípico que se produce sobre el organismo al reemplazar una de sus proteínas endógenas con sus ortólogos correspondientes. Estos estudios requieren de una cepa de levadura que carezca de la proteína a evaluar, de la clonación de las secuencias que codifican para las proteínas de interés en un vector que dirija su expresión en el organismo heterólogo, y de un fenotipo que valide que las proteínas estudiadas complementan la función de la proteína endógena [11,38].

La estrategia para generar cepas mutantes de levadura es sencilla debido a que este organismo presenta una alta frecuencia de recombinación. El marco abierto de lectura blanco es sustituido fácilmente por módulos de resistencia a antibióticos como *natMX4* y *kanMX6*, o marcadores nutricionales tales como *URA3*, *MET15*, *HIS3*, *LEU2*, *TRP1*, entre otros [42,43]. Respecto a los vectores de expresión, en la actualidad se dispone de una enorme gama de plásmidos para levadura [44–47]. Para ensayos de complementación funcional, es común emplear vectores centroméricos porque disminuyen el riesgo de producir falsos positivos provocados por sobreexpresión. Por último, el fenotipo más informativo a evaluar en esta clase de experimentos es la viabilidad de las células de *S. cerevisiae*. Las mutaciones que afectan la función de proteínas esenciales generalmente provocan

defectos en el crecimiento de la levadura, en consecuencia, sólo los ortólogos que logran representar el papel funcional de la proteína endógena afectada son capaces de reestablecer el crecimiento normal del organismo modelo [5].

1.2.4 Biogénesis ribosomal y ribosomopatías

El ribosoma es uno de los complejos moleculares más importantes en la célula; su construcción debe ser precisa y se encuentra altamente regulada. Los defectos en sus componentes estructurales y en factores implicados en su biogénesis, provocan el desarrollo de un grupo de enfermedades conocidas como ribosomopatías. Cada una de ellas se caracteriza por presentar un cuadro clínico específico, sin embargo, comparten algunas características clínicas entre sí, como defectos en la médula ósea, alteraciones esqueléticas y, en su mayoría, una alta predisposición al desarrollo de cáncer [48–52].

Actualmente, diversos grupos de investigación en el mundo han comenzado a realizar estudios sobre el proceso de biogénesis ribosomal en eucariontes (principalmente en organismos como *S. cerevisiae*, *Danio rerio* y *Mus musculus*). El objetivo es obtener la mayor cantidad de información sobre este proceso para comprender mejor su relación con otros eventos celulares, su posible uso como blanco terapéutico y el mecanismo específico por el cual se originan cada una de las características clínicas que definen a cada una de las ribosomopatías. Definitivamente, lo anterior no sólo será importante durante el desarrollo de nuevas estrategias terapéuticas para combatir a las ribosomopatías, sino también para el tratamiento de una de las enfermedades que más afecta a la humanidad en el siglo XXI: el cáncer [48]. A continuación se describe la función, composición y biogénesis del ribosoma eucarionte.

1.2.5 El ribosoma: traductor celular universal

Los ribosomas son las máquinas moleculares encargadas de sintetizar todas las proteínas requeridas para el crecimiento y el mantenimiento celular. Este complejo ribonucleoprotéico es capaz de decodificar la información contenida en los RNA mensajeros (mRNA) y sintetizar proteínas específicas. La mayor parte de los

estudios sobre el proceso de ensamblaje y función del ribosoma en eucariontes se han realizado en *S. cerevisiae*. En este organismo, cada ribosoma está conformado por dos subunidades. La subunidad menor (40S) está compuesta por el RNA ribosomal (rRNA) 18S y 33 proteínas, mientras que la subunidad mayor (60S) se compone por tres rRNA (5.8S, 5S y 25S) y 46 proteínas [53]. Además de su composición, cada subunidad es funcionalmente diferente; mientras que la subunidad menor actúa como el centro de decodificación del mRNA, la subunidad mayor es el sitio que cataliza la formación del enlace peptídico entre aminoácidos mediando la formación de la proteína [54].

1.2.6 Biogénesis ribosomal en *S. cerevisiae*

La biogénesis ribosomal comienza en el nucléolo, en donde los rRNA pre-35S (precursor de las especies 18S, 5.8S y 25S) y 5S, son transcritos por la RNA polimerasa I y la RNA polimerasa III, respectivamente [55,56]. Mientras el pre-35S es transcrito, experimenta diferentes modificaciones químicas, como la pseudourinidilación y metilación; éstas reacciones químicas son realizadas por cerca de 75 ribonucleoproteínas pequeñas nucleolares (snoRNP). Se ha visto que los ribosomas que no presentan estas modificaciones se ensamblan correctamente pero presentan una reducción en su eficiencia y fidelidad de traducción [57–60]. Por otro lado, este proceso también requiere de la actividad de la RNA polimerasa II, quien se encarga de sintetizar los transcritos que codifican para las proteínas ribosomales y factores accesorios necesarios en la síntesis ribosomal [61]. Cerca del 60% de los mRNA producidos por la RNA polimerasa II, codifican para componentes del proceso de biogénesis del ribosoma; este porcentaje refleja la importancia del proceso de ensamblaje ribosomal en el mantenimiento de la función celular.

Conforme el pre-rRNA 35S es transcrito y químicamente modificado, un pequeño subgrupo de proteínas ribosomales (principalmente de la subunidad 40S) y proteínas accesorias se asocian con la molécula de ácido ribonucleico dando origen a la partícula 90S (Figura 2) [62]. Este complejo es procesado por una serie de endo y exonucleasas que realizan diversos cortes sobre el pre-rRNA 35S. El

corte sobre el sitio A2 permite la liberación de la subunidad pre-40S y el rRNA 27SA₂ (precursor de las especies 25S y 5.8S). En seguida, el rRNA 27SA₂ se asocia con proteínas ribosomales y se ensambla con el rRNA 5S para dar origen a la partícula pre-60S. En este punto, ambas pre-subunidades siguen vías de maduración independientes. Por un lado, la subunidad pre-40S es rápidamente exportada al citoplasma, en donde experimenta dos eventos de maduración principales: (1) un rearrreglo estructural producido por una serie de factores accesorios y (2) el corte endonucleolítico por la nucleasa Nob1 sobre el rRNA 20S para producir el rRNA 18S maduro [53]. El proceso de maduración de la subunidad 60S es mucho más complejo e involucra, entre otras cosas, el procesamiento de los rRNA, así como la disociación de diversos factores accesorios y asociación de diversas proteínas ribosomales [54,63–66]. En este trabajo nos enfocaremos brevemente en el proceso de maduración de la subunidad 60S. Para más información acerca del proceso de maduración de la subunidad 40S revisar las referencias [53,54,67].

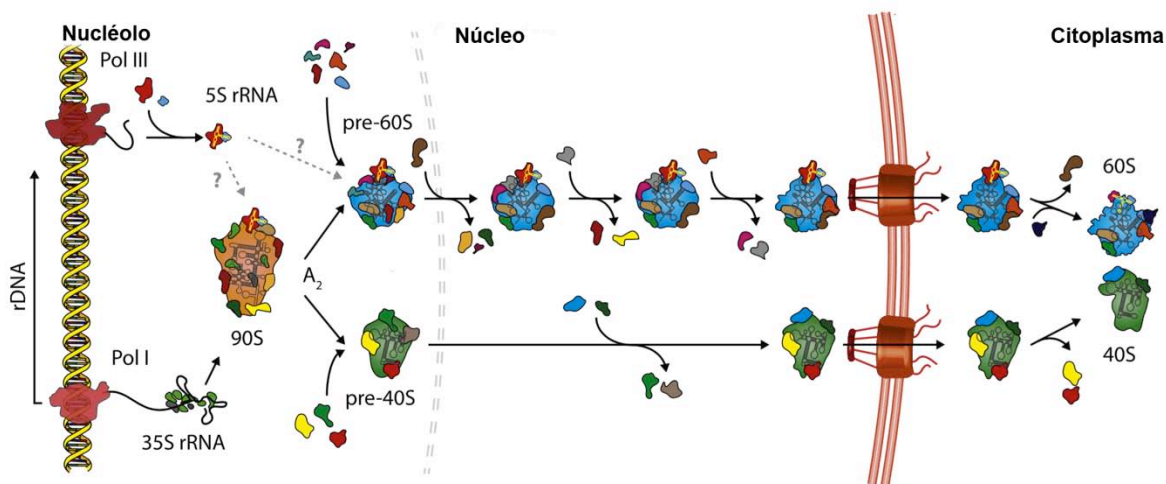


Figura 2. Proceso de biogénesis ribosomal. La asociación del rRNA ribosomal con proteínas ribosomales y accesorias da origen a la partícula 90S la cual al seguir su procesamiento produce las subunidades pre-60S y pre-40S. En seguida, ambas partículas son transportadas desde el nucléolo al nucleoplasma y posteriormente, exportadas al citoplasma, en donde sufren los últimos eventos de su maduración y se ensamblan para formar la partícula traduccionalmente activa 80S. Modificada de [68].

Exportación de la pre-subunidad ribosomal 60S

Debido a que en eucariontes el proceso de traducción ocurre en el citoplasma, ambas pre-subunidades ribosomales deben ser transportadas hasta este compartimento. Diversos factores accesorios tales como Npl3, Mex67, Mtr2, Ecm1, Gle2, Bud20, Arx1 y Abl1, presentan un papel relevante en la exportación de la pre-subunidad 60S. Sin embargo, la proteína que media principalmente este proceso es Nmd3. Este factor accesorio actúa como un adaptador que dirige la interacción entre la pre-subunidad 60S y la proteína Crm1, quien promueve el transporte de la partícula ribonucleoprotéica a lo largo del complejo del poro nuclear (CPN) hasta alcanzar el citoplasma [54,68].

Maduración de la subunidad ribosomal 60S

Una vez que la pre-subunidad 60S llega al citoplasma, se encuentra lista para enfrentar los últimos eventos de su maduración. Las últimas modificaciones que esta subunidad experimenta comprenden el procesamiento final del rRNA para producir a las especies 25S y 5.8S, el ensamblaje de las últimas proteínas ribosomales y la liberación de las proteínas accesorias que se encargaron de facilitar su ensamblaje, su exportación y de prevenir una interacción prematura con la subunidad 40S [69].

Estudios acerca de la composición de la partícula pre-ribosomal han permitido identificar al menos dos vías que convergen en la maduración de la subunidad pre-60S [70,71]. Una de ellas (Figura 3) involucra el reciclaje de Mtr4, una RNA helicasa dependiente de ATP encargada del procesamiento del rRNA 5.8S [72–74]. Mtr4 es un parálogo de P0, un componente esencial del tallo ribosomal. El tallo es una estructura característica de la subunidad 60S funcional, que se encarga de reclutar a los factores de activación y de elongación durante el proceso de traducción [75]; su ensamblaje sobre la subunidad mayor depende de la liberación de Mtr4 por parte de la fosfatasa Yvh1. Cuando la disociación de Mtr4 no ocurre, afecta el ensamblaje del tallo y provoca que las proteínas Tif6 y Nmd3 no sean recicladas al núcleo [69]. Tif6, conocida en mamíferos como eIF6 y en arqueas como aIF6, es un factor de anti-asociación que permanece unido a la

subunidad pre-60S e impide que se asocie prematuramente con la subunidad 40S [76]. La disociación de Tif6 de la subunidad 60S es esencial para permitir la liberación de Nmd3 por parte de la GTPasa Lsg1, y la asociación de las subunidades 60S y 40S, para dar origen al ribosoma funcional 80S.

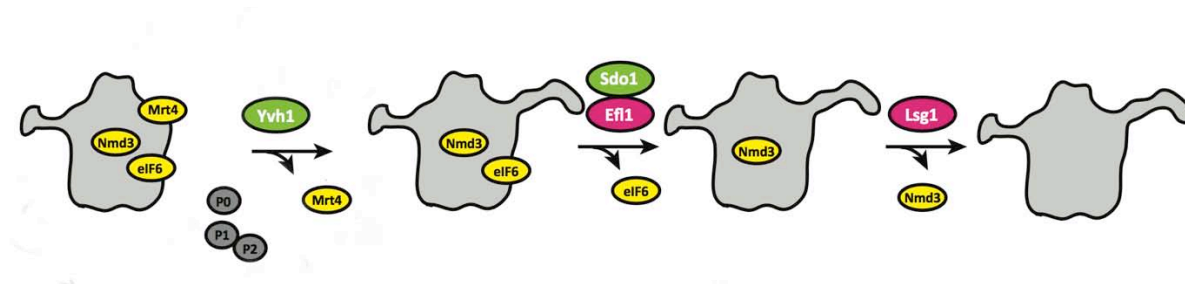


Figura 3. Últimos eventos de maduración de la subunidad ribosomal 60S. Tif6 es representada por su ortólogo eucariote eIF6, y P0, P1 y P2, corresponden a las proteínas que componen al tallo ribosomal. Modificada de [77].

Liberación de la proteína Tif6 de la subunidad ribosomal 60S

La liberación de Tif6 es catalizada por la acción conjunta de las proteínas Efl1 y Sdo1 [78–80] (Figura 4). Inicialmente se forma el complejo Efl1·GTP·Sdo1 que contacta posteriormente a la subunidad ribosomal mayor. El nucleótido trifosfatado es hidrolizado por la actividad de la GTPasa Efl1. Diversos investigadores han propuesto que la energía aportada por la hidrólisis de la molécula de alta energía, provoca un cambio conformacional en Efl1 que impacta directamente sobre la estructura de subunidad ribosomal 60S permitiendo que Tif6 sea liberada y posteriormente reciclada al núcleo. La GTPasa Efl1, de manera independiente, presenta mayor afinidad por GDP que por GTP. Estudios de unión entre Efl1 y ambos nucleótidos en presencia de Sdo1, han demostrado que ésta última promueve el recambio de GDP por GTP en la GTPasa [81] Nuestro trabajo está dirigido al estudio de este evento de la maduración ribosomal. En particular nos interesa estudiar al grupo de proteínas que constituye a la familia EFL1 y Sdo1 (SBDS). A continuación se describe la información más relevante sobre ambos factores accesorios.

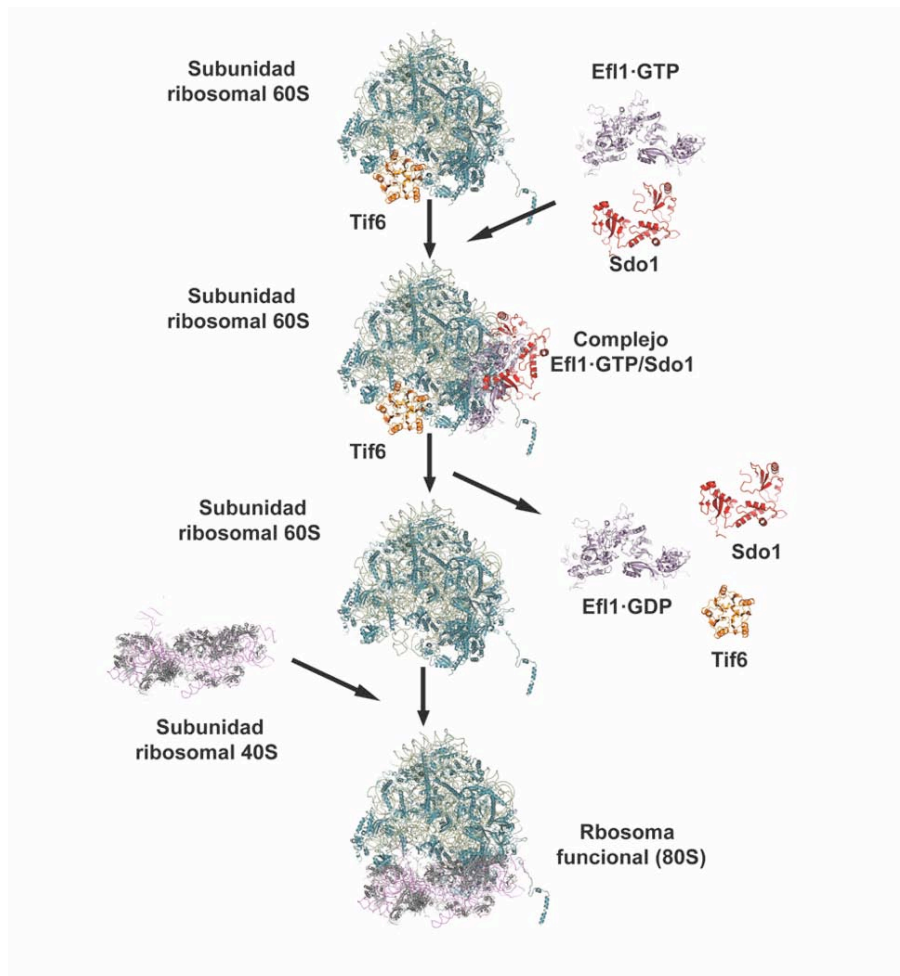


Figura 4. Modelo de la liberación de la proteína Tif6 de la subunidad ribosomal 60S y ensamblaje de las subunidades ribosomales 60S y 40S para dar origen al ribosoma funcional [82].

1.3 Las proteínas SBDS

1.3.1 Conservación evolutiva y características estructurales

La proteína Sdo1 de *S. cerevisiae* (ScSdo1) y su ortólogo humano HsSBDS, pertenecen a una familia de proteínas altamente conservada en Archaea y Eukarya [11]. Estudios de resonancia magnética nuclear de la proteína humana, y de difracción de rayos X de los miembros de *Archaeoglobus fulgidus* (AfSbds) y *Methanothermobacter thermotrophicus* (MtSbds) [83–85], revelaron que este conjunto de proteínas se organiza en tres dominios estructurales (Figura 5). La estructura del dominio I o N-terminal presenta el mismo plegamiento que la proteína de levadura Yhr87wp [84], y se ha demostrado que junto con el dominio 2 o central, interactúa con la subunidad ribosomal 60S *in vitro* [86].

Adicionalmente, se ha demostrado que la función de ambos dominios es esencial, mientras que la del dominio 3 resulta dispensable. De manera interesante, también se ha observado que mientras el dominio N-terminal es intercambiable entre especies, el dominio central funciona de manera especie específica [11]. Por otra parte, resultados obtenidos en nuestro laboratorio y en otros grupos de investigación, han demostrado que la interacción entre la proteína Efl1 de *S. cerevisiae* y ScSdo1, se lleva a cabo principalmente por el dominio 2 [87].

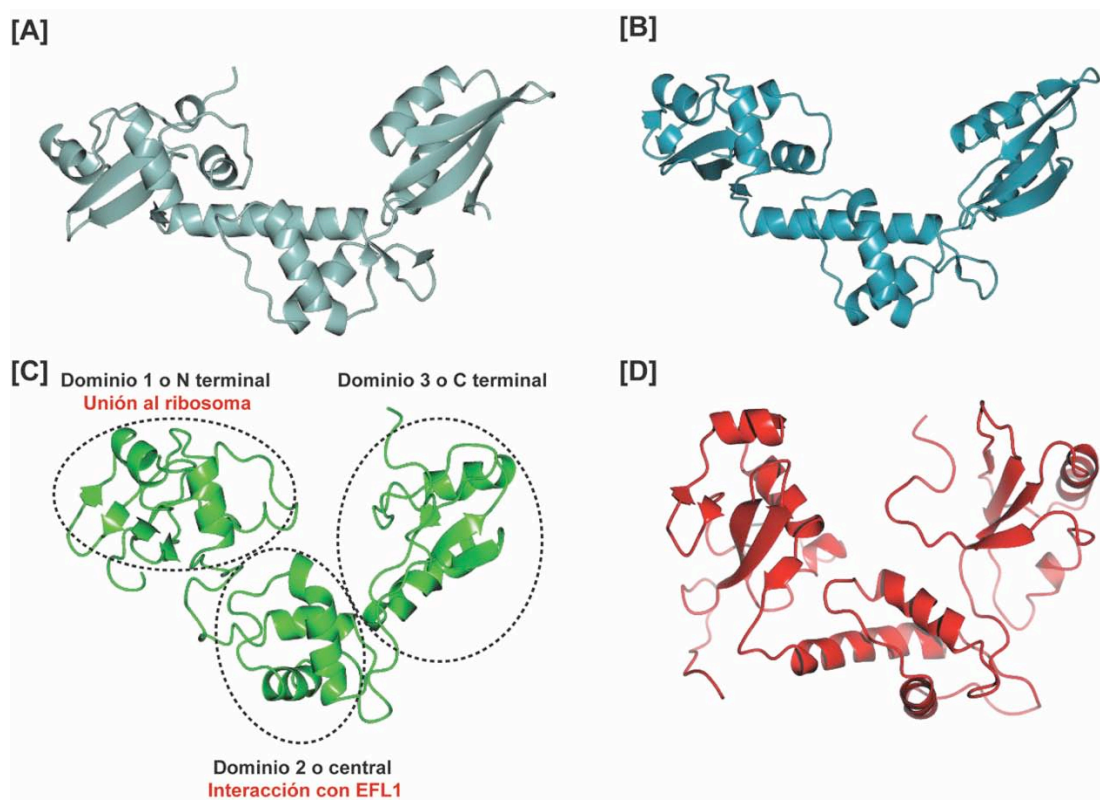


Figura 5. La estructura tridimensional de las proteínas SBDS se encuentra conservada entre miembros del dominio Archaea y Eukarya. [A] Estructura cristalográfica de la proteína SBDS de *Methanothermobacter thermautotrophicus* (PDB ID:4WBM), [B] Estructura cristalográfica de la proteína SBDS de *Archaeoglobus fulgidus* (PDB ID:1T95), [C] Estructura tridimensional de la proteína SBDS humana obtenida por RMN (PDB ID:2KDO) [D] Modelo tridimensional de la proteína Sdo1 de *Saccharomyces cerevisiae* obtenido por homología empleando como moldes las estructuras A, B y C.

1.3.2 La función de las proteínas SBDS es esencial

Diversos experimentos genéticos han revelado que las células de levadura haploides *sdo1* Δ tienen un defecto severo de crecimiento [83], y las deleciones bialélicas en ratón son embrionicamente letales [88]. Se ha comprobado que la

pérdida de función de la proteína HsSBDS es el principal origen del Síndrome Shwachman-Diamond (SDS por sus siglas en inglés) [89,90]. Esta es una enfermedad autosómica recesiva, clasificada como ribosomopatía, que se caracteriza por insuficiencia pancreática exocrina, alteraciones esqueléticas, deficiencia en la función de la médula ósea y una alta predisposición a desarrollar leucemia [52,91]. El 90 % de los pacientes con SDS presenta mutaciones en el gen *SBDS*. La mayoría de las mutaciones introducen un codón de paro prematuro que resulta en una proteína trunca; también se han descrito otras mutaciones no sinónimas en el dominio N-terminal y central, por lo que varios investigadores sugieren que estos cambios podrían perturbar tanto la interacción con la subunidad ribosomal mayor como con la GTPasa EFL1, provocando que el proceso de maduración ribosomal de la subunidad 60S no ocurra adecuadamente [83,85,89,90,92].

Actualmente, existen varios grupos de investigación trabajando para descubrir nuevos genes blanco que sean capaces de explicar el desarrollo del síndrome en aquellos pacientes que no presentan defectos en *SBDS*; dada la relación intrínseca que tiene con HsSBDS, diversos investigadores han propuesto a la proteína EFL1 humana (HsEFL1) como uno de los candidatos principales [93,94].

1.4 Las proteínas EFL1

1.4.1 Características generales

La proteína Efl1 de *S. cerevisiae* (ScEfl1), también conocida en mamíferos como EFTUD1, es una GTPasa citoplasmática que presenta 28% de identidad con los factores de traducción EF-2 (eEF-2 para eucariontes y aEF-2 para arqueas) y EF-G (bacteria) [78]. El alineamiento entre secuencias de aminoácidos de ScEfl1 y Efl-2 de levadura (ScEfl-2) sugiere que la proteína de maduración ribosomal podría tener una organización estructural similar a estas translocasas: cinco dominios numerados del I-V (Figura 6). Estudios genéticos en *S. cerevisiae* han mostrado que, de manera análoga a un proceso de control de calidad, ScEfl1 prueba que componentes ribosomales como el tallo y el sitio P, hayan sido ensamblados adecuadamente y se encuentren íntegros [77,95]. La similitud estructural entre

ScEfl1 y ScEf-2, además de sugerir un origen evolutivo común, propone que ScEfl1, asistida por ScSdo1, promueve la liberación de Tif6 de la subunidad ribosomal 60S mediante un cambio conformacional similar al que realiza ScEf-2 durante la fase de translocación en la síntesis de proteínas [95–97].

1.4.2 Organización y función de los dominios de las proteínas EFL1

Estudios *in vivo* e *in vitro* de miembros de EFL1 de levadura y mamíferos, han revelado algunas características importantes acerca de la función de cada uno de los dominios estructurales que conforman a estas proteínas (Figura 6). Se ha observado que el dominio I (al que llamaremos dominio G en este trabajo) contiene un subdominio G encargado de reconocer específicamente e hidrolizar a la molécula de GTP. Este subdominio está compuesto por cinco motivos (representados como G1-G5) que están conservados en las P-loop GTPasas [98]. Cada uno de estos sitios presenta papeles diferentes; mientras que los motivos G4 y G5, reconocen específicamente la base de guanina, y el motivo G1, interactúa con los fosfatos α y β del GTP, los motivos G2 y G3, conocidos también como switch I y switch II, respectivamente, están implicados en el mecanismo de hidrólisis de GTP por estas enzimas.

Por otro lado, estudios de doble híbrido realizados en nuestro grupo de investigación han mostrado que el dominio II de ScEfl1 es el sitio principal de interacción con ScSdo1. El alineamiento de secuencias con ScEf-2 también revela que ScEfl1 presenta una serie de inserciones a lo largo de su secuencia; una de las más prominentes se encuentra embebida en este dominio. Estudios *in vitro* demostraron que esta región se encuentra desordenada y es esencial para la interacción entre ScEfl1 y ScSdo1 [87]. Finalmente, Becam *et al.* [78] revelaron que la integridad de la región C-terminal es esencial para la función de la GTPasa; y estudios por criomicroscopia electrónica mostraron que ScEf-2 contacta a la subunidad ribosomal mayor a través de los dominios IV y V [97,99], lo anterior sugiere fuertemente que Efl1 podría estar utilizando las mismas regiones para interactuar con la subunidad ribosomal 60S.

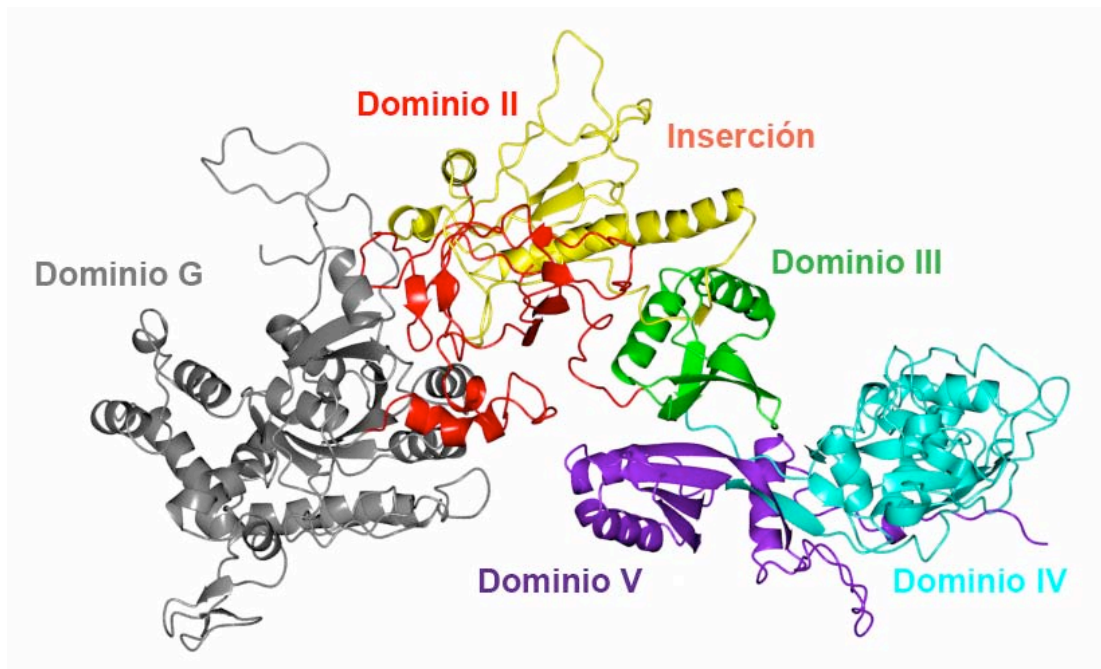


Figura 6. Modelo tridimensional de ScEfl1. El modelo fue obtenido por homología empleando MODELLER y la estructura cristalográfica de ScEf-2 como molde (PDB ID: 1N0V), así como Rosetta *Ab initio* para modelar la inserción [100]. De acuerdo con nuestro modelo, ScEfl1 se organiza en un dominio catalítico (dominio G) dentro del dominio I (gris), y cuatro dominios más numerados del II al V. El dominio II (rojo) contiene una región intrínsecamente desordenada (amarilla) implicada en la interacción con ScSdo1.

2 Justificación

El estudio de las proteínas EFL1 de *S. cerevisiae* y *H. sapiens*, ha llamado la atención de diversos investigadores en el mundo en los últimos años; sus resultados han permitido caracterizar aspectos importantes sobre su función y su estructura [93]; sin embargo, aún no se ha logrado comprender completamente cómo llevan a cabo su papel en el proceso de biogénesis ribosomal y su posible relación con el SDS. Su estudio, desde una perspectiva evolutiva, aportará información relevante respecto a su origen, su especialización y su función. Algunos investigadores han propuesto que estas proteínas son exclusivas de eucariontes y que su función está conservada evolutivamente [79,101]. Sin embargo, no existen investigaciones que aclaren completamente esta cuestión. Por lo tanto, para responder lo anterior, en este trabajo decidimos buscar ortólogos de EFL1, caracterizar su distribución filogenética, identificar características estructurales evolutivamente conservadas que podrían ser relevantes para su plegamiento y función, y evaluar si la función de estas proteínas se encuentra conservada entre especies. Además, a partir de ensayos de complementación funcional empleando pares de proteínas EFL1/SBDS buscaremos revelar aspectos importantes sobre el fenómeno de reconocimiento entre ambas proteínas. Finalmente, para obtener información acerca del proceso evolutivo de las proteínas EFL1 y la ruta de maduración ribosomal en la que participa, evaluaremos si las proteínas EF-2 y SbdS de la arquea *A. fulgidus*, son capaces de complementar la función de las proteínas Efl1 y Sdo1 en *S. cerevisiae*.

3 Hipótesis

- Si la familia de proteínas EFL1 se encuentra altamente conservada en el dominio Eukarya, entonces sus ortólogos correspondientes deben presentar una función conservada que puede ser evaluada mediante complementación funcional en levadura.
- Si las proteínas EFL1 y SBDS trabajan de forma conjunta en el mismo paso de la biogénesis ribosomal, entonces es posible que hayan co-evolucionado de manera tal que funcionen como un par ortogonal cuya función puede ser evaluada mediante doble complementación funcional en levadura.

4 Objetivos

4.1 Objetivo general

- Evaluar la conservación evolutiva y funcional de la familia de proteínas EFL1 y SBDS.

4.2 Objetivos particulares

- Buscar ortólogos de EFL1 y caracterizar su distribución filogenética de la mediante la construcción de un perfil filogenético.
- Identificar residuos altamente conservados en la familia EFL1 a partir del análisis de alineamientos múltiples de secuencias.
- Estudiar la relación filogenética entre las GTPasas EFL1, EF-2 y EF-G, mediante la construcción de un árbol filogenético.
- Construir las cepas de *S. cerevisiae* *efl1Δ/EFL1* y *efl1Δ/EFL1, sdo1Δ/SDO1*.
- Construir los vectores de expresión para las proteínas EFL1 y SBDS, necesarios para los ensayos de complementación.
- Construir los plásmidos para realizar los ensayos de complementación genética con los genes correspondientes a los ortólogos de diferentes organismos de EFL1.
- Evaluar la complementación genética de los ortólogos EFL1 mediante selección de la progenie meiótica viable en *S. cerevisiae efl1Δ*.
- Construir los plásmidos para realizar los ensayos de complementación genética con los genes correspondientes a los ortólogos de diferentes organismos de SBDS.
- Evaluar la complementación genética de los ortólogos SBDS mediante selección de la progenie meiótica viable en *S. cerevisiae sdo1Δ*.

- Evaluar la doble complementación genética de los diferentes ortólogos de EFL1 y SBDS mediante selección de la progenie meiótica viable en *S. cerevisiae efl1Δ/sdo1Δ*.
- Evaluar la doble complementación genética del factor de elongación EF-2 y SBDS de *Archaeoglobus fulgidus* mediante selección de la progenie meiótica viable en *S. cerevisiae efl1Δ/sdo1Δ*.

5 Materiales y métodos

5.1 Estudio bioinformático de las proteínas EFL1

Perfil filogenético de la familia de proteínas EFL1

Con el objetivo de identificar a los miembros de la familia EFL1 y observar su distribución filogenética, el primer paso fue construir un perfil de HMMER (pHMM) con las secuencias de aminoácidos de EFL1 correspondientes a *Arabidopsis thaliana*, *Caenorhabditis elegans*, *Dictyostelium discoideum*, *Homo sapiens*, *Mus musculus*, *Ratus norgevus*, *Gallus gallus*, *Saccharomyces cerevisiae*, *Schizosaccharomyces pombe* y *Drosophila melanogaster*. Las secuencias fueron obtenidas de la Princeton Protein Orthology Database (P-POD) utilizando como entrada el nombre de la proteína EFL1 de *S. cerevisiae* (Ria1) [5] y el perfil fue construido usando el programa HMMER 3.1b2 [102]. En seguida, se realizó una búsqueda contra la base de datos de UniprotKB empleando como consulta el pHMM; este paso fue realizado con la herramienta *hmmsearch* del servidor HMMER [103,104]. Los resultados de la búsqueda fueron analizados con el programa OrthoMCL [33] para asignar las secuencias a diferentes grupos de ortólogos. Por último, el grupo en donde se encontraban los miembros de EFL1 conocidos (marcadores de ortología), fue analizado manualmente mediante la construcción de árboles filogenéticos; las secuencias que se agruparon con los miembros de EFL1 conocidos, fueron seleccionadas para integrar nuestro grupo de estudio. La información taxonómica de cada miembro recopilado fue obtenida a través de UniprotKB [105].

Identificación de los residuos evolutivamente conservados en la familia de proteínas EFL1

Para identificar sitios evolutivamente conservados entre los miembros de la familia de proteínas EFL1, seleccionamos 68 miembros que representan su diversidad filogenética dentro del dominio Eukarya y alineamos sus secuencias de aminoácidos [106]. Los alineamientos fueron obtenidos empleando los programas

T-COFFEE [107] y PRALINE [108], y analizados con el software Jalview 2.4 [109]. Los residuos invariables, los sitios con un alto índice de conservación y la estructura secundaria predicha con el módulo de PSIPRED de PRALINE, se marcaron sobre la secuencia de aminoácidos y en el modelo tridimensional de ScEfl1. Este modelo fue generado por homología en trabajos anteriores, usando como molde la estructura cristalográfica de la proteína ScEf-2 (PDB ID 1N0V) y los programas Rosetta y MODELLER [110].

Análisis filogenético de los factores de elongación EFL1, EF-2 y EF-G

Se ha sugerido que las proteínas EFL1 están relacionadas evolutivamente con la familia de proteínas EF-2 y EF-G. Para obtener información importante sobre su proceso evolutivo, decidimos estudiar su relación filogenética empleando las secuencias de aminoácidos correspondientes a las proteínas EFL1, EF-2, EF-G y EF-4 (grupo externo). Las proteínas EFL1 utilizadas para este análisis corresponden con las reportadas en P-POD, mientras que las secuencias de eEF2, aEF2, EF-G y EF-4 (lepA) seleccionadas, se identificaron durante la primera búsqueda empleando el pHMM. Las secuencias de todas las proteínas se recopilaron de la base de datos de UniprotKB después de asegurarnos que habían sido manualmente anotadas y revisadas (Tabla Suplementaria 1). Para la construcción del árbol filogenético, las secuencias fueron alineadas con el programa MAFFT usando el método de refinamiento iterativo E-INS-i [111]. El árbol fue inferido con el programa PhyML del servidor Trex-online [26] usando como matriz de sustitución Blosum62, la cual fue seleccionada por medio del servidor ProtTest 2.4 [112] y se evaluó mediante Bootstrapping (100 replicas). Finalmente, el árbol fue visualizado y editado con el software FigTree 1.4.2.

Caracterización de inserciones en miembros de la familia EFL1

Los dominios estructurales en los que están organizados las proteínas EFL1 de *H. sapiens*, *S. pombe*, *D. discoideum*, *A. thaliana* y *P. falciparum*, fueron delimitados a través de alineamientos de estructura primaria con ScEfl1. Posteriormente, la presencia de inserciones en los miembros de EFL1 se caracterizó a partir de alineamientos de cada una de sus secuencias, con la correspondiente a su

proteína EF-2. Todos los alineamientos se realizaron empleando el programa T-COFFEE.

5.2 Evaluación funcional de ortólogos de EFL1 y SBDS

5.2.1 Obtención de las cepas diploides heterocigotas para los ensayos de complementación genética funcional

Ensayos de complementación genética en *S. cerevisiae*

Obtención de la cepas *S. cerevisiae efl1Δ/EFL1* y *efl1Δ/EFL1 sdo1Δ/SDO1*

La cepa diploide de *S. cerevisiae* AM1 (*efl1Δ::natMX4/EFL1*) se obtuvo mediante la sustitución de uno de los alelos de *EFL1* por el módulo de resistencia a cloNAT, natMX4 (Figura 7A). El módulo se amplificó por PCR con el PCR Master Mix 2x (Fermentas), usando el vector p4339 [43,113] como molde y los oligonucleótidos F-EFL1::Nat y R-EFL1::Nat. Estos oligonucleótidos añaden en los extremos del amplicón 50 pb idénticos a las secuencias regulatorias río arriba y río abajo del gen *EFL1*. En seguida, se transformaron células competentes de la cepa diploide *S. cerevisiae* Y5538 (Tabla 1) con el producto de la reacción empleando el protocolo descrito en el Apéndice 7. Las clonas recombinantes se seleccionaron en medio YPD + cloNAT (Tabla Suplementaria 9). Posteriormente, se extrajo DNA genómico (gDNA) de diferentes clonas candidatas siguiendo el protocolo descrito en el

Apéndice 1 y mediante PCR se comprobó 1) la sustitución de uno de los alelos *EFL1* por el módulo de resistencia usando los oligonucleótidos EFL1+500 pb y R-EFL1::Nat, así como 2) la presencia de un alelo *EFL1* integro empleando los oligonucleótidos ScFEFL1 y ScREFL1. La composición y condiciones de incubación de las reacciones de amplificación se muestran en las Tablas 2-5. Las secuencias de los oligonucleótidos utilizados se encuentran en la Tabla Suplementaria 5.

Tabla 1. Genotipos relevantes de las cepas de *S. cerevisiae* empleadas en este trabajo.

Cepa	Genotipo relevante	Referencia
Y5538	S288c, MATa/α, lypΔ/+, mfa1Δ::MFα1pr-LEU2/+; can1Δ::MFA1pr-HIS3/+, his3Δ1/his3Δ1, leu2Δ0/leu2Δ0, ura3Δ0/ura3Δ0, met15Δ0/met15Δ0	[113,114]
AM1	S288c, MATa/α, lypΔ/+, mfa1Δ::MFα1pr-LEU2/+; can1Δ::MFA1pr-HIS3/+, efl1Δ::natMX4/EFL1, his3Δ1/his3Δ1, leu2Δ0/leu2Δ0, ura3Δ0/ura3Δ0, met15Δ0/met15Δ0	Este trabajo
AM2	S288c, MATa/α, lypΔ/+, mfa1Δ::MFα1pr-LEU2/+; can1Δ::MFA1pr-HIS3/+, efl1Δ::natMX4/EFL1, sdo1Δ::kanMX6/SDO1 his3Δ1/his3Δ1, leu2Δ0/leu2Δ0, ura3Δ0/ura3Δ0, met15Δ0/met15Δ0	Este trabajo

Tabla 2. Composición de la PCR empleada para la amplificación del módulo de resistencia *NatMX4*.

Componente	Cantidad
PCR Master Mix 2x	25 μL
F-EFL1::Nat (10 μM)	2 μL
R-EFL1::Nat (10 μM)	2 μL
P4339 (10 ng/μL)	2 μL
DMSO	2.5 μL
Agua	16.5 μL
Volumen total	50 μL

Tabla 3. Condiciones de incubación de las PCR empleadas para amplificar el módulo de resistencia *NatMX4*.

Número de ciclos	Temperatura	Tiempo
1	95 °C	3 min
30	95 °C	30 s
	53 °C	1 min
	72 °C	1 min 10 s
1	72 °C	10 min
1	4 °C	∞

Tabla 4. Composición de las PCR empleadas para corroborar la integración del módulo *NatMX4* en el locus *EFL1* y la presencia de un alelo *EFL1* integro en la cepa de levadura de interés.

Componente	Cantidad
PCR Master Mix 2x	12.5 μ L
Oligonucleótido sentido (10 μ M)	1 μ L
Oligonucleótido reverso (10 μ M)	1 μ L
gDNA	1 μ L
DMSO	1.25 μ L
Agua	8.75 μ L
Volumen total	25 μL

Tabla 5. Condiciones de incubación de las PCR empleadas para amplificar el módulo de resistencia *NatMX4*, confirmar su integración en el DNA genómico de *S. cerevisiae* y comprobar la integridad de un alelo *EFL1*.

Número de ciclos	Temperatura	Tiempo
1	95 °C	3 min
30	95 °C	30 s
	55 °C	1 min
	72 °C	*
1	72 °C	10 min
1	4 °C	∞

* La Taq polimerasa contenida en el PCR Master Mix 2x amplifica 1 kb/min, por lo tanto, el tiempo de extensión para amplificar el módulo de resistencia más 500 pb río arriba (~1700 pb) y la secuencia de *EFL1* (3300 pb) fue de 1 min 45 s y 3 min 30 s, respectivamente.

La cepa *S. cerevisiae* AM2 (*efl1 Δ ::natMX4/EFL1*, *sdol1 Δ ::kanMX4/SDO1*) se generó a partir de la sustitución de uno de los alelos *SDO1* por el módulo de resistencia a G418, *kanMX6* (Figura 7B). El módulo fue amplificado usando el PCR Master Mix (Fermentas), usando el vector pFA6a-kanMX6 como molde y los oligonucleótidos F-SDO1::Kan y R-SDO1::Kan. Estos oligonucleótidos añaden en sus extremos 20 pb de bases idénticos a las secuencias regulatorias río arriba y río abajo del gen *SDO1*. En seguida, células competentes de la cepa AM1 (Tabla 1) se transformaron con el producto de la PCR siguiendo el protocolo descrito en el Apéndice 7. Las posibles clonas recombinantes se seleccionaron en medio YPD + G418 + cloNAT (Tabla Suplementaria 9). Para verificar la sustitución de uno de

los alelos *SDO1* por el módulo *kanMX6*, así como la presencia de un alelo *SDO1* integro, se obtuvo gDNA de diferentes clonas candidatas siguiendo el protocolo descrito en el Apéndice 1. Se amplificó 1) el módulo *kanMX6* más 100 pb rio arriba usando los oligonucleótidos SDO1-100pb y R-SDO1::Kan, así como 2) la secuencia codificante de *SDO1* empleando los oligonucleótidos FScSDO1 y RScSDO1. La composición y condiciones de incubación de las reacciones de amplificación se muestran en las Tablas 6-9. Las secuencias de los oligonucleótidos utilizados se encuentran en la Tabla Suplementaria 6.

Tabla 6. Composición de las PCR empleadas para la amplificación del módulo de resistencia *KanMX4*.

Componente	Cantidad
PCR Master Mix 2x	25 μ L
F-SDO1::Kan	2 μ L
R-SDO1::Kan	2 μ L
pFA6a- <i>kanMX6</i> (10 ng/ μ L)	2 μ L
DMSO	2.5 μ L
Agua	16.5 μ L
Volumen total	50 μL

Tabla 7. Condiciones de incubación de las PCR empleadas para amplificar el módulo de resistencia *KanMX4*.

Número de ciclos	Temperatura	Tiempo
1	95 °C	3 min
30	95 °C	30 s
	55 °C	1 min
	72 °C	1 min 30 s
1	72 °C	10 min
1	4 °C	∞

Tabla 8. Composición de las PCR empleadas para corroborar la integración del módulo *kanMX4* en el locus *SDO1*.

Componente	Cantidad
PCR Master Mix 2x	12.5 μ L
Oligonucleótido sentido (10 μ M)	1 μ L
Oligonucleótido reverso (10 μ M)	1 μ L
gDNA	1 μ L
DMSO	1.25 μ L
Agua	8.75 μ L
Volumen total	25 μL

Tabla 9. Condiciones de incubación de las PCR empleadas para amplificar el módulo de resistencia *natMX4*, confirmar su integración en el DNA genómico de *S. cerevisiae*, así como la integridad de un alelo *SDO1*.

Número de ciclos	Temperatura	Tiempo
1	95 °C	5 min
30	95 °C	30 s
	55 °C	1 min
	72 °C	*
1	72 °C	10 min
1	4 °C	∞

* La Taq polimerasa contenida en el PCR Master Mix 2x amplifica 1 kb/min, por lo tanto, el tiempo de extensión para amplificar el módulo de resistencia más 100 pb río arriba (1600 pb) y la secuencia de *SDO1* (750 pb), fue de 1 min 40 s y 45 s, respectivamente.

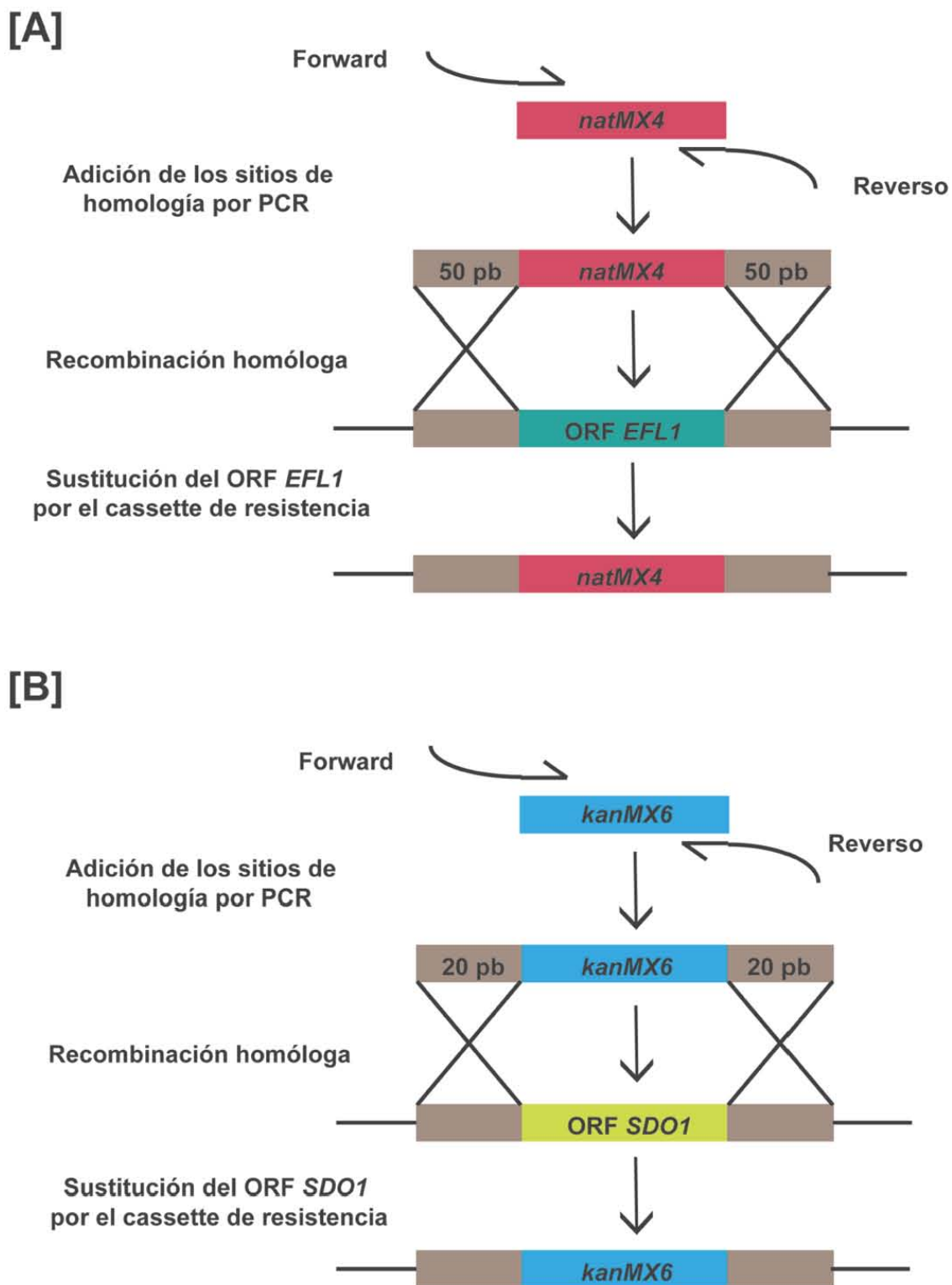


Figura 7. Estrategia experimental implementada para reemplazar las secuencias codificantes de *EFL1* con el módulo de resistencia a cloNAT (*natMX4*) [A] y de *SDO1* con el módulo de resistencia a G418 (*kanMX6*) [B].

5.2.2 Vectores de expresión

Construcción del vector pAM1

El vector pAM1 (Figura 8) empleado para la expresión de los miembros de la familia EFL1 se construyó mediante la adición de un módulo de expresión al vector pRS316. El módulo de expresión se sintetizó y clonó en el vector pUC57 por la compañía GenScript USA, Inc. Este fragmento de DNA comprende la secuencia del promotor constitutivo de la Alcohol Deshidrogenasa 1 (-750 pb, *pr-ADH1*), un sitio de clonación múltiple seguido de la secuencia codificante para la etiqueta V5 (GKPIPPLLGLDST), un codón de termino ocre (TAA) y la secuencia del terminador de la ADH1 (*ter-ADH1*).

Tanto la construcción sintetizada por GenScript, como el pRS316, se digirieron con las enzimas de restricción *KpnI* (New England Biolabs) y *SacI* (New England Biolabs). Ambas reacciones de restricción se resolvieron en un gel de agarosa al 1%; los fragmentos correspondientes al módulo de expresión (1.2 kb) y pRS316 (4.8 kb), se escindieron del gel y purificaron con el kit Agarose Gel Extraction (Jena Bioscience). Posteriormente, los fragmentos correspondientes al inserto y al vector se ligaron en una relación 3:1 durante 18 h a 8 °C empleando T4 DNA ligasa (New England Biolabs). Al finalizar el tiempo de incubación, 50 µL de células quimiocompetentes de *E. coli* DH10B se transformaron con 8 µL de reacción de ligación siguiendo el protocolo descrito en el Apéndice 5. Las clonas transformantes se seleccionaron en medio LB + Ampicilina y el DNA plasmídico (pDNA) de cada una de las clonas candidatas se extrajo empleando el método EasyPrep (Apéndice 3). Finalmente, cada plásmido se restringió con *KpnI* y *SacI* para identificar las construcciones de interés.

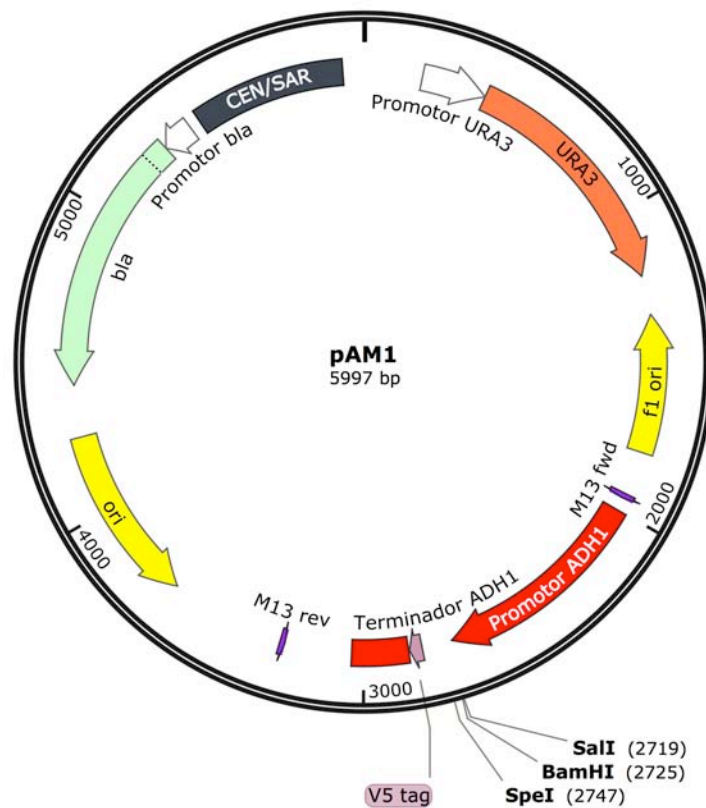


Figura 8. Características generales del vector de expresión centromérico pAM1.

Construcción del vector pAM2

El vector pAM2 (Figura 9) empleado para expresar los miembros de la familia SBDS se construyó a partir de la adición de un módulo de expresión al pRS411 (Figura 10). El módulo, al igual que en el pAM1, comprende la secuencia del promotor constitutivo y terminador de la ADH1, un sitio de clonación múltiple y un codón de termino ocre. La principal diferencia es que mientras el módulo en el pAM1 contiene la secuencia que codifica para la etiqueta V5, el módulo en el pAM2 contiene la secuencia que codifica la etiqueta c-Myc (EQKLISEEDL).

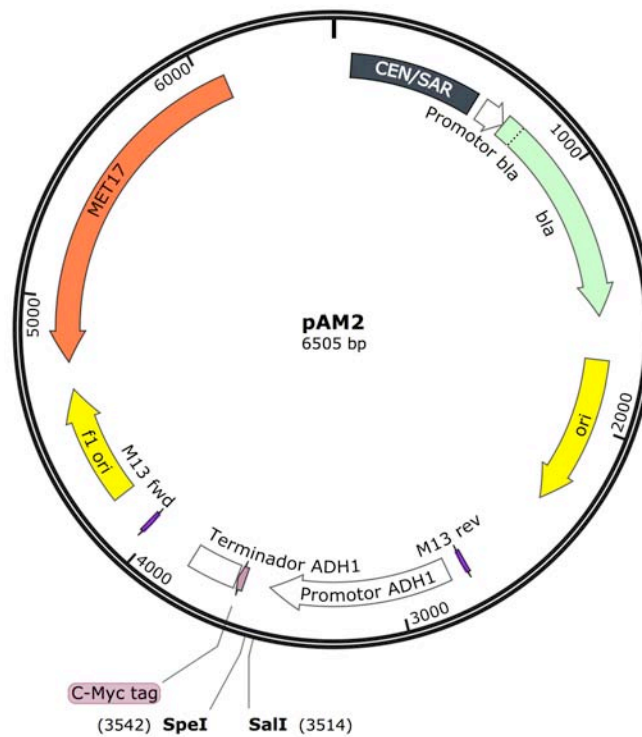


Figura 9. Características generales del vector de expresión centromérico pAM2.

El módulo de expresión se sintetizó por PCR de traslape empleando el PCR Master Mix 2X (Fermentas) y el vector pAM1 como molde (Figura 10). El *pr-ADH1* y el sitio de clonación múltiple se amplificaron con los oligonucleótidos M13-Forward y R-Cmyc (fragmento P1), mientras que el *ter-ADH1* se amplificó con los oligonucleótidos M13-Reverse y el F-Cmyc (fragmento P2). Ambos productos de PCR contienen la secuencia codificante de la etiqueta c-Myc como un sitio de traslape, lo que permitió ensamblar ambos fragmentos a partir de una tercera PCR empleando los oligonucleótidos M13-Forward y M13-Reverse. La composición y condiciones de incubación para cada una de las reacciones se describen en las Tablas 10-12. Las secuencias de los oligonucleótidos se encuentran en la Tabla Suplementaria 2.

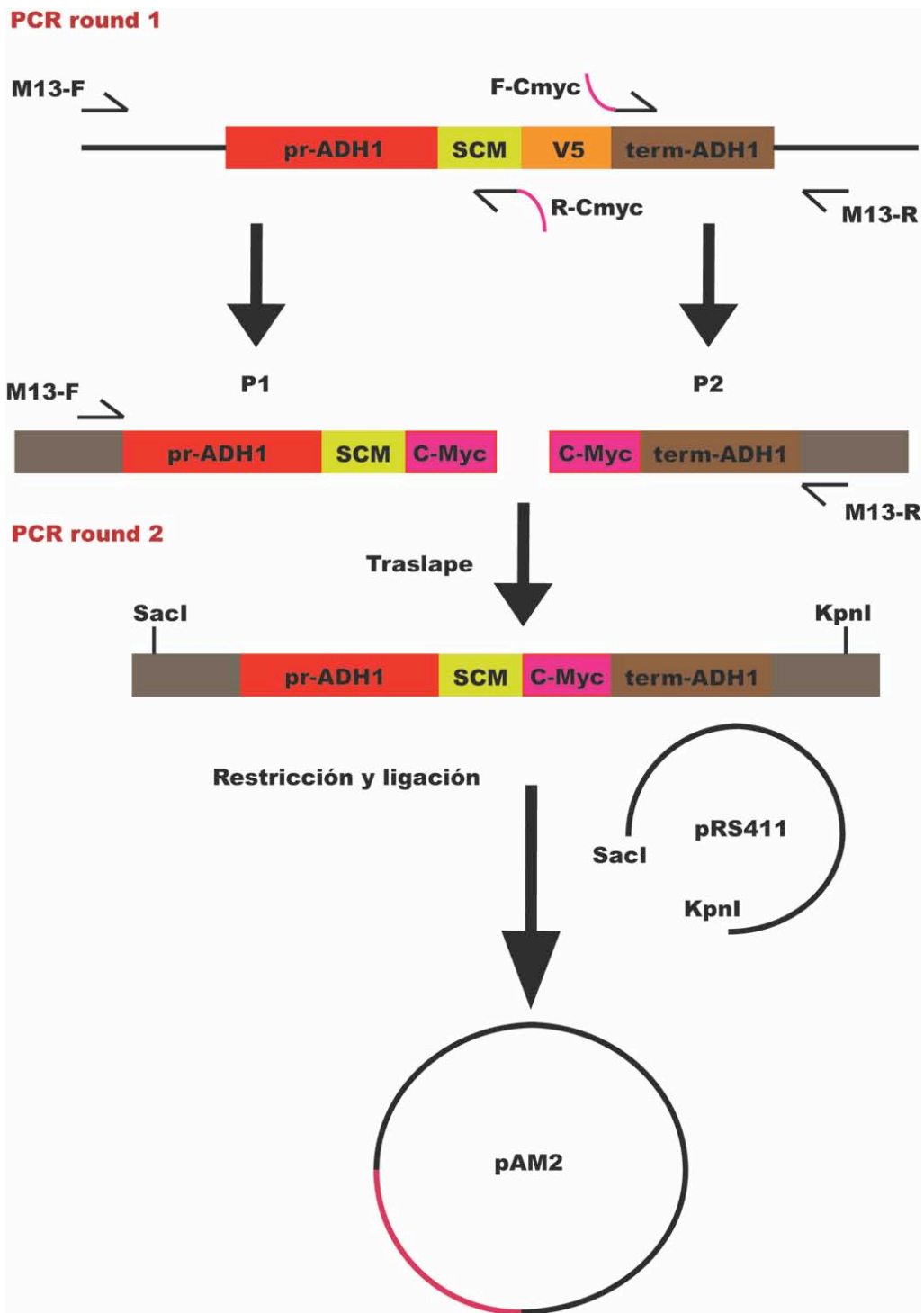


Figura 10. Estrategia implementada para construir el vector de expresión pAM2. pr-ADH1: promotor Alcohol Deshidrogenasa 1. SCM: sitio de clonación múltiple. C-Myc: Etiqueta para detección de la proteína por Inmunoblot. term-ADH1: terminador de la alcohol deshidrogenasa 1. P1: Producto de PCR 1, P2: Producto de PCR 2.

Una vez amplificado el modulo completo, se purificó con el kit PCR Purification (Jena Bioscience). A continuación, el amplicón y el vector pRS411 se digirieron con las enzimas *KpnI* y *SacI* (New England Biolabs). Las reacciones de restricción se resolvieron en gel de agarosa al 1% y los fragmentos correspondientes al inserto (1.2 kb) y vector (4.5 kb) se escindieron y purificaron de gel usando el kit Gel Extraction (Jena Bioscience). Posteriormente, ambos fragmentos se ligaron en una relación 3:1 durante 18 h a 8 °C empleando T4 DNA ligasa (New England Biolabs). Al terminar el periodo de incubación, 50 µL de células quimiocompetentes de *E. coli* DH10B se transformaron con 8 µL de reacción de ligación siguiendo el protocolo descrito en el Apéndice 5. Las células transformantes se seleccionaron en medio LB + ampicilina (Tabla Suplementaria 8). Finalmente, se obtuvo pDNA de cada una de las clonas trasformantes empleando el método de EasyPrep (Apéndice 3) y se digirió con las enzimas de restricción *KpnI* y *SacI* para identificar las clonas con la construcción de interés.

Tabla 10. Composición de las PCR empleadas para la amplificación de los fragmentos P1 y P2.

Componente	Cantidad
PCR Master Mix 2x	12.5 µL
Oligonucleótido sentido (10 µM)	1 µL
Oligonucleótido reverso (10 µM)	1 µL
pAM1 (10 ng/µL)	2 µL
Agua libre de nucleasas	8.5 µL
Volumen total	25 µL

Tabla 11. Composición de las PCR empleadas para la amplificación del módulo de expresión del pAM2.

Componente	Cantidad
PCR Master Mix 2x	25 µL
M13-Forward (10 µM)	2 µL
M13-Reverso (10 µM)	2 µL
P1 (10 ng/µL)	1 µL
P2 (10 ng/µL)	3 µL
Agua libre de nucleasas	17 µL
Volumen total	50 µL

Tabla 12. Condiciones de incubación de las PCR para construir el módulo de expresión del pAM2.

Número de ciclos	Temperatura	Tiempo
1	95 °C	3 min
30	95 °C	30 s
	55 °C	45 s
	72 °C	*
1	72 °C	10 min
1	4 °C	∞

* La Taq polimerasa contenida en el PCR Master Mix 2x amplifica 1 kb/min, por lo tanto el tiempo de extensión para amplificar el fragmento P1 (~ 900 pb), P2 (400 pb) y módulo completo (1300 pb) fue de 25 s, 54 s y 1 min 20 s, respectivamente.

5.2.3 Construcciones de los miembros de EFL1 para los ensayos de complementación funcional

Obtención de las secuencias codificantes de los miembros EFL1

Todas las secuencias codificantes se obtuvieron por PCR empleando el PCR Master Mix 2x (Fermentas), excepto la secuencia codificante de EF-2 de *A. fulgidus* (*AfEF2*) que se sintetizó con un uso de codones optimizado para su expresión en *S. cerevisiae* y clonó en el pUC57 por la empresa GenScript USA Inc. (<http://www.genscript.com>).

La secuencia codificante de *S. cerevisiae* (*ScEFL1*) se amplificó usando el plásmido pRS426-EFL1 como molde y los oligonucleótidos FScEFL1 y RScEFL1, con los que se adicionó en los extremos del amplicón un sitio *Sall* y *SpeI* (en dirección 5'-3'). Para amplificar la secuencia de EFL1 isoforma 1 (*HsEFL1-1*) e isoforma 2 (*HsEFL1-2*) de *H. sapiens* se utilizaron como moldes las construcciones pRS426-EFTUD1-1 y pRS426-EFTUD1-2, respectivamente, y los oligonucleótidos FHsEFL1 y RHsEFL1, con los que se flanqueó cada uno de los productos de PCR con los sitios de reconocimiento por las enzimas *Sall* y *BamHI*. Para la amplificación de la secuencia de *M. musculus* (*MmEFL1*), utilizamos como molde el pDNA de una clona comercial (Clone ID: 4983234, Dharmacon, GE Healthcare) y los oligonucleótidos FMmEFL1 y RMmEFL1. Por otro lado, para obtener la secuencia codificante de *S. pombe* (*SpEFL1*) se usó como molde cDNA

producido a partir de RNA total aislado de células en fase de crecimiento exponencial (Apéndice 2), y los oligonucleótidos FSpEFL1 y RSpEFL1. A los amplicones *MmEFL1* y *SpEFL1*, además de adicionárseles los sitios *Sall* y *SpeI*, también se les adicionó 17 pb idénticos a las secuencias que flanquean el sitio de inserción en el pAM1 útiles para su clonación por recombinación homóloga. La composición de las reacciones de PCR, así como sus condiciones de incubación, se describen en las Tablas 13-14 y las características de los oligonucleótidos empleados se describen en la Tabla Suplementaria 3.

Tabla 13. Composición de las PCR empleadas para la amplificación de las secuencias codificantes de los miembros de EFL1 evaluados en este trabajo.

Componente	Cantidad
PCR Master Mix 2x	25 μ L
Oligonucleótido sentido (10 μ M)	2 μ L
Oligonucleótido reverso (10 μ M)	2 μ L
Molde	1 μ L
Agua libre de nucleasas	17 μ L
Volumen total	50 μL

Tabla 14. Condiciones de incubación de las PCR para amplificar las secuencias codificantes de los miembros de la familia EFL1 evaluados en este trabajo.

Número de ciclos	Temperatura	Tiempo
1	95 °C	5 min
30	95 °C	30 s
	55 °C	1 min
	72 °C	3 min 40 s
	72 °C	10 min
1	4 °C	∞

Clonación de las secuencias codificantes de los miembros de la familia EFL1 en el vector de expresión pAM1

Para obtener la construcción pAM1-ScEFL1 y pAM1-AfEF2, la secuencia codificante de la proteína ScEFL1, la construcción pUC57-AfEF2 y el pAM1 fueron digeridos con las enzimas de restricción *Sall* y *SpeI* (Fermentas) por 12 h a 37 °C. Por otro lado, para clonar las secuencias HsEFL1-1 y HsEFL1-2 en el pAM1,

vector e insertos fueron digeridos con la enzima *Sall* durante 12 h. Posteriormente los fragmentos de DNA se purificaron siguiendo el protocolo del Apéndice 4 y digeridos 6 h con *BamHI* (Fermentas). Las reacciones de restricción se resolvieron en un gel de agarosa al 1% y los fragmentos correspondientes a los insertos y vector fueron escindidos del gel y purificados usando el kit Agarose Gel Extraction (Jena Bioscience). Posteriormente se ligaron durante 18 h a 8 °C, en una relación 2:1 usando T4 DNA ligasa (New England Biolabs). Para obtener las construcciones pAM1-MmEFL1 y pAM1-SpEFL1 los amplicones correspondientes se purificaron con el kit PCR Purification (Jena Bioscience) y clonaron por recombinación en el pAM1 previamente digerido con *Sall* y *SpeI* usando el kit CloneEZ PCR Cloning (GenScript USA, Inc.) siguiendo las condiciones del proveedor.

Finalmente, se tomaron 10 µL de cada reacción de ligación y recombinación para transformar 50 µL de células quimiocompetentes de *E. coli* DH10B siguiendo el protocolo descrito en el Apéndice 5. Las células transformantes se seleccionaron en medio LB + ampicilina (Tabla Suplementaria 8) y de cada una de las clonas se extrajo pDNA por el método de EasyPrep (Apéndice 3). Las construcciones de interés se identificaron mediante la digestión de cada pDNA con *Sall/SpeI* y *Sall/BamHI*. Las construcciones con el patrón de restricción esperado se secuenciaron (GenScript USA, Inc.) para confirmar que la secuencia correspondía al inserto de interés y que se encontraban integras (sin mutaciones).

Adición del codón de paro a las construcciones con los miembros de la familia EFL1 para los ensayos de complementación funcional

Evidencia experimental obtenida por nuestro laboratorio y otros estudios [78], sugiere que deleciones en la región C-terminal de las proteínas EFL1, así como la presencia de residuos extra, etiquetas o proteínas de fusión, pueden afectar su función. Por lo tanto, todas las construcciones descritas en el punto anterior, con excepción de la pAM1·AfEF2 (esta construcción se diseñó para expresar a la proteína de interés con la etiqueta V5 fusionada a la región N-terminal), se modificaron empleando un protocolo de mutagénesis sitio dirigida no convencional

[115] para adicionarles un codón de paro en el extremo C-terminal e impedir la traducción de la etiqueta V5.

Para modificar cada construcción, se realizaron dos PCR en paralelo empleando KOD polimerasa (EMD Millipore). La composición de cada reacción se muestra en la Tabla 15. Al terminar las reacciones de polimerización, los pares correspondientes fueron mezclados entre sí e incubados durante 12–14 h a 37 °C con 30 U de *DpnI* (Fermentas). Posteriormente, la mezcla se expuso a un gradiente de temperatura (Tabla 16) para promover la hibridación entre cadenas recién sintetizadas (modificadas). Se utilizaron 10 µL de cada mezcla para transformar 50 µL de células quimiocompetentes de *E. coli* DH10B siguiendo el protocolo descrito en el Apéndice 5. Las clonas candidatas se seleccionaron en medio LB + ampicilina (Tabla Suplementaria 8). Por último, la identificación de las construcciones de interés se realizó a partir de la purificación del pDNA extraído de cada clona candidata seguido de digestión con *HindIII* (Fermentas). La adición del codón de termino, así como la integridad de cada una de las secuencias codificantes, se confirmó por secuenciación (GenScript USA, Inc.). Los oligonucleótidos utilizados para agregar el codón de termino a cada una de las construcciones se muestran en la Tabla Suplementaria 7.

Tabla 15. Composición de las reacciones de polimerización para modificar las construcciones que expresan a los miembros EFL1.

Componente	Reacción sentido	Reacción antisentido
DNA molde	500 ng	500 ng
Oligonucleótido sentido (10 µM)	4 µL	-
Oligonucleótido antisentido (10 µM)	-	4 µL
dNTP 2 mM cada uno	2 µL	2 µL
Buffer KOD 10X	2.5 µL	2.5 µL
KOD polimerasa (2 U/µL)	0.5 µL	0.5 µL
Agua libre de nucleasas	*	*
Volumen final	25 µL	25 µL

*Cuanto baste para alcanzar el volumen final.

Tabla 16. Gradiente de temperaturas empleado para realinear las cadenas de DNA recién sintetizadas durante la modificación de las construcciones que expresan a los miembros EFL1.

Temperatura	Tiempo
95 °C	5 min
90 °C	2 min
80 °C	2 min
70 °C	2 min
60 °C	1 min
50 °C	1 min
40 °C	1 min
37 °C	1 min
4 °C	∞

5.2.4 Construcciones de los miembros de SBDS para los ensayos de complementación funcional

Obtención de las secuencias codificantes de los miembros de la familia SBDS

Las secuencias codificantes de los miembros SBDS se obtuvieron por PCR utilizando el PCR Master Mix 2x (Fermentas). Los oligonucleótidos empleados para su amplificación se muestran en la Tabla Suplementaria 4.

Los fragmentos que codifican Sdo1 de *S. cerevisiae* (ScSDO1), SBDS de *H. sapiens* (HsSBDS), y las quimeras HsSBDS·D2Sc y ScSDO1·D2Hs, se amplificaron empleando como moldes las construcciones (obtenidas anteriormente en nuestro grupo de investigación) pRSET-SDO1, pRSET-SBDS, pRSET-SBDS 2Y y pRSET-SDO1 2H, respectivamente. Para amplificar el marco abierto de lectura de *A. fulgidus* SBDS (AfSBDS) se empleó la construcción pET15b-AfSBDS, amablemente donada por el Dr. Alexei Savchenko de la Universidad de Toronto. Por otro lado, para obtener el fragmento codificante de SBDS de *M. musculus* (MmSBDS) usamos cDNA obtenido a partir de RNA total de hígado de ratón.

Por otra parte, la secuencia codificante de SDO1 de *S. pombe* (SpSDO1) se obtuvo por PCR de traslape utilizando como molde gDNA que se purificó

empleando el protocolo descrito en el Apéndice 1. Los exones 1 y 2 se amplificaron independientemente usando los pares de oligonucleótidos SpFSDO1/Rex1 SpSDO1, y Fex2 SpSDO1/SpRSDO1, respectivamente. A partir del reverso Rex1 SpSDO1, se generó un sitio de traslape entre el exón 1 y el exón 2, por lo tanto, el marco abierto de lectura completo se obtuvo en una tercera PCR usando los exones 1 y 2 como molde, y los oligonucleótidos SpFSDO1 y SpRSDO1. Todos los amplicones obtenidos en esta sección se flanquearon con los sitios *Sa*I y *Sp*I durante la PCR. La composición y condiciones de incubación de las reacciones se muestran en las Tablas 17-20.

Tabla 17. Composición de las PCR empleadas para la amplificación de las secuencias codificantes de los miembros de la familia SBDS.

Componente	Cantidad
PCR Master Mix 2x	25 μ L
Oligonucleótido sentido (10 μ M)	2 μ L
Oligonucleótido reverso (10 μ M)	2 μ L
Molde (Aprox. 20 ng/ μ L)	1 μ L
Agua libre de nucleasas	17 μ L
Volumen total	50 μL

Tabla 18. Condiciones de incubación de las PCR para amplificar las secuencias codificantes de los miembros de la familia SBDS.

Número de ciclos	Temperatura	Tiempo
1	95 °C	5 min
30	95 °C	30 s
	55 °C	1 min
	72 °C	50 s
	72 °C	10 min
1	4 °C	∞

Tabla 19. Composición de las PCR empleadas para la amplificación de los exones 1 y 2 de la secuencia que codifica SpSDO1.

Componente	Cantidad
PCR Master Mix 2x	25 μ L
Oligonucleótido sentido (10 μ M)	2 μ L
Oligonucleótido reverso (10 μ M)	2 μ L
DNA genómico (50 ng/ μ L)	1 μ L
Agua libre de nucleasas	17 μ L
Volumen total	50 μL

Tabla 20. Condiciones de incubación de las PCR para amplificar los exones 1 y 2 de la secuencia que codifica SpSDO1.

Número de ciclos	Temperatura	Tiempo
1	95 °C	5 min
30	95 °C	30 s
	55 °C	1 min
	72 °C	*
	72 °C	10 min
1	4 °C	∞

* La Taq polimerasa contenida en el PCR Master Mix 2x amplifica 1 kb/min, por lo tanto el tiempo de extensión para amplificar el exón 1 (130 pb) y el exón 2 (620 pb) fue de 15 s y 40 s, respectivamente.

Clonación de las secuencias codificantes de los miembros de la familia de SBDS en el vector de expresión pAM2

Las secuencias codificantes de cada uno de los miembros de SBDS se resolvieron en un gel de agarosa al 1%. La banda correspondiente a cada una de ellas se escindió y purificó empleando el kit Agarose Gel Extraction (Jena Bioscience). Posteriormente, los amplicones se clonaron en el pGEM T easy (Promega) en una relación vector inserto 1:3. Las reacciones se prepararon como se establece en el instructivo del producto y se incubaron durante 18 h a 8 °C. Terminado el tiempo de incubación, se transformaron 50 μ L de células quimiocompletentes de *E. coli* DH10B con 5 μ L de cada reacción de ligación. Las células transformantes se seleccionaron en LB + ampicilina (Tabla Suplementaria 8) y las clonas candidatas a portar el plásmido con el inserto de interés, se identificaron a partir del sistema

reportero de pérdida de función de *LacZ* (colonias blancas en presencia de x-gal). A continuación, se purificó pDNA a partir de las células transformantes blancas, siguiendo el protocolo de EasyPrep (Apéndice 3) y se digirió con la enzima *EcoRI*. Los plásmidos con el patrón de restricción esperado se secuenciaron (GenScript USA, Inc.) para verificar la integridad de cada uno de los insertos.

Para subclonar los fragmentos de interés en el pAM2, las construcciones de cada una de las secuencias codificantes de los miembros de SBDS en el pGEM-T, así como el pAM2 se digirieron con *Sall* y *SpeI* (Fermentas) durante 12 h a 37 °C. Posteriormente, las reacciones de restricción se resolvieron en un gel de agarosa al 1% y los fragmentos correspondientes a los insertos y al vector se escindieron del gel. En seguida, cada uno se purificó empleando el kit Agarose Gel Extraction (Jena Bioscience). Los insertos y el vector se ligaron durante 18 h a 8 °C en una relación 5:1 usando T4 DNA ligasa (New England Biolabs). Posteriormente se tomaron 5 µL de cada reacción de ligación para transformar 50 µL de células competentes de *E. coli* DH10B siguiendo el protocolo descrito en el **Apéndice 5**. Las células transformantes se seleccionaron en medio LB + ampicilina (Tabla Suplementaria 8) y de cada una de las clonas se extrajo pDNA por el método de EasyPrep (Apéndice 3). Las construcciones se identificaron a partir de la restricción de los plásmidos candidatos con *Sall* y *SpeI*.

5.2.5 Ensayos de complementación funcional en *S. cerevisiae*

Transformación de DNA plasmídico en células de *Saccharomyces cerevisiae*

Las células de levadura se transformaron con pDNA siguiendo el protocolo acetato de litio/acarreador de DNA/Polietilenglicol (Apéndice 7). Para evaluar la complementación funcional de los miembros de la familia EFL1, el vector vacío pAM1 (control negativo) y las construcciones pAM1·ScEFL1 (Control positivo), pAM1·SpEFL1, pAM1·HsEFL1-1, pAM1·HsEFL1-2, pAM1·MmEFL1 y pAM1·AfEF-2 se emplearon para transformar células diploides de la cepa de *S. cerevisiae* AM1 (Tabla 1). Las clonas transformantes se seleccionaron en medio SD-Ura (Tabla Suplementaria 9). Sólo las células de *S. cerevisiae* transformadas con los plásmidos que expresan a las proteínas sin la etiqueta V5 se emplearon para los

ensayos de esporas al azar, el resto se utilizaron para detectar la expresión de cada proteína.

Por otro lado, para probar la complementación funcional de los miembros de SBDS, el vector vacío pAM2 (control negativo) y las construcciones pAM2-ScSDO1 (control positivo), pAM2-SpSDO1, pAM2-MmSBDS, pAM2-HsSBDS, pAM2-AfSBDS, pAM2-HsSBDS D2Sc y pAM2-ScSDO1 D2Hs se emplearon para transformar células diploides de la cepa AM2 (Tabla 1). las células transformantes se seleccionaron en medio SD-Met (Tabla Suplementaria 9).

Finalmente, con el objetivo de evaluar un fenómeno de co-evolución entre los miembros de la familia EFL1 y SBDS, los vectores vacíos pAM1/pAM2 (controles negativos) y los pares pAM1-ScEFL1/pAM2-ScSDO1 (Control positivo), pAM1-SpEFL1/pAM2-SpSDO1, pAM1-MmEFL1/pAM2-MmSDO1, pAM1-HsEFL1-1/pAM2-HsSBDS, pAM1-HsEFL1-2/pAM2-HsSBDS, pAM1-AfEF-2/pAM2-AfSBDS, pAM1-HsEFL1-1/pAM2-HsSBDS·D2Sc, pAM1-HsEFL1-1/pAM2-ScSdo1·D2Hs, pAM1-ScEFL1/pAM2-HsSBDS·D2Sc y pAM1-ScEFL1/pAM2·ScSdo1·D2Hs, se usaron para cotransformar células diploides la cepa de *S. cerevisiae* AM2 (Tabla 1). Las células cotransformantes fueron seleccionadas en medio SD-Ura-Met (Tabla Suplementaria 9).

Extracción de proteína total de células de *Saccharomyces cerevisiae*

Se inocularon 3 mL de medio de selección adecuado con cada una de las clonas de *S. cerevisiae* transformantes (en el caso de EFL1, aquellas que se transformaron con los plásmidos que expresan a las proteínas fusionadas con la etiqueta V5) y se mantuvieron a 30 ° C con agitación constante (250 rpm) durante 18 h. Posteriormente, se inoculó 50 mL de medio de selección fresco con la cantidad de preinóculo suficiente para alcanzar una DO_{600nm} de 0.2. Los cultivos se incubaron a 30 ° C con agitación constante (250 rpm) durante 4-5 h hasta llegar a una DO_{600nm} de 0.7-0.8. Una vez alcanzada la densidad óptica deseada, se obtuvo el paquete celular de cada cultivo por centrifugación (5 min a 14,500 rpm) y se resuspendió en 200 μ L de solución de lisis (Tris·HCl 50 mM pH 7, $MgCl_2$ 5 mM, NaCl 300 mM, glicerol 10%) adicionado con inhibidor de proteasas (ProBlock™, Gold Biotechnology). A continuación, se agregaron aproximadamente 100 μ L de

perlas de vidrio (0.45 μm de diámetro) y se provocó la lisis celular en un equipo FastPrep-24 (7 ciclos de 20 s). En seguida, las muestras fueron centrifugadas a 14,500 rpm durante 10 min y el sobrenadante se recuperó en un tubo nuevo. La cantidad de proteína total se cuantificó por el método de Bradford, y finalmente las muestras se congelaron en nitrógeno líquido y se almacenaron a - 80 °C.

Detección de la expresión proteica por Inmunoblot

Para cada una de las muestras se tomaron 20 μg de proteína total y se resolvieron en un gel de poliacrilamida (SDS-PAGE) al 12%. En seguida, las proteínas se transfirieron a una membrana de PVDF (Millipore) durante 27 min a 20 V en una cámara de transferencia semiseca (Bio-Rad). Para verificar que la transferencia de las proteínas fue exitosa, la membrana se teñó con rojo de Ponceau. Posteriormente, la membrana se destiñó y bloqueó en buffer TBST-10% leche (NaCl 140 mM, Trizma base 100 mM, Tween-20 0.05 %) durante 1 h a temperatura ambiente. La detección de las proteínas EFL1 y SBDS se realizó con los anticuerpos α -V5 (1:5000 TBST+ 5 % leche, Invitrogen) y α -C-Myc (1:5000 TBST + 5 % leche, Invitrogen), respectivamente. Para detectar el control de carga, se usó el anticuerpo α -GAPDH (1:15000 TBST + leche 5 %, Invitrogen). Todos los anticuerpos utilizados en este trabajo estaban acoplados a una peroxidasa de rábano, por lo cual, empleamos como revelador el reactivo SuperSignal West Pico Chemiluminiscent Substrate (Thermo Scientific).

Inducción de esporas de las células de *S. cerevisiae* transformantes

Las clonas de *S. cerevisiae* transformadas con las construcciones que expresan a cada uno de los miembros de la familia EFL1 se estriaron en medio de esporulación sólido (Tabla Suplementaria 9) y se incubaron durante 8 días a 23 °C. Las células de levadura transformadas con los plásmidos que codifican los diferentes miembros de la familia SBDS, así como las células cotransformadas con los plásmidos que codifican ambas familias de proteínas, presentaron retardo en crecimiento y esporulación. Por lo tanto, las transformantes se estriaron en

medio de esporulación sólido y se incubaron a 23 °C durante 24 días. El proceso de esporulación de todas las células se monitoreó bajo el microscopio óptico.

Ensayos de esporas al azar

Las esporas de las células de *S. cerevisiae* transformantes se resuspendieron en agua estéril y posteriormente se plaquearon sobre cajas de medio de crecimiento sintético SD/MSG-His/Arg/Lys suplementado con canavanina 50 µg/mL, tialisina 50 µg/mL sin ningún antibiótico para seleccionar a la progenie meiótica silvestre y observar la viabilidad de las esporas, 2) SD/MSG-His/Arg/Lys+canavanina/tialisina con 100 µg/mL de cloNAT para seleccionar a la progenie meiótica *efl1Δ::natMX6* y ensayar la conservación funcional de las proteínas EFL1, 3) SD/MSG-His/Arg/Lys+canavanina/tialisina con 200 µg/mL de G418 para seleccionar a la progenie meiótica *sdo1Δ::kanMX4* y ensayar la complementación funcional de las proteínas SBDS y, 4) SD/MSG-His/Arg/Lys+canavanina/tialisina/cloNAT/G418/ para seleccionar a la progenie meiótica *efl1Δ::natMX6 sdo1Δ::kanMX4* y ensayar la complementación funcional de los pares EFL1/SBDS de cada especie. Los medios adicionados con cloNAT fueron incubados durante 3 días a 30 ° C, mientras que los medios que contenían G418, se mantuvieron a la misma temperatura durante 5 días.

6 Resultados

6.1 Estudio bioinformático de las proteínas EFL1

Perfil filogenético de la familia de proteínas EFL1

Para iniciar este estudio se buscaron secuencias que fueran posibles ortólogos de ScEfl1. La búsqueda en la base de datos UniprotKB usando como entrada el pHMM generado a partir de miembros conocidos de la familia EFL1 arrojó 94,835 resultados (E -value < 0.01). Este conjunto de secuencias contemplaba miembros de EFL1, EF2, EF-G y otras proteínas, principalmente GTPasas. El análisis de los resultados con el programa OrthoMCL identificó 1,269 secuencias como miembros de EFL1. Los fragmentos más pequeños y secuencias repetidas fueron eliminadas resultando en un total de 750 proteínas, todas exclusivas de organismos diferentes del dominio Eukarya. La Tabla 21 muestra la distribución filogenética de la familia EFL1 de acuerdo a la información brindada en UniprotKB y a la clasificación propuesta por [11,15,106,116]. Los miembros de EFL1 se presentan en la mayoría de los subgrupos de los phyla que componen el dominio Eukarya, con un enriquecimiento particular en Fungi, Metazoa, Streptophyta y Alveolata.

La longitud de las proteínas recopiladas varía entre especies, el valor promedio es de alrededor de 1040 aminoácidos. Las proteínas más grandes están presentes en los alveolados *Gregarina niphandrodes* y *Pseudocohnilembus persalinus*, y el hongo *Mycena chlorophos*, cuya longitud es de 2,715, 2,752 y 2,393 aminoácidos, respectivamente. Otras proteínas con más de 1,200 residuos fueron encontradas en la mayoría de los protistas y en algunos hongos, plantas y metazoarios. Alineamientos múltiples de secuencias respecto a ScEfl1 y HsEFL1, revelaron que estos miembros presentan diferencias estructurales importantes, entre las que destacan inserciones en los subdominios del dominio catalítico y el dominio IV de ScEfl1, así como expansiones en la región C o N terminal. Por ejemplo, los miembros de *Vitis vinifera* y *Trichinella nelsoni* presentan residuos extra en ambos extremos, mientras que *Drosophila busckii* y *Phytophthora nicotianae*, varían en el

C-terminal. De especial mención es la proteína EFL1 de *P. nicotianae* que presenta una expansión de aproximadamente 800 aminoácidos en el C-terminal.

Tabla 21. Distribución filogenética de la familia de proteínas EFL1.

Reino/Phylum	Subgrupo	No. de especies
Opisthokonta	Metazoa	244
	Fungi	337
	Ichthyosporea	2
	Choanoflagellida	2
	Nucleariidae	1
Amoebozoa	Mycetozoa	5
	Archamoebae	3
	Apusozoa	1
Archaeplastida	Chlorophyta	10
	Streptophyta	61
	Rhodophyta	3
Chromalveolata	Stramenopila	19
	Haptophyta	2
	Alveolata	37
	Cryptophyta	1
Excavata	Euglenozoa	17
	Heterolobosea	1
	Parabasalia	1
Rhizaria	Cercozoa	2
	Foraminifera	1
Total		750

Adicionalmente, identificamos dos miembros de EFL1 en algunos organismos como *Arabidopsis thaliana*, *Macaca mulatta*, *Drosophila melanogaster*, *Homo sapiens*, *Danio rerio*, *Felis catus*, *Gasterosteus aculeatus*, *Glarea lozoyensis*, *Apergillus niger*, *Botryotinia fuckeliana*, *Moniliophthora roreri*, *Rhodospodium toruloides*, *Rhipicephalus pulchellus*, *Ajellomyces dermatitidis*, *Callithrix jacchus*, *Anas platyrhynchos*, *Cricetulus griseus*, *Plasmodium falciparum*, *Takifugu rubripes*, *Drosophila simulans*, *Gasterosteus aculeatus*, *Sporisorium scitamineum*,

Echinococcus granulosus, *Rhipicephalus pulchellus*, *Schistocephalus solidus*, *Desmodus rotundus*, *Loxodonta africana*, *Meleagris gallopavo*, *Poeciliopsis prolifica*, *Phytophthora infestans*, *Ciona savignyi*, *Fundulus heteroclitus*, *Oryza barthii*, *Pongo abelii*, *Pan troglodytes* y *Gorilla gorilla*. El análisis entre ambas secuencias de aminoácidos para cada organismo reveló que comparten un alto porcentaje de identidad con diferencias principalmente en componentes importantes de los subdominios catalíticos. Esto ya había sido observado en la isoforma 2 de la proteína humana (HsEFL1-2) que respecto a la isoforma 1 carece de 51 residuos correspondientes a la parte final del subdominio G1 y el G2 (Figura 11). Otros organismos en los que encontramos dos posibles candidatos fueron *Sorghum bicolor*, *Leersia perrieri*, *Oryza glumipatula*, *Oryza nivara*, *Oryza punctata*, *Oryza rufipogon*, *Oryza sativa*, *Angomonas deanei* y *Brachypodium distachyon*. Sin embargo, el análisis de cada par de proteínas mediante su alineamiento de secuencia primaria, mostró un mayor grado de divergencia entre ellas, sugiriendo que probablemente no son isoformas (Figura 12). Un análisis filogenético (dato no mostrado) en el que incluimos miembros conocidos de EFL1 y eEF-2, demostró que los dos miembros de cada organismo son evolutivamente más cercanos a la familia EFL1 que a eEF-2, indicando que probablemente el gen que codifica a EFL1 en estos organismos se duplicó.

[A]

```

                                     G1
HsEFL1-1_Q7Z2Z2  MVLNSLDKMIQLQKNTANIRNICVLAHVDHGKTTLADCLISSNGI ISSRLAGKLRYMDSREDE
HsEFL1-2_Q7Z2Z2-2 MVLNSLDKMIQLQKNTANIRNICVLAHVDHG-----
*****

                                     G2                                     G3
HsEFL1-1_Q7Z2Z2  QIRGITMKSSAISLHYATGNEEYLINLIDSPGHVDFSSEVSTAVRICDGCIIVVDAVEGVCPQ
HsEFL1-2_Q7Z2Z2-2 -----NEEYLINLIDSPGHVDFSSEVSTAVRICDGCIIVVDAVEGVCPQ
*****

                                     G4
HsEFL1-1_Q7Z2Z2  TQAVLRQAWLENIRPVLVINKIDRLIVELKFTPEAYSHLKNILEQINALTGTFLTFSKVLEER
HsEFL1-2_Q7Z2Z2-2 TQAVLRQAWLENIRPVLVINKIDRLIVELKFTPEAYSHLKNILEQINALTGTFLTFSKVLEER
*****

                                     G5
HsEFL1-1_Q7Z2Z2  AERETESQVNPNSEQGEQVYDWSTGLEDTDDSHLYFSPEQGNVFTSAIDGWGFGIEHFARIY
HsEFL1-2_Q7Z2Z2-2 AERETESQVNPNSEQGEQVYDWSTGLEDTDDSHLYFSPEQGNVFTSAIDGWGFGIEHFARIY
*****
```

[B]

```

                                     G1
DmEFL1_Q9VV61  MPVVEGSDLVLLQRRRQQRNICILAHVDHGKTTLADSLVASNGIISQRMAGKLRYLDRSDE
DMEFL1_Q95TH3  -----

                                     G2                                     G3
DmEFL1_Q9VV61  QERGITMKSSSISLYYQEAEMAGNPDYLINLIDSPGHVDFSSEVSTAVRLCDGAIIVVDVVE
DMEFL1_Q95TH3  -----MAGNPDYLINLIDSPGHVDFSSEVSTAVRLCDGAIIVVDVVE
*****

                                     G4
DmEFL1_Q9VV61  GVGPPQTRACLRQIYEEQLKPVVLNKLDRILEKQMDPLDAYFHLCQVLEQVNAVLSIFASD
DMEFL1_Q95TH3  GVGPPQTRACLRQIYEEQLKPVVLNKLDRILEKQMDPLDAYFHLCQVLEQVNAVLSIFASD
*****

                                     G5
DmEFL1_Q9VV61  ILAKEDITKKDNYESALEEVDDSELYFSPSSGNVIFCSAYDGWAFSVRDFAAAMYAKRLEMSRK
DMEFL1_Q95TH3  ILAKEDITKKDNYESALEEVDDSELYFSPSSGNVIFCSAYDGWAFSVRDFAAAMYAKRLEMSRK
*****
```

Figura 11. Alineamiento de secuencia primaria del dominio catalítico correspondiente a las dos posibles proteínas EFL1 de [A] *H. sapiens* y [B] *D. melanogaster*.

Residuos conservados evolutivamente en la familia de proteínas EFL1

El alineamiento de secuencias de los miembros de EFL1 permitió identificar 51 residuos invariables. Para observar su distribución, cada uno de los residuos identificados fue marcado en rojo sobre la estructura primaria de la proteína ScEfl1 (Figura 13). A pesar de que estos aminoácidos se presentan en todos los dominios que componen a esta proteína, se concentran particularmente en los subdominios G1-G4 del dominio catalítico, el dominio III y el dominio V (principalmente en la región C-terminal).

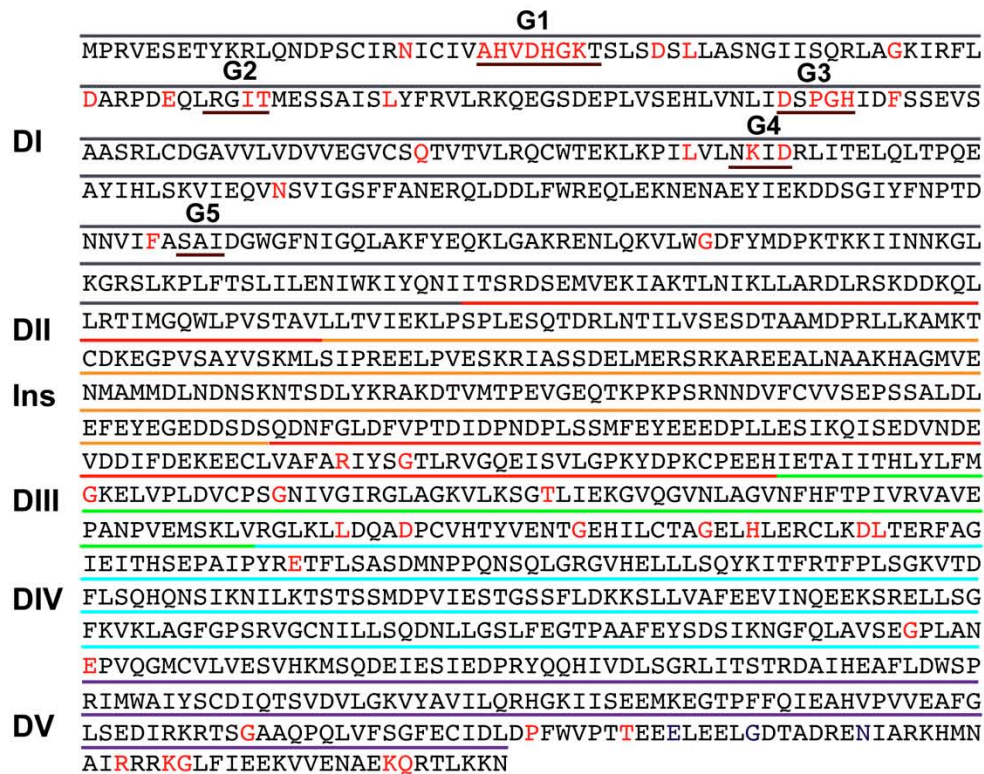


Figura 13. Análisis de los residuos invariables en la familia de proteínas EFL1. Los residuos invariables se identificaron mediante el alineamiento múltiple de secuencias de miembros de EFL1 y se muestran en rojo. Los cinco dominios que conforman a esta proteína, así como los subdominios G que componen el dominio catalítico, se encuentran delimitados con líneas de colores.

Adicionalmente, decidimos identificar las posiciones que presentaban sustituciones conservativas, es decir, sitios que de acuerdo al alineamiento, eran sustituidos por residuos con propiedades fisicoquímicas similares. Por lo tanto, aquellos residuos con un valor de conservación mayor o igual a 8, de acuerdo con

el análisis del alineamiento realizado con el programa T-COFFEE y Jalview 2.4, fueron representados en rojo sobre la secuencia de aminoácidos de ScEfl1 junto con los residuos invariables que son resaltados por encima de la tipografía normal (Figura 14). Esta clase de residuos es muy abundante en los dominios G, III y V, y en menor proporción en los dominios II y IV. Es posible observar que la inserción que caracteriza a las proteínas EFL1 de las EF-2 no presenta residuos invariables o cambios en residuos conservados, por lo que es probable que este fragmento sea característico de especie o de un linaje en específico. Asimismo, es interesante notar que la mayoría de las posiciones con sustituciones neutras se concentran en los alrededores de los residuos invariables; dada la conservación evolutiva de los residuos en estos sitios es probable que representen un papel importante en la función y plegamiento de esta familia de proteínas.

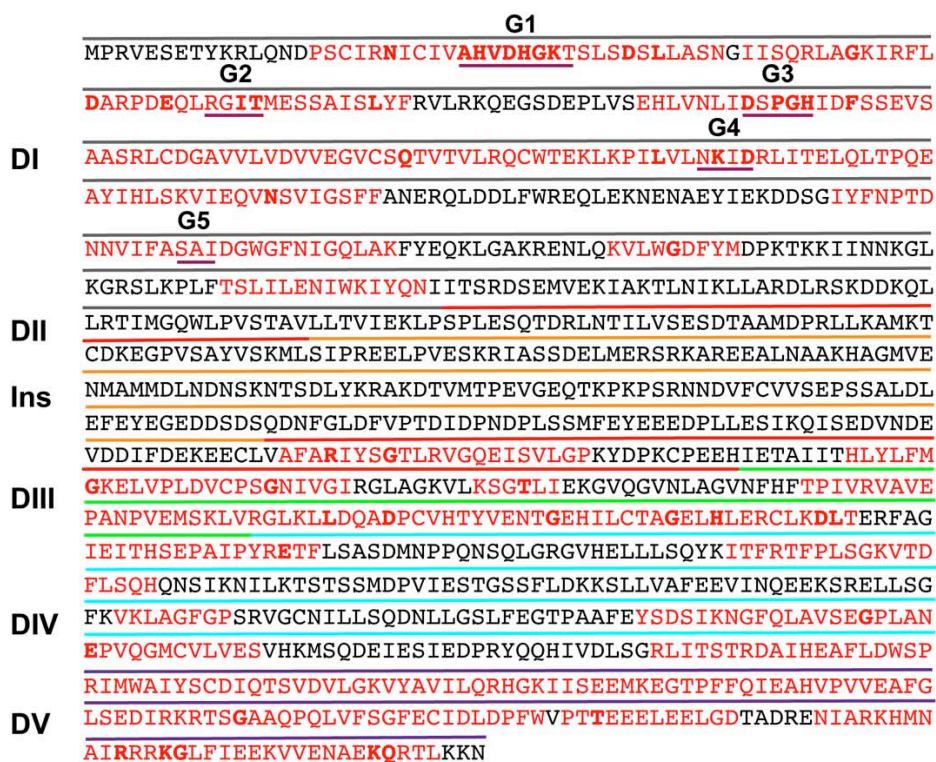


Figura 14. Análisis de los residuos con un alto índice de conservación en la familia de proteínas EFL1. Los residuos con un porcentaje de identidad alto fueron identificados mediante el alineamiento múltiple de secuencias de miembros de la familia de proteínas EFL1 y se muestran en rojo. Los cinco dominios por los que la proteína está conformada, se encuentran delimitados con líneas de colores.

La distribución global de los residuos invariables y las posiciones con sustituciones conservativas en esta familia de proteínas sugiere que su plegamiento tridimensional está conservado evolutivamente. La predicción de estructura secundaria a partir de las secuencias de aminoácidos en el alineamiento de diversos miembros de EFL1 empleando el módulo de PSIPRED contenido en el programa PRALINE se muestra en la Figura 15, en donde las estructuras de α -hélice y hojas β conservadas por los miembros de EFL1 fueron marcadas en rojo y en azul, respectivamente, sobre la secuencia de aminoácidos de ScEfl1. Si se comparan las Figura 14 y Figura 15 es posible notar que gran parte de las regiones de estructura secundaria predichas por PSIPRED coinciden con posiciones que presentan residuos invariables o sustituciones conservativas. Interesantemente, puede observarse que la mayoría de estas regiones de estructura secundaria corresponden con las presentadas en la estructura de ScEfl1 modelada por homología (Figura 16). Esto no sólo indica que la probabilidad de que el plegamiento de los miembros de EFL1 se encuentre conservado evolutivamente es alta, sino que además, dado que el modelo tridimensional fue obtenido usando la estructura cristalográfica de ScEf-2, es probable que este plegamiento sea compartido por ambas familias de proteínas, soportando la hipótesis de que ambas proteínas podrían compartir el mismo origen evolutivo.

MPRVESE**TYKRLQ**ND**PSCIRNICIVAHVDHGKTSLS**SDSLLASNGIIS**QRL**LAGKIRFL
 DAR**PDEQLRGI**TM**ESSAISLYFRVLRKQEGSDEPLVSEHLVNLID**SPGHIDFS**SSEVS**
DI **AASRLCDGAVVLVDV**VEGVCS**QTVT**VLR**QCWTEK**LKP**ILVLN**KID**RLITELQ**LT**PQE**
AYIHLSKVIEQVNSVIGSFFANERQLDDLFWRE**QLEKNENAEYIEK**DDSGIY**FNPTD**
NNVIFASAIDGWGF**NI**Q**LAKFYEQKL**GAKREN**LQKVLW**GDFYMDPKTKKIINN**KGL**
KGRSLKPLFTSLILENIWKIYQNIITSRD**SEMVEKIAKTL**NIKLLARDLRSKDD**KQL**
DII **LRTIMGQWLPVSTAVLLTVIEKLPSPLESQTDRLNTILV**SESDTAAMDPRLLKAMKT
 CDKEGPVSAYVSKMLSIPREELPVESKRIASSDEL**ERSK**KAREEALNAAKHAGMVE
Ins NMAMMDLNDNSKNTSDLYKRAKDTVMTPEVGEQTKPKPSRNNDFVFCVVSEPSSALDL
 EFEYEGEDDSDSQDNFGLDFVPTDIDPNDPLSSMFYEEEDPLLESIKQISEDVND
 VDDIFDEKEECL**VAFARIYSGTLRVGQEI**SVLGPKYDPKCPEEH**IETAIITHLYLFM**
DIII **GKELVPLDVCPSGNIVGIRGLAGKVLKSGTLIEKGVQGVNLAGVNFHFTPIVRVAVE**
PANPVEMSKLVRGLKLLDQADPCVHTYVENTGEHILCTAGELHLERCLKDLTERFAG
IEITHSEPAIPYRETFLSASDMNPPQNSQLGRGVHELLLSQYKITFRTFPLSGKVTD
DIV **FLSQHQNSIKNILKTSTSSMDPVIESTGSSFLDKKSLLVAFEEVINQEEKSRELLSG**
FKVKLAGFGPSRVGCNILLSQDNLLGSLFEGTPAAFEYSDSIKNG**FQLAV**SEGPLAN
 EPVQGM**CVLVESVHKMSQDEIESIEDPRYQQHIVDL**SGRLI**TSTRDAIHEAFLD**WSP
RIMWAIYSCDIQTSVDVLGKVYAVILQRHGKIISEEMKEGTP**FFQIEAHVPVVEAFG**
DV **LSEDIRKR**TS**GAAQPQLVFSGFEC**IDLDPFWVPT**EEEELEELGDTADRENIARKHMN**
AIRRRKGLFIEEKVVENAE**KQRTLKKN**

Figura 15. Análisis de estructura secundaria en la familia de proteínas EFL1. En rojo y en azul se muestran las regiones predichas como α hélices y hojas β , respectivamente, de acuerdo al alineamiento múltiple de secuencias de los miembros de EFL1 obtenido mediante el programa PSIPRED. Los cinco dominios por los que la proteína está conformada, se encuentran delimitados con líneas de colores.

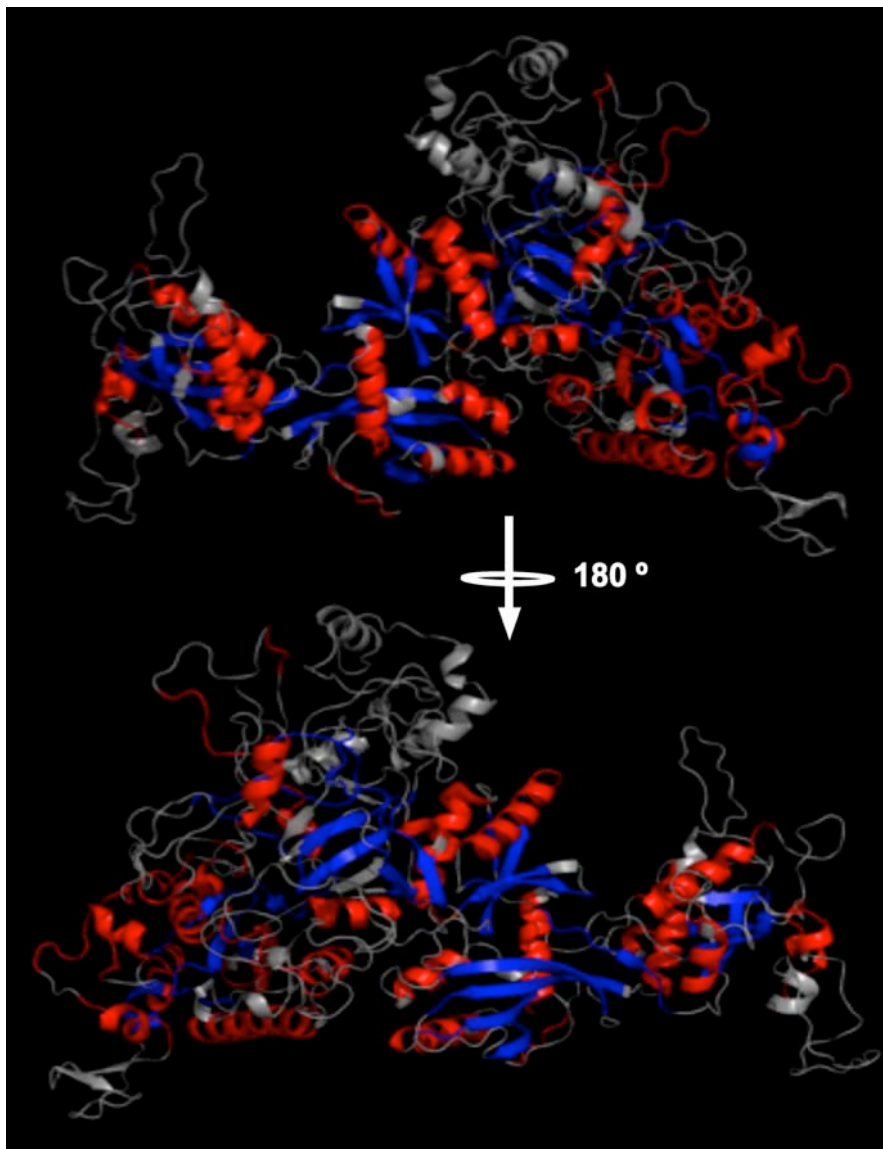


Figura 16. Predicción del contenido de estructura secundaria de la familia de proteínas EFL1. Modelo del plegamiento tridimensional de ScEfl1. En rojo y en azul se muestran las regiones de α -hélices y hojas- β , respectivamente, que de acuerdo al alineamiento de múltiples secuencias de los miembros de EFL1, y el predictor de estructura secundaria PSIPRED se encuentran conservadas en la familia EFL1.

Análisis filogenético de las GTPasas EFL1, EF-2 y EFG

Durante la búsqueda de los miembros de la familia EFL1 observamos consistentemente la presencia de secuencias correspondientes a miembros de EF-2 y EF-G. Algunos autores ya habían sugerido, dado el porcentaje de similitud que presentan entre sí estas proteínas, una posible relación evolutiva, sin embargo, hasta la fecha no había reportes sobre este tipo de investigaciones. Por

lo tanto, para estudiar la relación entre estas proteínas, decidimos construir un árbol filogenético.

El árbol se organizó en cinco grupos monofiléticos, uno de ellos correspondiente al grupo externo (EF-4), y cuatro más que corresponden a los miembros de eEF-2 de Eukarya y aEF-2 de Archaea, a EF-G y EFL1, que conforman nuestro grupo de estudio. Interesantemente, este análisis reveló que todas estas proteínas presentan un origen evolutivo en común. Además, el análisis filogenético muestra que las proteínas EFL1 y eEF-2 son taxa hermanos debido a que comparten el ancestro en común más cercano. La presencia de miembros de ambas familias de proteínas en un mismo organismo indica que las proteínas EFL1 se originaron por duplicación génica después de la separación de los dominios Eukarya y Archaea (Figura 17). Adicionalmente, un alineamiento múltiple entre secuencias de eEF2 de organismos evolutivamente lejanos mostró que existen pocas diferencias en la secuencia primaria de los miembros de esta familia, mientras que el mismo análisis con los miembros de EFL1, reveló que estas últimas proteínas presentan mayores diferencias entre sí. En conjunto, estos resultados sugieren que la familia eEF-2 ha experimentado pocos cambios en su estructura primaria, y muy probablemente en su estructura tridimensional a lo largo de su evolución; manteniendo su papel en el proceso de traducción. Sin embargo, al parecer las proteínas EFL1 han adquirido mayor cantidad de cambios que favorecieron su especialización en la vía de maduración de la subunidad ribosomal 60S.

Finalmente, diversos estudios han revelado que las proteínas SBDS y eIF6/Tif6 presentan ortólogos en Archaea, además de que eIF6 también funciona como un factor de anti-asociación en estos organismos [11,83,117,118]. Lo anterior sugiere que este paso en la vía de maduración ribosomal observado en eucariontes podría conservarse en arqueas. Por otro lado, se ha demostrado que la función de las proteínas EFL1 en mamíferos y en levadura, es esencial para la liberación del factor de anti-asociación de la subunidad ribosomal mayor. Debido a que nuestros resultados han demostrado que la familia EFL1 es exclusiva del dominio Eukarya, no resulta trivial inferir qué proteína del dominio Archaea podría presentar una función similar. Sin embargo, la relación evolutiva entre las proteínas EFL1, eEF-2 y aEF-2 revelada en el árbol filogenético, sugiere que estas últimas, además de

participar en el proceso de translocación durante la síntesis de proteínas, podrían representar el papel de la familia EFL1 durante la biogénesis ribosomal en Arqueas.

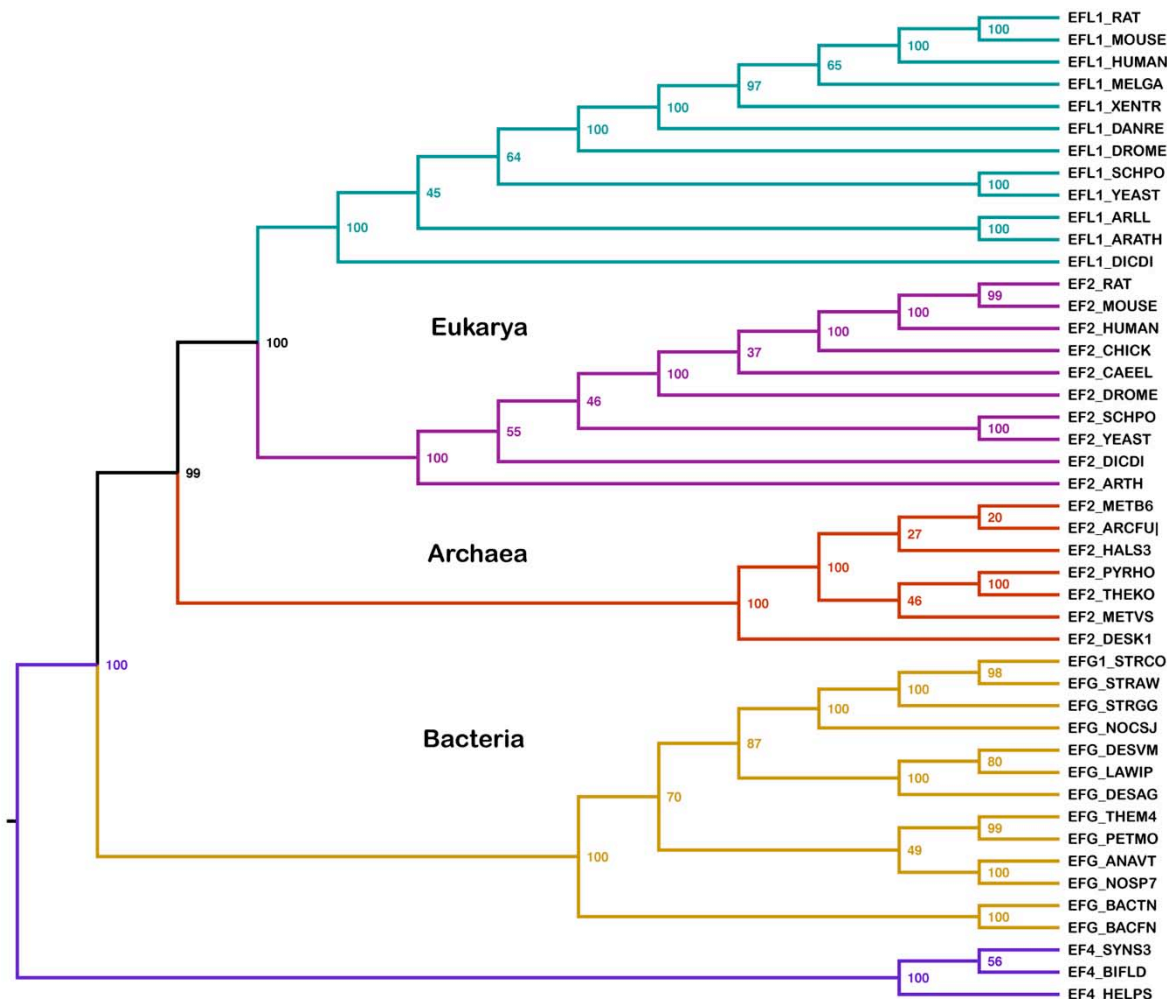


Figura 17. Árbol filogenético de las proteínas EFL1 (verde), eEF-2 (rosa), aEF-2 (naranja), EF-G (amarillo) y EF-4 (grupo externo, morado). Los valores de Bootstrap se muestran para cada una de las ramas.

Diferencias estructurales entre EFL1 y eEF-2

Estudios previos en los que se compara la secuencia de ScEfl1 y ScEF-2, mostraron la presencia de una inserción de 162 residuos en el dominio II. Asano *et al.* [87] estudiaron que esta región se encuentra desordenada y es esencial para la interacción entre las proteínas ScEfl1 y ScSdo1 *in vitro*. Su importancia en la

función que desempeña ScEfl1 durante el proceso de biogénesis ribosomal pone de manifiesto la posibilidad de que otras diferencias estructurales con las proteínas eEF-2 también contribuyan con su especialización.

Para identificar los cambios estructurales que la familia EFL1 ha sufrido durante su proceso evolutivo respecto a eEF-2, decidimos detectar la presencia de inserciones en las proteínas EFL1 de *S. cerevisiae*, *S. pombe*, *H. sapiens*, *A. thaliana* y *P. falciparum* mediante el alineamiento de sus secuencias de aminoácidos con las secuencias de su proteína eEF-2 correspondiente (Figura 18). Este análisis reveló que el dominio III es el único sitio en todas las proteínas que no presenta inserciones. Adicionalmente, mostró que el número y longitud de las inserciones permanece prácticamente constante en los dominios G y V, mientras que varía en los dominios II y IV, por lo que nos concentramos particularmente en estos sitios.

En cuanto a la cantidad de inserciones en el dominio II, las proteínas AtEFL1 y PfEFL1 presentaron el mayor número con 2 y 3, respectivamente; sin embargo, es importante resaltar que las de mayor longitud, se encontraron en ScEfl1 (162 aa), HsEFL1 (79 aa) y PfEFL1 (74 aa). Independientemente de su tamaño, dada la cercanía del dominio II con el sitio catalítico, permite sospechar que estas regiones podrían conservar la misma función que la inserción identificada en ScEfl1. El dominio IV fue otro de los dos sitios con mayor cantidad de cambios, particularmente, el dominio IV de la proteína PfEFL1, cuyo número y longitud de las inserciones es mayor que en el resto. En general, es posible suponer que los cambios estructurales identificados en estas proteínas respecto a eEF-2, particularmente los observados en los dominios II y IV, además de promover la especialización de la función de las proteínas EFL1 podrían también otorgarles propiedades específicas de especie.

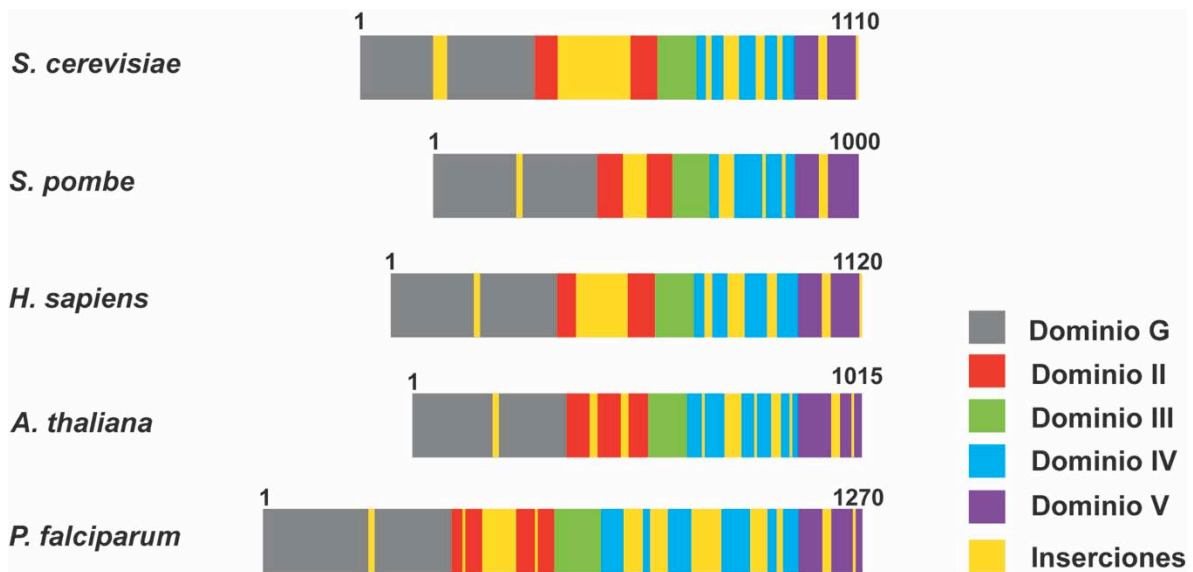


Figura 18. Diferencias estructurales en miembros de la familia EFL1 respecto a sus eEF-2 correspondientes. Los dominios de cada proteína fueron inferidos mediante su alineamiento con ScEfl1 y posteriormente representados en bloques de colores. Las inserciones observadas en los alineamientos fueron localizadas sobre cada proteína y marcadas en amarillo.

6.2 Evaluación funcional de las proteínas EFL1 y SBDS

Obtención de las cepas de *Saccharomyces cerevisiae* *efl1Δ/EFL1* y *efl1Δ/EFL1 sdo1Δ/SDO1*

Para evaluar la conservación funcional de la familia EFL1, así como un posible fenómeno de co-evolución entre las familias EFL1/SBDS, ensayamos la capacidad de complementación funcional de las proteínas candidatas en células de levadura *efl1Δ*, *sdo1Δ* y *efl1Δ/sdo1Δ*, respectivamente. Debido a que la función de los genes *SDO1* y *EFL1* es necesaria para el crecimiento óptimo de *S. cerevisiae*, resulta complicado obtener células haploides con este genotipo. Por lo tanto, aprovechamos que este organismo presenta un ciclo de vida diploide y generamos las cepas diploides heterocigotas AM1 y AM2 (Tabla 1). A partir de lo anterior, fue posible transformar a las células de cada cepa con los plásmidos de interés, detectar por inmunoblot la expresión de las proteínas EFL1 y SBDS candidatas, generar esporas y evaluar la complementación genética de los miembros de la familia EFL1, SBDS y de los pares ortogonales EFL1/SBDS correspondientes,

mediante la selección de progenie meiótica viable *efl1Δ*, *sdo1Δ* y *efl1Δ/sdo1Δ*, respectivamente.

La Figura 19A muestra el amplicón del módulo de resistencia a cloNAT (~1,500 pb) obtenido al adicionar en sus extremos 50 pb idénticos a las secuencias regulatorias río arriba y abajo del gen *EFL1* de *S. cerevisiae*. Este producto fue empleado para transformar células de la cepa *S. cerevisiae* Y5538 (Tabla 1) con el fin de remplazar con este marcador uno de los alelos *EFL1* (cepa AM1, Tabla 1). A partir de la selección de transformantes sobre medio YPD + cloNAT (Tabla Suplementaria 9) logramos detectar cinco clonas candidatas. La sustitución de la secuencia codificante de Efl1 por el marcador de selección, así como la presencia de un alelo *EFL1* integro, se evaluaron mediante la amplificación por PCR de un fragmento que comprendía el módulo de resistencia más 500 pb río arriba del codón de inicio del marco abierto de lectura (~1,700 pb) y la secuencia codificante de Efl1 (~3,300 pb), respectivamente. En la Figura 19B se puede observar la presencia de una banda de 1,700 pb para las cinco clonas analizadas, indicando que el módulo de resistencia fue integrado en el locus esperado. Por otro lado, en la Figura 19C se muestra el análisis de la clona 1 (cepa de trabajo). La banda del carril 1, como se mencionó anteriormente, indica que el módulo *natMX4* fue integrado en el locus de interés, mientras que la banda de 3,300 pb en el carril 2, revela que esta cepa aún presenta un alelo *EFL1*.

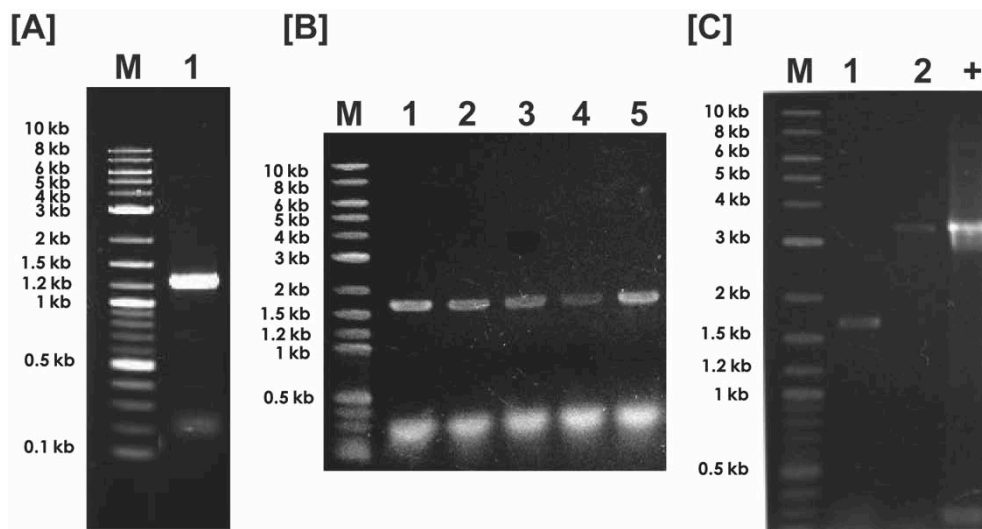


Figura 19. Construcción y caracterización de la cepa *S. cerevisiae* AM1. Geles de agarosa al 1% teñidos con bromuro de etidio. [A] M: marcador de peso molecular, 1: amplicón del módulo *natMX4*. [B] M: marcador de peso molecular, 1-5: amplicones del módulo *natMX4* más 500 pb río arriba de la secuencia codificante de Efl1, obtenidos usando como molde el gDNA de las clonas recombinantes 1-5. [C] M: marcador de peso molecular, 1: amplicón correspondiente al módulo *natMX4* más 500 pb del locus *EFL1* 2: amplicón de la secuencia codificante de Efl1, obtenidos empleando como molde el gDNA de la clona recombinante 1, + : control positivo, amplicón de *EFL1* obtenido empleando como molde la construcción pAM1-*ScEFL1*.

Para generar la cepa *S. cerevisiae* AM2, el módulo *kanMX6* flanqueado por 20 pb idénticos a las secuencias río arriba y abajo de la secuencia codificante de la proteína Sdo1 (~1,500 pb) fue obtenido por PCR (Figura 20A) y posteriormente utilizado para reemplazar el marco abierto de lectura de *SDO1* a través de la transformación de células de la cepa AM1. Después del tratamiento, obtuvimos seis clonas candidatas sobre medio YEPD + cloNAT + G418. De cada una se obtuvo gDNA y, usando un oligonucleótido que hibrida 100 pb río arriba del codón de inicio de *SDO1* y otro complementario a la secuencia final del módulo de resistencia se comprobó que el marcador de selección había sido integrado al genoma de la levadura al detectar como producto final de PCR un fragmento en ~ 1,600 pb. Como se muestra en la Figura 20B, todas las clonas analizadas integraron el módulo de resistencia en el locus de interés. Posteriormente, se demostró que las clonas recombinantes 1 y 2 mantenían un alelo de *SDO1* íntegro, a partir de la amplificación de la secuencia codificante de *SDO1* (Figura 20C).

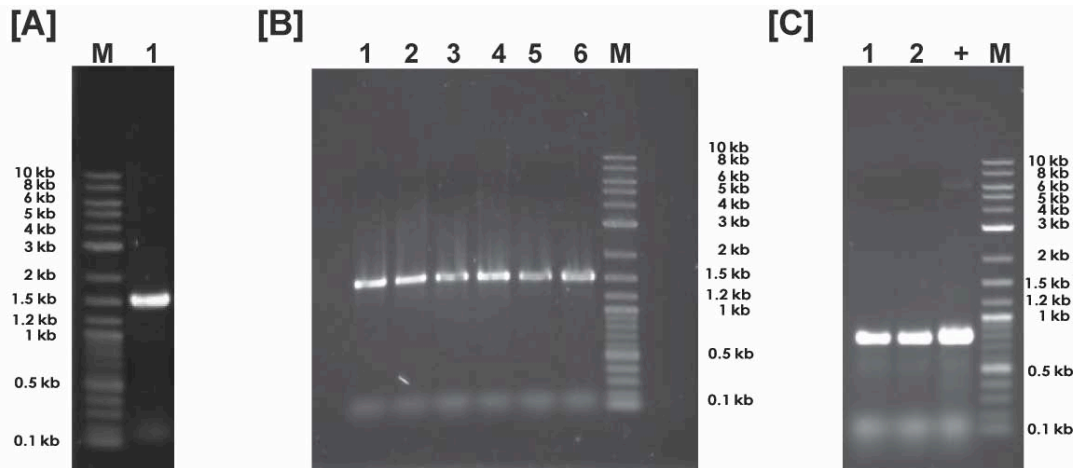


Figura 20. Construcción y caracterización de la cepa *S. cerevisiae* AM2. Geles de agarosa al 1% teñidos con bromuro de etidio. [A] M: marcador de peso molecular, 1: amplicón del módulo *kanMX6* [B] M: marcador de peso molecular. 1-6: amplicones del módulo *kanMX6* más 100 pb rio arriba de la secuencia codificante de *Sdo1*, obtenidos usando como molde el gDNA de las clonas recombinantes 1-6. [C] M: marcador de peso molecular, 1-2: amplicones del gen *SDO1* (750 pb) obtenidos empleando como molde el gDNA de las clonas recombinantes 1-2 y +: control positivo, amplicón de *SDO1* obtenido empleando como molde la construcción pAM2-ScSDO1.

Debido a que el proceso de esporulación no es completamente eficiente, la mayoría de las células empleadas durante los ensayos de complementación para las proteínas EFL1, SBDS y los pares ortogonales EFL1/SBDS, son diploides heterocigotas *efl1Δ::natMX6/EFL1*, *sdo1Δ::kanMX4/SDO1* y *efl1Δ::natMX6/EFL1 sdo1Δ::kanMX4/SDO1*, respectivamente. Estas células son capaces de crecer en medio suplementado con los antibióticos correspondientes, resultando en la presencia de falsos positivos durante la evaluación de la conservación funcional de los ortólogos de interés. Para evitar la presencia de esta clase de células y asegurar que la complementación funcional de los ortólogos seleccionados se evalúa específicamente sobre progenie haploide, las cepas AM1 y AM2 cuentan con marcadores genéticos como *MFA1pr-HIS3*, *MFα1-LEU2*, *can1Δ* y *lyp1Δ* (Tabla 1). Los reporteros *MFA1pr-HIS3* y *MFα1-LEU2*, permiten el crecimiento de las células haploides MATa o MATα en medio sin histidina o leucina, respectivamente, evitando el desarrollo de las células diploides MATa/MATα debido a que la expresión de estos marcadores se encuentra reprimida en ellas. Por otro lado, la

selección de células haploides se hace más astringente al implementar los marcadores recesivos *can1Δ* y *lyp1Δ*. El gen *CAN1* codifica un permeasa de arginina que permite a la canavanina, un análogo tóxico de este aminoácido, entrar y asesinar a la célula. De manera similar, el gen *LYP1* codifica un permeasa de lisina que facilita que la tialisina, su análogo tóxico, entre a la célula y la asesine. Por lo tanto, sólo las células haploides *can1Δ/lyp1Δ*, son capaces de crecer en medio suplementado con este par de análogos de aminoácidos tóxicos, mientras que las células diploides MATa/MATα, que portan una copia íntegra de los genes *CAN1* and *LYP1*, son asesinadas [113].

Vectores de expresión para los ensayos de complementación funcional

Para llevar a cabo los ensayos de complementación genética en *S. cerevisiae*, es fundamental que las proteínas a evaluar sean expresadas en una forma y cantidad similar a las proteínas endógenas. Para ello construimos los vectores pAM1 y pAM2 (Figura 8 y Figura 9) para expresar las proteínas correspondientes a los miembros de las familias EFL1 y SBDS, respectivamente.

Los dos plásmidos tienen ciertas características en común, por ejemplo, ambos pueden ser propagados y seleccionados en *E. coli* debido a que cuentan con un origen de replicación para este organismo (derivado del pMB1) y el gen *bla* que permite a las células transformantes crecer en medio adicionado con ampicilina. Por otro lado, son capaces de replicarse en levadura dado que portan una secuencia de replicación autónoma (ARS por sus siglas en inglés). Adicionalmente, se trata de dos plásmidos sumamente estables, ya que contienen una secuencia centromérica que promueve correctamente su segregación. Además, esta región permite que el número de copias se mantenga entre 1-2 por célula. Finalmente, para la expresión de las secuencias codificantes de interés ambos cuentan con un módulo de expresión que comprende el *pr-ADH1* (-750 pb), un sitio de clonación múltiple, una etiqueta para detección de la expresión de la proteína seguida de un codón de paro y el *ter-ADH1*. Una de las principales ventajas de emplear este promotor, es que no cuenta con la secuencia que inhibe su activación por etanol, por lo que los marcos abiertos de lectura bajo su control

se expresan constitutivamente, además, es un promotor de fuerza media, lo que evita la sobre expresión de las proteínas [119,120].

Por otra parte, estos plásmidos también muestran diferencias significativas entre sí. El vector pAM1 tiene como marcador de selección al gen *URA3* y el pAM2 cuenta con el gen *MET15*. Otra diferencia importante, es que las proteínas expresadas a través del pAM1 son fusionadas a una etiqueta V5, mientras que las proteínas provenientes del pAM2, son fusionadas a una etiqueta c-myc. Ambas marcas son útiles para detectar la presencia de las proteínas empleando los anticuerpos anti-V5 y anti-c-myc, respectivamente.

Para obtener el pAM1, el módulo de expresión fue clonado entre los sitios *SacI* y *KpnI* del vector pRS316. En la Figura 21 se muestra el patrón electroforético de los plásmidos pRS316 (control) y pAM1 obtenido después de su digestión con las enzimas mencionadas anteriormente. En los carriles 1 y 2 es posible observar dos bandas de ~ 4,800 pb correspondientes al pRS316 abierto, mientras que sólo en el carril 2 se detecta una banda de ~ 1,200 pb que representa las secuencias regulatorias insertadas; esto indica que el módulo fue clonado correctamente en el pRS316 y por tanto, que la construcción del vector de expresión pAM1 resultó exitosa.

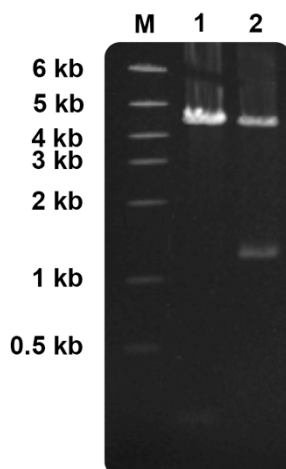


Figura 21. Construcción del vector de expresión pAM1. Gel de agarosa al 1% teñido con bromuro de etidio. M: marcador de peso molecular, 1: Patrón de restricción del pRS316 digerido con *KpnI* y *SacI*, 2: Patrón de restricción del pAM1 restringido con *KpnI* y *SacI*.

La estrategia seguida para la construcción del pAM2 fue más elaborada que la implementada para obtener el pAM1. En la Figura 10 se describe la estrategia experimental efectuada para construir el vector de expresión para las proteínas SBDS. Primero amplificamos el fragmento P1 (~900 pb) que comprendía el promotor *ADH1*, el sitio de clonación múltiple y la secuencia codificante de la etiqueta c-myc, y enseguida se amplificó el fragmento P2 (~500 pb) que estaba conformado por la etiqueta c-myc y el del terminador *ADH1* (Figura 22A); las secuencias c-myc complementarias entre los fragmentos P1 y P2 permitieron ensamblar el módulo completo (~ 1,400 pb) por traslape a partir de un tercer ciclo de PCR (Figura 22B). Una vez que se obtuvo el módulo completo, se clonó entre los sitios *SacI* y *KpnI* del pRS411. Para confirmar que las secuencias regulatorias habían sido insertadas exitosamente en el pRS411, el pDNA obtenido de 9 clonas bacterianas seleccionadas sobre medio adicionado con ampicilina después de su transformación con las reacciones de ligación, fue digerido con *KpnI* y *SacI*. De las nueve clonas, 6 mostraron el patrón esperado: una banda de ~5 kb, que pertenece al pRS411, y una banda de ~1,200 pb, correspondiente al módulo de expresión (Figura 22C). Los resultados anteriores revelaron que la construcción del pAM2 había sido exitosa. Finalmente, para confirmar la presencia de la secuencia c-myc y la integridad del resto de las regiones del módulo, la región correspondiente del plásmido obtenido de la clona 6 se secuenció. El resultado de este análisis mostró la presencia de la etiqueta c-myc y la ausencia de mutaciones en el módulo.

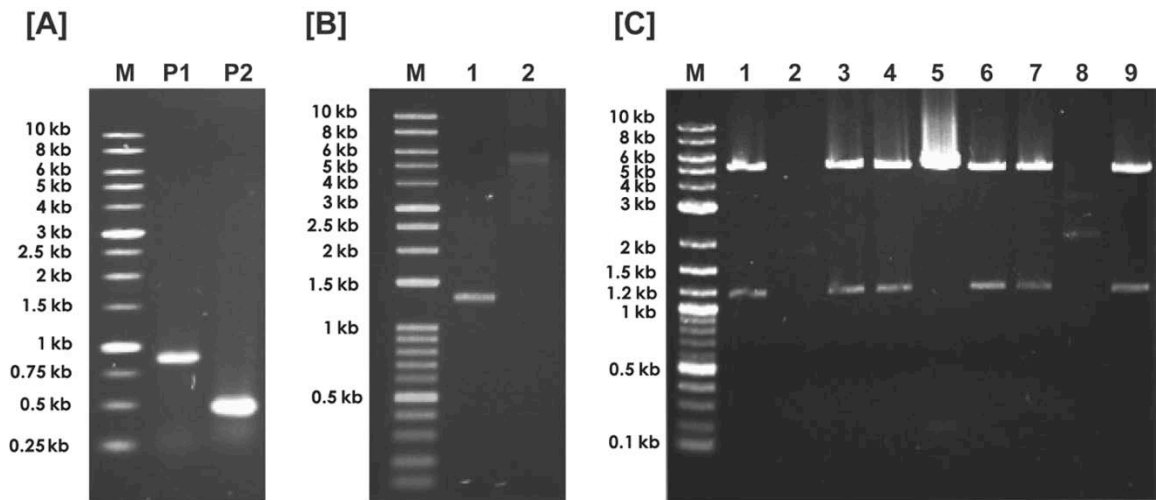


Figura 22. Construcción del vector de expresión pAM2. Geles de agarosa al 1% teñidos con bromuro de etidio [A] M: marcador de peso molecular, P1 y P2: Amplicones correspondientes de los fragmentos P1 (900 pb) y P2 (500 pb) [B] M: marcador de peso molecular, 1: amplicón correspondiente al módulo de expresión del pAM2 digerido con *KpnI* y *SacI*, 2: pRS411 digerido con *KpnI* y *SacI* [C] M: marcador de peso molecular, 1-9: Patrón de restricción con *KpnI* y *SacI*, del pDNA obtenido de las clonas candidatas para obtener el pAM2.

Construcciones para evaluar la complementación funcional de los diferentes miembros de la familia EFL1

Para evaluar si la función de las proteínas EFL1 se encontraba evolutivamente conservada, decidimos obtener la secuencia codificante de al menos un miembro de cada phyla del dominio Eukarya y ensayar su complementación genética en levadura. Sin embargo, a pesar de que intentamos obtener la secuencia codificante de la proteína EFL1 de organismos como *Toxoplasma gondii*, *Dictyostelium discoideum* y *Arabidopsis thaliana*, sólo fuimos capaces de clonar las secuencias correspondientes a ambas isoformas humanas, la proteína de ratón, la proteína de *S. cerevisiae* (control positivo) y la proteína de *S. pombe*.

En la Figura 23A se pueden observar los amplicones correspondientes a las secuencias codificantes de las proteínas ScEfl1, HsEFL1-1 y HsEFL1-2, mientras que en la Figura 23B se muestran las secuencias codificantes de MmEFL1 y SpEfl1. Como puede notarse, todas estos fragmentos de ADN oscilan entre los 3 kb y 3.5 kb de longitud y codifican proteínas que van de los 1,000 a los 1,127 aminoácidos.

Es claro que algunos de los amplicones presentaban fragmentos inespecíficos; para eliminarlos las bandas de interés se purificaron del gel y posteriormente clonadas en el vector pAM1. Las secuencias codificantes de las proteínas humanas se insertaron entre los sitios *Sall* y *BamHI*, y la secuencia *ScEFL1* se clonó entre los sitios *Sall* y *SpeI*. Esta estrategia resultó demasiado complicada, consumió mucho tiempo y después de la transformación de las células bacterianas con las reacciones de ligación el número de clones transformantes para escrutinio resultaba demasiado bajo. Debido a lo anterior, decidimos implementar la estrategia de clonación por recombinación homóloga para las secuencias *MmEFL1* y *SpEFL1*. A diferencia de la primer estrategia, ésta última tomó menor cantidad de tiempo, sin embargo, no resolvió el problema que experimentamos con el número de clones transformantes. A pesar de estas complicaciones experimentales, y como puede observarse en la Figura 24 (carriles 1-5), todas estas construcciones fueron obtenidas exitosamente. El patrón electroforético obtenido después de la digestión de los plásmidos correspondientes con *Sall/SpeI* y con *Sall/BamHI* para las construcciones humanas, muestra que los insertos de interés (entre 3 kb – 3.5 kb) fueron clonados adecuadamente en el pAM1 (~ 6 kb). Por otro lado, para evaluar un papel de los aEF2 relacionado con la función de las proteínas EFL1, decidimos evaluar su función mediante complementación genética. La secuencia codificante de la proteína AfEf-2 se sintetizó por GenScript USA Inc. y posteriormente, se subclonó en nuestro laboratorio, entre los sitios *Sall* y *SpeI* del vector pAM1. De la misma manera que las construcciones anteriores, el patrón de restricción de esta construcción con las enzimas *Sall* y *SpeI* se muestra en la Figura 24. Pueden notarse la presencia de un par de fragmentos, uno de ellos correspondiente al pAM1 abierto (6 kb) y otro que pertenece a la secuencia *AfEF-2* (2.2 kb); lo anterior demuestra que el marco abierto de lectura de la proteína AfEf-2 fue insertado correctamente en el pAM1. Finalmente, es importante mencionar que una vez identificadas las construcciones de interés, todos los insertos fueron secuenciados. Este análisis reveló que las secuencias de interés habían sido clonadas exitosamente y que no presentaban mutaciones.

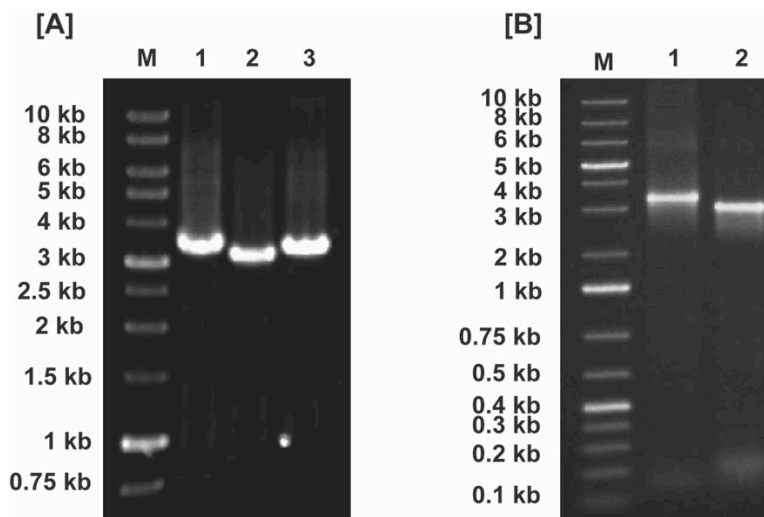


Figura 23. Amplificación de las secuencias codificantes de diferentes miembros de la familia de proteínas EFL1. Geles de agarosa al 1 % teñidos con bromuro de etidio. [A] M: marcador de peso molecular, amplicones correspondientes a los miembros de EFL1 de 1: *S. cerevisiae*, 2: *H. sapiens* isoforma 2 y 3: *H. sapiens* isoforma 1. [B] M: marcador de peso molecular, amplicones correspondientes a los miembros de EFL1 de 1: *M. musculus* y 2: *S. pombe*.

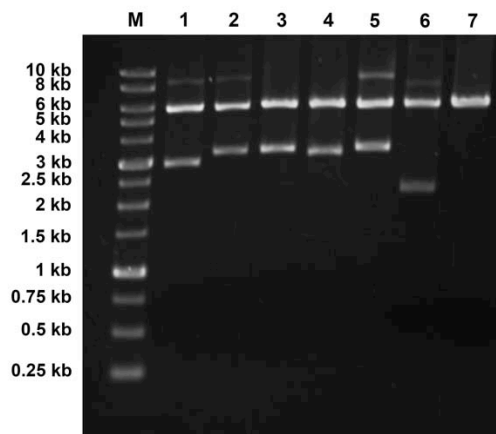


Figura 24. Clonación de las secuencias codificantes de diferentes miembros de la familia de proteínas EFL1 en el vector pAM1. Geles de agarosa al 1% teñidos con GelRed (Biotium, Inc.). Patrón de restricción de las construcciones que expresan los miembros de la familia EFL1, empleando *Sall/Bam*HI para los miembros humanos, y *Sall/Spel* para el resto.. M: marcador de peso molecular, 1: pAM1·SpEfl1, 2: pAM1·ScEfl1, 3: pAM1·MmEFL1, 4: pAM1·HsEFL1-2, 5: pAM1·HsEFL1-1, 6: pAM1·AfEF2 y 7: pAM1.

Modificación de las construcciones que expresan a las proteínas EFL1

Para detectar la expresión de los miembros de EFL1 estudiados en este trabajo, a través de la clonación en el vector pAM1, se agregó una etiqueta V5 en el C-terminal de cada una de las secuencias codificantes. Sin embargo, en ensayos de complementación funcional previos, observamos que el control positivo (ScEfl1) fusionado a esta marca no lograba rescatar el defecto de crecimiento en las levaduras provocado por la pérdida de su proteína endógena. Reportes previos sugerían que etiquetas como la proteína verde fluorescente, así como la pérdida de algunos residuos en el C-terminal de ScEfl1 afectaban su función [78]. Para evitar que la etiqueta impidiera la función normal de los miembros de EFL1 evaluados, decidimos agregar una señal de paro (TAA) entre el último codón de cada una de las secuencias codificantes EFL1 y la secuencia de la etiqueta V5, permitiendo así que las proteínas fueran expresadas de manera nativa. Lo anterior se realizó empleando un método de mutagénesis dirigida mediante PCR de cadena sencilla [115]. Por otra parte, dado que este comportamiento se observó antes de obtener la secuencia codificante de la proteína AfEf-2, se pidió a GenScript que la etiqueta V5 fuera adicionada en el N-terminal, por esta razón no fue necesario corregir la construcción que la expresaba.

Antes de evaluar por secuenciación si la adición del codón de paro había sido exitosa, los plásmidos candidatos fueron identificados mediante patrón de restricción con la enzima *HindIII*. Las Figuras 25A-B muestran los resultados de este escrutinio. El patrón de restricción observado en los geles de agarosa para las construcciones pAM1·ScEFL1 (fragmentos en 6.5 kb y 3 kb), pAM1·SpEFL1 (fragmentos en 6 kb, 1.5 kb, 0.8 kb y 0.6 kb), pAM1·MmEFL1 (7.8 kb, 2 kb y 0.3 kb), pAM1·HsEFL1-2 (6.2 kb, 1.2 kb, 1 kb, 0.5 kb y 0.3 kb) y pAM1·HsEFL1-1 (6.3 kb, 1.2 kb, 1 kb, 0.5 kb y 0.3 kb), con excepción de las clonas 2, carriles 3 y 8, de los plásmidos pAM1·SpEFL1 y pAM1·HsEFL1-2, respectivamente, correspondió con el análisis teórico realizado con el software Serial Cloner 2.6.1.

A partir del escrutinio anterior, decidimos secuenciar los insertos en las construcciones pAM1·ScEFL1 (clona 1, carril 1), pAM1·SpEFL1 (clona 2, carril 4), pAM1·MmEFL1 (clona 1, carril 5), pAM1·HsEFL1-2 (clona 1, carril 7) y pAM1·HsEFL1-1 (clona 1, carril 1), (Figura 25B). Los resultados de la secuenciación

mostraron que el codón de paro se había agregado adecuadamente a cada construcción y que la secuencia del resto del inserto no presentaba mutaciones. Estas construcciones, junto con la construcción pAM1·AfEF-2, se emplearon para realizar los ensayos de complementación funcional.

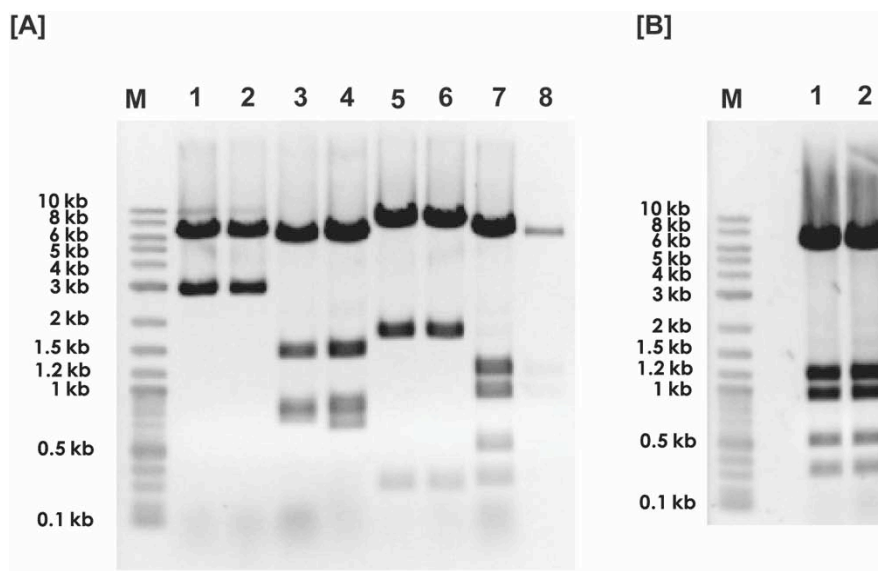


Figura 25. Análisis de las construcciones para la expresión de los miembros de la familia EFL1 de manera nativa. Geles de agarosa al 1% teñidos con bromuro de etidio. Patrón de restricción obtenido con *HindIII*, de las construcciones que expresan los miembros de la familia EFL1 de manera nativa. [A] M: marcador de peso molecular, 1-2: pAM1·ScEfl1, 3-4: pAM1·SpEfl1, 5-6: pAM1·MmEFL1, 7-8: pAM1·HsEFL1-2. [B] M: marcador de peso molecular, 1-2: pAM1·HsEFL1-1.

Construcciones para evaluar la complementación funcional de los diferentes miembros de la familia SBDS

Con el objetivo de evaluar la conservación funcional de las proteínas SBDS, decidimos clonar en el vector de expresión pAM2 (Figura 9) la secuencia codificante de la proteína de ScSdo1, SpSdo1, HsSBDS, MmSBDS, AfSbds y las proteínas quiméricas HsSBDS D2Sc y ScSdo1 D2Hs, en las que se intercambiaron entre sí su dominio 2 correspondiente.

El marco abierto de lectura de SpSdo1 se construyó por PCR de traslape. En la Figura 26C se muestran los amplicones correspondientes a los exones 1 (150 pb, carril 3) y 2 (600 pb, carril 2) que componen esta secuencia codificante. A partir del oligonucleótido reverso para amplificar el exón 1, se generó un sitio

complementario al exón 2 que mediante un tercer ciclo de PCR permitió ensamblar el fragmento codificante completo. El resto de las secuencias codificantes de los miembros de SBDS evaluados se obtuvo mediante PCR convencional. En la Figura 26A-B se muestran los productos de PCR correspondientes a las secuencias *HsSBDS*, *MmSBDS* y *ScSDO1*. Todos los amplicones SBDS tienen una longitud aproximada de 750 pb y codifican proteínas cuya longitud oscila alrededor los 250 aminoácidos. Los fragmentos codificantes de la proteína AfSbds y las quimeras *HsSBDS* D2Sc y *ScSDO1* D2Hs no se muestran, sin embargo, es importante aclarar que los resultados obtenidos para estos fragmentos fueron idénticos que para las secuencias anteriores.

Una vez obtenidos los amplicones de los miembros SBDS, se clonaron en el vector de tránsito pGEM-T easy. El sitio de clonación en este vector está flanqueado por dos secuencias de reconocimiento por la enzima de restricción *EcoRI*. Dado que ninguna de las secuencias codificantes de los miembros SBDS presenta sitios de corte por esta enzima decidimos utilizar esta endonucleasa para identificar los plásmidos con los insertos de interés. Los resultados de la digestión del pDNA obtenido de 4 clonas candidatas para cada construcción se muestran en las Figura 27A-D. Como puede observarse, en todos los carriles se presenta una banda de 3 kb que corresponde al pGEM T easy abierto, y una banda más de aproximadamente 750 pb que pertenece a cada uno de los insertos. Estos resultados revelaron que todos los insertos correspondientes a los fragmentos codificantes de las proteínas SBDS habían sido clonados exitosamente en el vector de tránsito.

Posteriormente, los insertos de las construcciones anteriores se secuenciaron y sólo aquellos fragmentos codificantes que no presentaron cambios en su secuencia de nucleótidos fueron subclonados entre los sitios *Sall* y *SpeI* del pAM2. En la Figura 28 se muestran los patrones de restricción obtenidos con *Sall* y *SpeI* de las construcciones pAM2·ScSDO1, pAM2·SpSDO1, pAM2·HsSBDS, pAM2·MmSBDS, pAM2·AfSBDS, pAM2·ScSDO1 D2Hs, pAM2·HsSBDS D2Sc, comparado con el patrón obtenido para el pAM2 vacío. Como puede notarse, todos los carriles muestran una banda correspondiente al vector de expresión

abierto (6.5 kb) y una banda más perteneciente a cada una de las secuencias codificantes de los miembros SBDS (~750 pb).

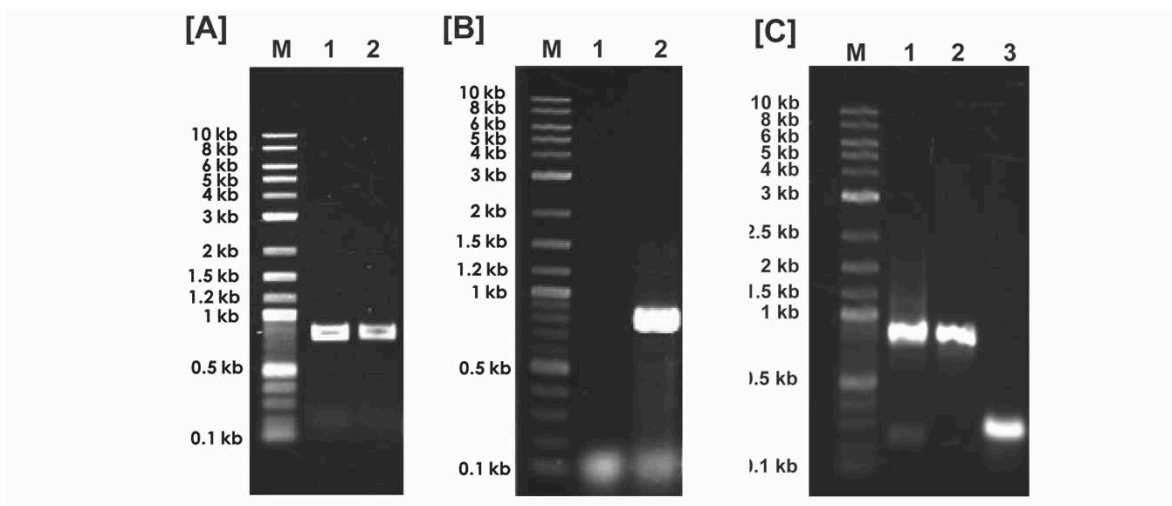


Figura 26. Amplificación de las secuencias codificantes de diferentes miembros de la familia de proteínas SBDS. Geles de agarosa al 1% teñidos con bromuro de etidio. [A] M: marcador de peso molecular, amplicones correspondientes a las secuencias codificantes de las proteínas SBDS de 1: *M. musculus* y 2: *H. sapiens*. [B] M: marcador de peso molecular. 1: control negativo, 2: Amplicón correspondiente a la secuencia codificante de la proteína Sdo1 de *S. cerevisiae*. [C] M: marcador de peso molecular, amplicón correspondiente 1: a la secuencia codificante completa de la proteína Sdo1 de *S. pombe*, 2: al exón 2 de la secuencia codificante de la proteína Sdo1 de *S. pombe* y 3: al exón 1 de la secuencia codificante de la proteína Sdo1 de *S. pombe*.

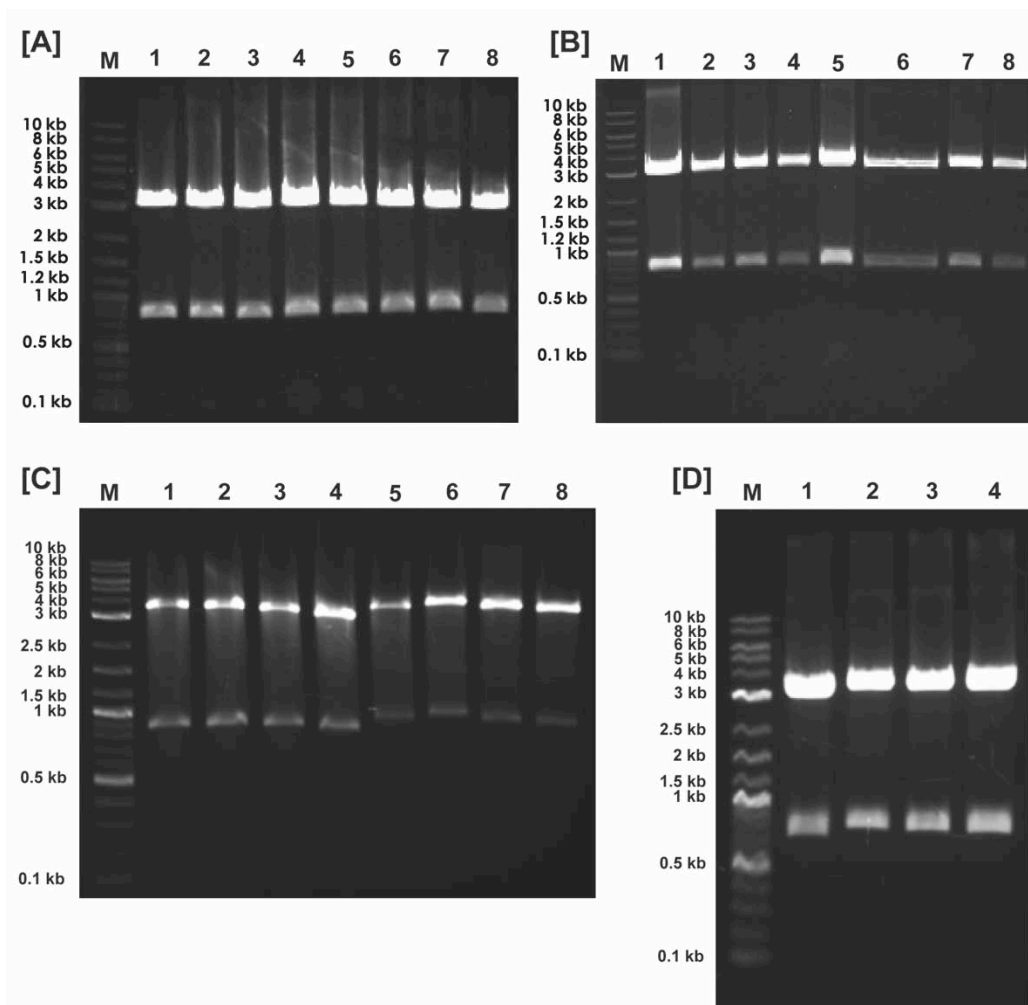


Figura 27. Clonación de las secuencias codificantes de diferentes miembros de la familia de proteínas SBDS en el vector pGEMTeasy. Geles de agarosa al 1% teñidos con GelRed (Biotium, Inc.). [A] M: marcador de peso molecular, 1-4: pDNA digerido con *EcoRI*, para identificar las construcciones pGEM T·HsSBDS y 5-8: pGEM T·MmSBDS [B] M: marcador de peso molecular, pDNA digerido con *EcoRI* para identificar las construcciones 1-4: pGEM T·ScSDO1 y 5-8: pGEM T·SpSDO1. [C] M: Marcador de peso molecular, pDNA digerido con *EcoRI* para identificar las construcciones 1-4: pGEM T·ScSDO1 D2Hs y 5-8: pGEM T·HsSBDS D2Sc [D] M: Marcador de peso molecular, pDNA digerido con *EcoRI* para identificar las construcciones 1-4: pGEM T·AfSBDS.

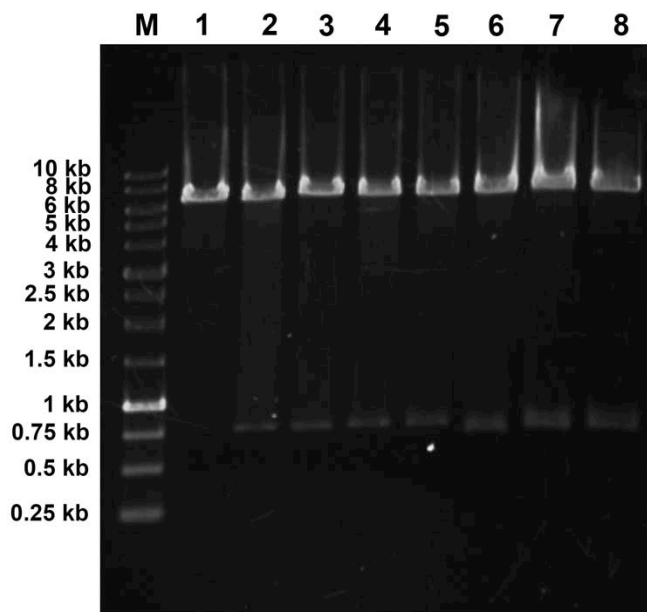


Figura 28. Clonación de las secuencias codificantes de diferentes miembros de la familia de proteínas SBDS en el vector pAM2. Geles de agarosa al 1 % teñidos con GelRed (Biotium, Inc.). M: marcador de peso molecular, pDNA digerido con *SalI* y *SpeI* de las construcciones 1: pAM2, 2: pAM2·ScSDO1, 3: pAM2·SpSDO1, 4: pAM2·HsSBDS, 5: pAM2·MmSBDS 6: pAM2·AfSBDS, 7: pAM2·HsSBDS D2Sc, 8: pAM2·ScSDO1 D2Hs.

Detección de la expresión de las proteínas EFL1 y SBDS

En la Figura 29A se muestra el resultado del inmunoblot realizado para detectar la expresión de los miembros EFL1 y la proteína AfEF-2 empleando un anticuerpo contra le etiqueta V5 acoplado a una peroxidasa de rábano. Mientras que el extracto protéico obtenido de células transformadas con el pAM1 vacío (control negativo) no mostró ninguna señal, los extractos purificados de levaduras transformadas con los plásmidos que contienen las secuencias codificantes de los miembros EFL1 presentaron bandas en ~127 kDa que corresponden con el peso teórico de estas proteínas. Por otro lado, el extracto obtenido de levaduras transformadas con la construcción pAM1·AfEF-2 mostró una señal en ~98 kDa que coincide con el peso calculado para la GTPasa Ef-2 de esta arquea. Es posible observar otras señales por debajo de las bandas de interés. Debido a que el extracto del control negativo no presenta dichas señales se sospecha que estos fragmentos son producto de la degradación de las proteínas EFL1 y no de una reacción inespecífica del anticuerpo.

Para evaluar si las proteínas SBDS y las proteínas quiméricas se expresaban en levadura, obtuvimos extractos proteicos derivados de células transformadas con los plásmidos que contienen sus secuencias codificantes y detectamos su presencia mediante inmunoblot usando un anticuerpo contra la etiqueta c-myc acoplado a una peroxidasa de rábano (Figura 29B). Con excepción del extracto control (vector pAM2 vacío), el resto de ellos presentaron bandas de ~ 36 kDa y 26 kDa. El peso teórico para las proteínas SBDS es de 28 kDa, sin embargo, diversos trabajos han reportado que estas proteínas migran en un gel SDS-PAGE con un peso molecular experimental de 36 kDa [86,87]. Lo anterior indica que las bandas observadas en los carriles 2-7 corresponden a los miembros SBDS y quimeras evaluados en este trabajo. Por otro lado, la banda en el extracto 8 de 26 kDa corresponde con el peso teórico de la proteína AfSbds.

Es claro que los niveles de expresión de estas proteínas son diferentes entre sí; mientras que las proteínas SpSdo1 y quimeras HsSBDS 2Dsc y ScSdo1 D2Hs presentan la señal más intensa, las proteínas ScSdo1, HsSBDS, MmSBDS y AfSbds apenas son detectables. Estos cambios en los niveles de expresión podrían ser influenciados por la estabilidad de cada proteína y las condiciones metabólicas en las que se encontraban las células cuando fueron cosechadas para la obtención de los extractos proteicos.

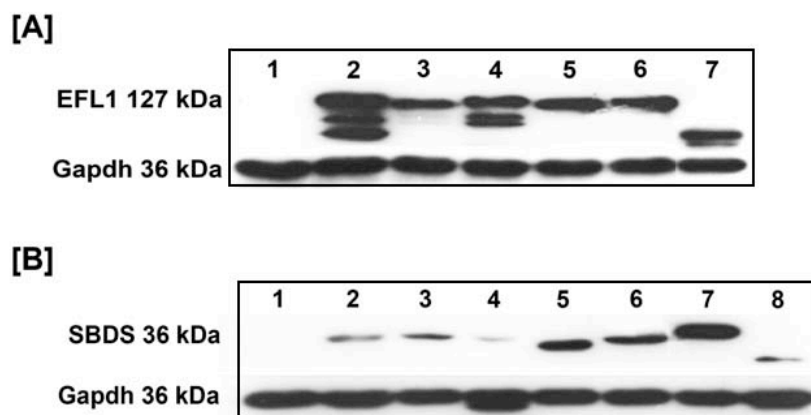


Figura 29. Detección de la expresión de distintos ortólogos de EFL1 y SBDS en células de levadura mediante inmunoblots [A] Detección las proteínas EFL1 en extractos de proteína total de las levaduras transformantes. 1: vector vacío, 2: Efl1 *S. cerevisiae*, 3: Efl1 *S. pombe*, 4: EFL1 *H. Sapiens* isoforma 1 5: EFL1 *H. Sapiens* isoforma 2, 6: EFL1 *M. Musculus*, 7: Ef-2 *A. fulgidus* (98 kDa). [B] Detección de la expresión de las proteínas SBDS en extractos de proteína total de las levaduras transformantes. 1: vector vacío 2: Sdo1 *S. cerevisiae* 3: Sdo1 *S. pombe* 4: SBDS *M. Musculus* 5: SBDS *H. Sapiens*, 6: quimera HsSBDS D2Sc 7: quimera ScSdo1 D2Hs 8: SbdS *A. fulgidus* (26 kDa). Gapdh: control de carga.

Ensayos de complementación funcional de los miembros de la familia EFL1

La búsqueda de ortólogos de EFL1 en UniprotKB reveló que estas proteínas se encuentran altamente conservadas y son exclusivas del dominio Eukarya. Para demostrar, además, que su función estaba evolutivamente conservada, nosotros llevamos a cabo ensayos de complementación genética en células de levadura *efl1Δ::natMX4*, con las proteínas EFL1 de ratón, la isoforma humana 1 e isoforma humana 2 (para observar un posible papel en biogénesis ribosomal), y el ortólogo de *S. pombe* (Figura 30). Por otro lado, la ausencia de proteínas EFL1 en arqueas y la relación filogenética de estas proteínas con GTPasas aEF2, sugiere que además de participar en el proceso de translocación durante la síntesis de proteínas, las proteínas aEF2 podrían estar implicadas en la liberación de aIF6 de la subunidad 50S durante su biogénesis en estos procariotes; para evaluar esta función, decidimos realizar ensayos de complementación genética en levadura empleando la proteína EF-2 de *A. fulgidus*.

El crecimiento de progenie meiótica haploide *EFL1* en el medio de cultivo sin cloNAT demostró que las esporas utilizadas para estos ensayos se encontraban viables (Figura 30A). En la Figura 30B se muestra la selección de la progenie

meiótica *eff1Δ::natMX4* en el medio de cultivo suplementado con antibiótico. Los resultados revelaron que sólo la proteína ScEfl1 (control positivo) expresada de manera extracromosomal es capaz de rescatar el defecto de crecimiento provocado por la pérdida del gen *EFL1*. Lo anterior podría sugerir que la función de estas proteínas no está conservada evolutivamente o que su papel es especie específico, sugiriendo coevolución con las proteínas SBDS.

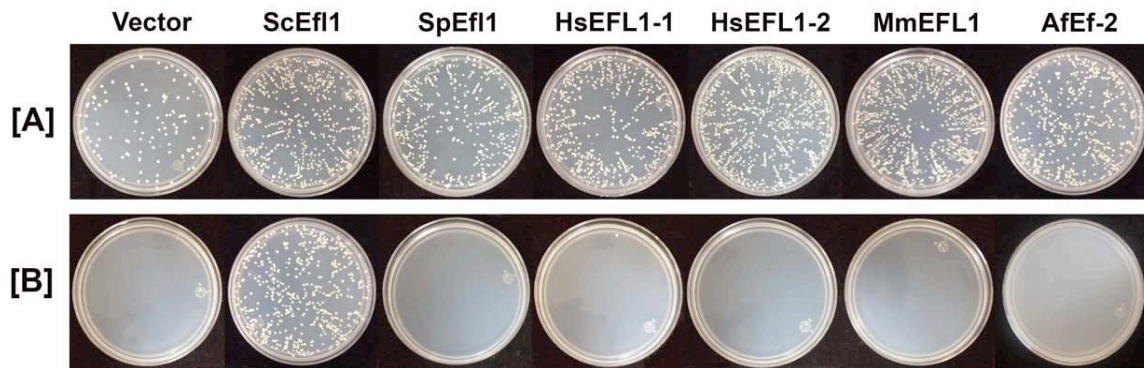


Figura 30. Ensayos de complementación genética de la familia EFL1. [A] Crecimiento de la progenie meiótica *EFL1* sobre medio de crecimiento sintético suplementado con canavanina y tialisina, después de 3 días de incubación a 37 °C. [B] Crecimiento de la progenie meiótica *eff1Δ* sobre medio de crecimiento sintético suplementado con canavanina, tialisina y cloNAT, después de 3 días de incubación a 37 °C.

Ensayos de complementación funcional de los miembros de la familia SBDS

Boocock *et. al* [11] ensayaron la complementación genética de las proteínas SBDS y observaron que su función se comportaba de manera especie específica. Debido a que nosotros queríamos demostrar coevolución en el reconocimiento entre las proteínas EFL1 (tal como sugirieron los resultados obtenidos de los ensayos de complementación funcional para estas proteínas) y SBDS, quisimos comprobar los resultados obtenidos por este grupo de investigación y evaluamos la conservación funcional de las proteínas SpSdo1, HsSBDS, MmSBDS, AfSbds y quimeras HsSBDS D2Sc y ScSdo1 D2Hs, en células *sdo1Δ::kanMX6* seleccionadas sobre medio adicionado con G418 (Figura 31). Como se reportó en [11], de las proteínas ensayadas, sólo ScSdo1, SpSdo1 y la proteína quimérica HsSBDS D2Sc, fueron capaces de complementar la función de la proteína endógena. Sin embargo,

observamos también, que la proteína quimérica ScSdo1 D2Hs, era capaz de complementar parcialmente la función de ScSdo1. El ensayo de complementación con la proteína quimérica HsSBDS D2Sc reveló que el dominio 2 es sumamente importante para el reconocimiento óptimo entre ScEfl1 y ScSdo1, sin embargo, los resultados con ScSdo1 D2Hs demuestran que los dominios 1 y 3 de ScSdo1 también promueven la interacción, aunque no eficientemente, con ScEfl1.

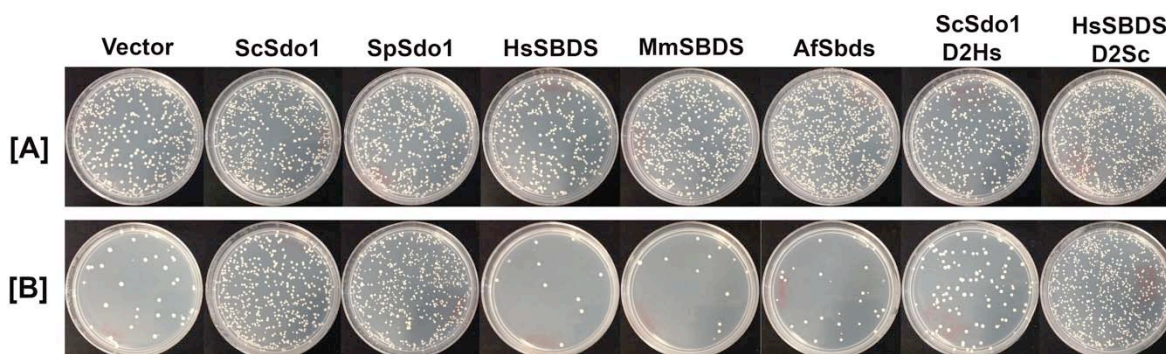


Figura 31. Ensayos de complementación genética de la familia SBDS. [A] Crecimiento de la progenie meiótica *SDO1* sobre medio de crecimiento sintético suplementado con canavanina y tialisina, después de 5 días de incubación a 37 °C. [B] Crecimiento de la progenie meiótica *sdo1Δ* sobre medio de crecimiento sintético suplementado con canavanina, tialisina y G418, , después de 5 días de incubación a 37 °C.

Ensayos de complementación funcional de los pares de proteínas EFL1/SBDS

Los resultados obtenidos de los ensayos de complementación genética con las proteínas EFL1 y SBDS individualmente, así como la especie especificidad observada para el dominio 2 de ScSdo1 y la alta variabilidad del dominio 2 de las proteínas EFL1 mostrada en su estudio bioinformático, indicaban que las proteínas heterólogas eran incapaces de complementar la función de las proteínas de levadura debido a que no eran capaces de reconocer a su contraparte en este organismo; esto sugería fuertemente, un posible fenómeno de coevolución entre ellas. Para demostrar este comportamiento, llevamos a cabo ensayos de complementación genética empleando los pares EFL1 y SBDS correspondientes a cada organismo ensayado previamente.

En la Figura 32 se muestran los ensayos de complementación de los pares ortogonales EFL1/SBDS de *S. pombe*, *H. sapiens*, *M. musculus* y *A. fulgidus*. La presencia de progenie meiótica *EFL1/SDO1* sobre medio sin antibióticos muestra que las esporas usadas en estos ensayos se encontraban viables. Por otro lado, en las cajas de Petri con G418 y cloNAT podemos notar que sólo el par ScEfl1/ScSdo1 (control positivo) complementa la función de los genes eliminados. Lo anterior puede indicar que a pesar de que ambas familias de proteínas se encuentran altamente conservadas en el dominio Eukarya, probablemente no participan en los mismos procesos celulares, o que el complejo EFL1·SBDS reconoce algún otro componente celular de manera especie-específica.

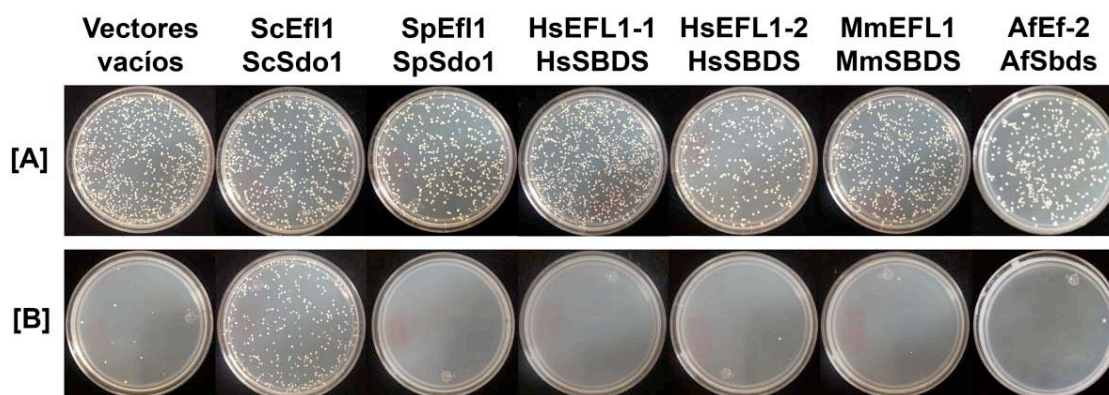


Figura 32. Ensayos de complementación genética de los pares de proteínas EFL1 y Sdo1/SBDS. [A] Crecimiento de la progenie meiótica *EFL1/SDO1* sobre medio de crecimiento sintético suplementado con canavanina y tialisina, después de 5 días de incubación a 37 °C. [B] Crecimiento de la progenie meiótica *efl1Δ/sdo1Δ* sobre medio de crecimiento sintético suplementado con canavanina, tialisina, cloNAT y G418, después de 5 días de incubación a 37 °C.

Finalmente, evaluamos el comportamiento de las quimeras HsSBDS D2Sc y ScSdo1 D2Hs con las proteínas ScEfl1 y HsEFL1 (Figura 33). Para estos ensayos, una vez más la progenie haploide transformada con los pares de plásmidos pAM1/pAM2 y pAM1·ScEFL1/pAM2·ScSDO1, se emplearon como control negativo y positivo, respectivamente. El crecimiento sobre medio de selección demostró que el par ScEfl1/HsSBDS D2Sc complementa eficientemente la función de las proteínas eliminadas, mientras que las proteínas ScEfl1/ScSdo1 D2Hs lo

hacen parcialmente. Lo anterior concuerda con los resultados observados durante la evaluación de la conservación funcional de la familia SBDS.

Por otro lado, la incapacidad de complementación funcional por parte de los pares HsEFL1-1/ScSdo1 D2Hs y HsEFL1-1/HsSBDS D2Hs indican que las proteínas quiméricas sólo complementan la función de ScSdo1 en presencia de la proteína ScEfl1 endógena o expresada de manera extracromosomal. En conjunto, estos ensayos demuestran que el defecto de crecimiento sólo es rescatado con proteínas de levadura o con proteínas que contienen fragmentos importantes de la misma, sugiriendo que este evento de maduración de la subunidad ribosomal 60S podría estar conservado evolutivamente y ocurre de manera especie específica.

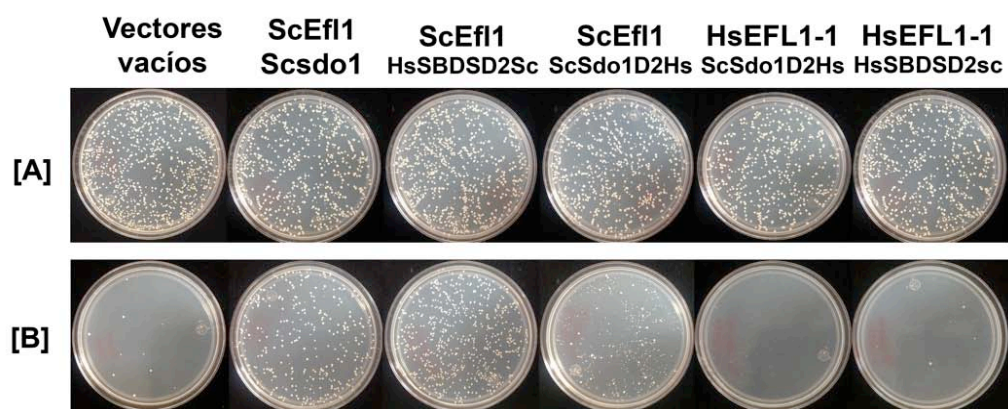


Figura 33. Ensayos de complementación genética de los pares de proteínas EFL1 y las quimeras SBDS. [A] Crecimiento de la progenie meiótica *EFL1/SDO1* sobre medio de crecimiento sintético suplementado con canavanina y tialisina, después de 5 días de incubación a 37 °C. [B] Crecimiento de la progenie meiótica *efl1Δ/sdo1Δ* sobre medio de crecimiento sintético suplementado con canavanina, tialisina, cloNAT y G418, después de 5 días de incubación a 37 °C.

7 Discusión de resultados

La biogénesis ribosomal es uno de los procesos con mayor importancia en el contexto celular [61]. A través de él, la célula dispone de la maquinaria que la provee de todas las proteínas que requiere para su desarrollo y mantenimiento. Gracias al establecimiento de técnicas como la purificación por afinidad en tándem, y dada la oportunidad de acoplarla a la espectrometría de masas, ha sido posible aislar diversos intermediarios de este proceso y caracterizar a detalle su composición; incluso, recientemente se obtuvo la estructura de la partícula pre-90S mediante criomicroscopia, aportando aspectos importantes acerca de su estructura y composición [121]. Estas investigaciones han revelado que el ensamblaje del ribosoma en eucariontes, además de la síntesis, procesamiento y asociación del rRNA ribosomal con proteínas estructurales, requiere de la participación de cerca de 75 snoRNA y más de 200 proteínas accesorias implicadas en su maduración [54,68,122]. La GTPasa Efl1 y su intercambiador de nucleótidos, Sdo1, corresponden a dos factores accesorios encargados de evaluar la integridad del tallo ribosomal y del sitio P, así como de liberar a Tif6 de la subunidad ribosomal mayor [80,95,123,124]. En este trabajo, caracterizamos la distribución filogenética de las proteínas EFL1, su relación evolutiva con otras GTPasas ribosomales, y evaluamos su conservación funcional de manera independiente y a la par con la familia de proteínas SBDS.

UniprotKB es una base de datos que reúne información funcional de proteomas correspondientes a organismos de todos los dominios de la vida [105]. A pesar de que este sitio cuenta con 478 proteomas de Archaea y 51,255 de Bacteria, en nuestra búsqueda sólo identificamos proteínas de la familia EFL1 en miembros del dominio Eukarya (Tabla 21). De acuerdo con [11,15,106,116], el dominio Eukarya puede organizarse en seis Phylum que corresponden a Opisthokonta, Amoebozoa, Excavata, Archaeplastida, Chromalveolata y Rhizaria. Los Opisthokonta corresponden a uno de los Phylum mejor soportados por estudios filogenéticos; este grupo está compuesto por organismos que presentan células

con un flagelo posterior simple. El grupo Amoebozoa consiste en organismos unicelulares que utilizan un pseudópodo tipo dedo para moverse y alimentarse [125]. Por otro lado, el Phylum Excavata corresponde a protistas que en su mayoría son anaerobios; generalmente, estos organismos se caracterizan por presentar hidrogenosomas o mitosomas en vez de mitocondria. Algunos patógenos humanos como especies de los géneros *Trypanosoma* y *Leishmania*, pertenecen a este Phylum [126]. El grupo Archaeplastida está compuesto por algas microscópicas, algas rojas y organismos verdes que corresponden a las algas verdes y plantas terrestres [127]. El Phylum Chromalveolata es uno de los más interesantes; dos de sus subgrupos más importantes son Stramenopila y Alveolata, que presentan protistas de importancia ecológica como algas y diatomeas, y patógenos humanos del género *Plasmodium* y *Toxoplasma*, respectivamente [128]. Finalmente, el grupo Rhizaria se compone por organismos unicelulares de vida libre que en su mayoría utiliza finos pseudópodos filosos para alimentarse. Adicionalmente, este Phylum contiene parásitos de plantas tipo hongos y algunos parásitos humanos [129]. Los miembros de EFL1 identificados se encuentran ampliamente distribuidos en organismos correspondientes a todos los subgrupos de los Phylum que componen el dominio Eukarya, sin embargo, observamos un aumento en el número de especies con miembros de EFL1 en Fungi, Metazoa, Streptophyta y Alveolata. Lo anterior es sin duda un reflejo de la disponibilidad de genomas secuenciados y anotados a la fecha en la base de datos que empleamos durante nuestra búsqueda.

En algunos organismos, principalmente pluricelulares, identificamos dos o más proteínas EFL1 posibles. Al comparar cada uno de los pares entre sí, observamos que éstos podían clasificarse en dos grupos: pares de proteínas EFL1 con un alto porcentaje de identidad y diferencias en una región particular (Figura 11), y pares de proteínas EFL1 con alto grado de divergencia en su estructura primaria (Figura 12). En los pares de proteínas que presentan un alto grado de identidad entre sí, generalmente, una de ellas carece de algunas regiones como los submotivos G1, G2 y G3, que son importantes para la catálisis. Esta clase de proteínas se habían observado anteriormente en humano y otros mamíferos, en donde existen dos

isoformas de EFL1. Particularmente, la segunda isoforma humana no cuenta con el motivo G2 y parte del G1 [130] (Figura 11). De acuerdo con la función de estas regiones, estas proteínas no podrían hidrolizar GTP [98,131], sin embargo, dado que cuentan con los motivos G4 y G5, cabe la posibilidad de que aun sean capaces de unir nucleótidos de guanina. Es importante mencionar, que esta clase de proteínas podría originarse a partir de dos mecanismos: (1) por empalme alternativo, dando origen a isoformas como en el caso de HsEFL1-1 y HsEFL1-2, o por un inicio de la traducción alternativo, como se sospecha para el par de proteínas EFL1 de *D. melanogaster*. Hasta el momento se desconoce la función de estas proteínas con características estructurales similares.

Por otra parte, se encontraron pares EFL1 con bajo porcentaje de identidad en miembros que pertenecen al phyla Archaeplastida. Estos pares de proteínas EFL1 cuentan con todos los componentes de su dominio catalítico y sólo divergen en la composición de su secuencia de aminoácidos (Figura 12). Recientemente ha resurgido la idea de ribosomas especializados que corresponden a subpoblaciones de partículas ribonucleoprotéicas con diferente composición [132–135]. Estas diferencias se originan de la expresión y modificaciones postraduccionales diferenciales de las proteínas estructurales, los cambios en el rRNA y la actividad de las proteínas accesorias que participan en su síntesis. Como resultado de los cambios en su composición, estos traductores celulares son capaces de ocupar mecanismos de traducción alternativos y cambiar la eficiencia con la que diversas proteínas son producidas dependiendo del estado celular. Por lo anterior, en el futuro el estudio de estas proteínas EFL1 adicionales, identificadas en organismos superiores, podría resultar sumamente interesante, ya que además de que podrían presentar funciones extraribosomales, éstas podrían participar en mecanismos relacionados con la especialización de las partículas ribosomales.

En nuestra búsqueda de miembros de EFL1 detectamos proteínas pertenecientes a las familias de factores de elongación EF-2 y EF-G. Estas GTPasas claramente correspondían a falsos positivos, sin embargo, su presencia resultó de gran interés para nosotros. Diversos autores habían sugerido con anterioridad la relación evolutiva entre EFL1 y estas translocasas [78,95,136]; en efecto, gracias a la

información obtenida mediante estudios sobre la estructura y función de EF-2, es como hemos logrado interpretar mejor el papel que EFL1 representa en la maduración ribosomal. El árbol filogenético construido con estas proteínas muestra claramente que comparten un origen evolutivo en común (Figura 17).

Se conoce muy bien, que durante la evolución de una proteína, su secuencia de aminoácidos cambia más rápido que su estructura tridimensional [137]. Por lo tanto, a pesar de la baja identidad en secuencia y dada su relación evolutiva, es posible que las GTPasas EFL1, EF-2 y EF-G, presenten una estructura terciaria similar. El alineamiento realizado con las estructuras cristalográficas de las proteínas ScEf-2 (PDB ID: 1N0V) y EF-G de *Thermus thermophilus* (TtEF-G, PDB ID: 4M1K) soporta muy bien esta hipótesis. A pesar de que estas proteínas pertenecen a organismos evolutivamente lejanos y la identidad entre sus secuencias es relativamente baja, el plegamiento global de ambas enzimas se encuentra conservado; este comportamiento también se ha observado en otras proteínas como las lipocalinas, cuya relación evolutiva no se demostró hasta después de la comparación de sus estructuras tridimensionales [138–140].

De acuerdo con la regla de superposición de especies [4], la presencia de miembros de eEF-2 y EFL1, en especies mutuamente excluyentes, sugiere que los genes que codifican a ambas familias de proteínas fueron originados por duplicación antes de que los eucariontes actuales divergieran. James y Tawfik han descrito un modelo que explica cómo uno de los genes parálogos obtiene una nueva función. Ellos proponen que las proteínas pueden presentar promiscuidad funcional y diversidad conformacional. El estado conformacional predominante de una proteína le permite llevar a cabo su función nativa, mientras que la conformación alternativa, generalmente en menor proporción, es capaz de presentar una actividad secundaria. Después de un evento de duplicación, una de las copias puede experimentar cambios sin comprometer la actividad original; así estas mutaciones pueden optimizar la función alternativa y estabilizar la conformación promiscua [141].

Debido a que la síntesis de proteínas corresponde a un proceso esencial, es posible suponer que el ancestro común de las GTPasas EFL1 y eEF-2 participara en el movimiento del ribosoma durante la traducción, como lo hacen actualmente

las translocasas EF-2 y EF-G [97,99,142,143]. Por otra parte, el hecho de que organismos del dominio Archaea presenten miembros de proteínas como Sdo1/SBDS y Tif6/eIF6, pero no de EFL1, permite sospechar que la liberación del factor de antiasociación Tif6 era necesaria mucho antes de que las proteínas EFL1 aparecieran. En consecuencia, existe la posibilidad de que la célula aprovechara las características estructurales y funcionales de esta proteína ancestral, para evaluar el estado estructural de la subunidad ribosomal mayor durante su maduración y promover la liberación del factor de antiasociación, resultando en una función alternativa para este factor de elongación. Algunas características como el alto índice de conservación entre las proteínas eEF-2, la gran divergencia entre las GTPasas de maduración ribosomal, el mismo sitio de unión en el ribosoma [96] y probable mecanismo de acción [95,97], podría apoyar fuertemente la hipótesis de que EFL1 surgió a través de la optimización de una función promiscua del factor de elongación eEF-2 después de su duplicación.

Una de las diferencias principales entre las familias eEF-2 y EFL1, recae sobre su actividad [124]. Mientras que las translocasas catalizan adecuadamente la hidrólisis de GTP sin la necesidad de factores adicionales, las proteínas de maduración ribosomal requieren de la participación de las proteínas Sdo1/SBDS como intercambiadores de nucleótidos [81]. El alineamiento múltiple de secuencias entre miembros de EFL1 y eEF-2, no mostró diferencias importantes en los motivos que conforman el sitio catalítico, indicando que los cambios en su afinidad por los nucleótidos de guanina y su actividad catalítica, están determinadas por regiones alostéricas en EFL1 que probablemente se establecieron a lo largo de su divergencia. Esta clase de regiones se identificaron en EFL1 al comparar ambas GTPasas de organismos como *S. cerevisiae*, *S. pombe*, *H. sapiens*, *A. thaliana* y *P. falciparum*. El resultado de este análisis reveló que las proteínas EFL1 presentan inserciones en las regiones que corresponden a los dominios G, II, IV y V. Resulta interesante observar, que mientras el número y longitud de las inserciones se mantiene constante en los dominios G y V, varía en los dominios II y IV (Figura 18), que han sido identificados como los sitios que conducen el reconocimiento de las proteínas SBDS y la subunidad 60S,

respectivamente, lo que podría sugerir que estos fenómenos son conducidos de manera especie específica.

Debido a su localización, una de las principales diferencias que puede explicar los cambios en la actividad catalítica de EFL1 respecto a eEF-2 es la inserción observada en el dominio catalítico. Para conocer cómo su presencia impacta en la actividad de la familia GTPasas EFL1 será necesario realizar ensayos con mutantes carentes de esta región y estudios estructurales que permitan observar el efecto que tiene sobre el plegamiento del dominio G y su interacción con los nucleótidos de guanina. Por otra parte, las inserciones en el dominio II también resultan candidatas fuertes a explicar las diferencias funcionales entre ambas familias de enzimas. Se ha observado que este sitio está implicado en la interacción con ScSdo1 *in vitro* y por tanto resulta clara la importancia que representa para modular su actividad [87]. Llama la atención que la inserción del dominio II en ScEfl1 se encuentra intrínsecamente desordenada. El análisis de las secuencias de aminoácidos de HsEFL1, SpEFL1, MmEFL1, PfEFL1 y AtEFL1 con PONDR-FIT, un predictor de regiones intrínsecamente desordenadas, reveló que estas proteínas presentan una región sin estructura secundaria que corresponde a la inserción observada en el dominio II correspondiente. A pesar de que esto permite sospechar que las inserciones del dominio II en las proteínas EFL1 podrían ser regiones encargadas de modular su actividad catalítica al contacto con su correspondiente proteína SBDS, estudios recientes han revelado que la proteína ScEfl1 que carece de este fragmento, es capaz de complementar la función de la proteína endógena en un modelo de levadura [93]. Lo anterior demuestra que la inserción observada en el dominio II de las proteínas EFL1 es dispensable para su función *in vivo*. Nosotros sospechamos que este sitio podría hacer dependiente a la GTPasa del intercambiador de nucleótidos, por lo que al ser eliminada, existe la posibilidad de que EFL1 actúe de manera independiente a ScEfl1. Para establecer el papel de esta región sobre la función de EFL1, deberán realizarse estudios estructurales, funcionales y genéticos que impliquen el uso de mutantes que no contengan esta inserción.

Asimismo, como se describió anteriormente, la mayor diversidad entre los miembros de EFL1 analizados, respecto a la longitud y número de inserciones,

ocurre en los dominios II y IV. Es claro que estas regiones representan un papel relevante en la optimización de las proteínas EFL1, pero además, es posible proponer que dadas las diferencias que existen en estos sitios de una especie a otra, podrían estar relacionadas con un fenómeno de especificidad de especie tal como fue observado para el dominio 2 de las proteínas SBDS [11].

Por otra parte, el alineamiento múltiple de secuencias construido con miembros de EFL1 mostró 51 residuos invariables distribuidos particularmente dentro del dominio G, III y V (Figura 13). En el dominio catalítico, la mayoría de esta clase de residuos se presenta en los motivos G1-G4, que son importantes para el reconocimiento específico y catálisis de la hidrólisis de GTP. Además, pudimos notar que el motivo G5, implicado en la interacción con la base de guanina, no presenta residuos invariables. Estudios en GTPasas como Ras han mostrado que el grado de conservación en este sitio es bajo [98].

En el año 2001 Becam *et. al* caracterizaron por primera vez a la proteína ScEfl1. Durante su investigación observaron que la presencia de la proteína verde fluorescente en C-terminal de la GTPasa inhibía su función *in vivo*. Posteriormente, ensayos de complementación en donde se emplearon proteínas ScEfl1 carentes desde 13 a 29 residuos del C-terminal, mostraron que estas GTPasas se comportaban como mutantes nulas [78]. Nuestro alineamiento entre secuencias de EFL1 mostró la presencia de residuos invariables dentro de los últimos 29 residuos del C-terminal, que corresponden a la R1086, K1089, G1090, K1103 y Q1104 (Figura 13). Mientras que los aminoácidos K1103 y Q1104 son exclusivos de las proteínas EFL1, los residuos R1086, K1089 y G1090, se conservan en las proteínas eEF-2 y aEF-2, lo que además, indica que su papel, cualquiera que éste sea, se encuentra evolutivamente conservado. Estudios de criomicroscopía electrónica con ScEf-2 han revelado que el C-terminal de esta GTPasa está implicado en su interacción con el ribosoma [97], por lo que es posible sugerir que esta parte conservada podría presentar el mismo papel en EFL1.

El alineamiento entre proteínas EFL1, a pesar de que globalmente presentan un porcentaje de identidad cercano al 30%, mostró un alto grado de conservación entre sus secuencias, sugiriendo que su plegamiento podría ser similar. El análisis del alineamiento con la herramienta de predicción de estructura secundaria PSIPRED, y la posterior extrapolación de las regiones de estructura secundaria conservadas al modelo tridimensional de la proteína ScEfl1 obtenido por homología usando las coordenadas de ScEf-2, mostró un alto grado de coincidencia para los dominios G, III, IV y V (Figura 15 y Figura 16), por lo que es probable que el plegamiento entre proteínas EFL1 se encuentre conservado a pesar de su divergencia.

Otra característica que resulta de gran interés, es la ausencia de inserciones, la presencia de un gran número de residuos invariables y el alto grado de conservación en el dominio III de las proteínas EFL1. Para fines prácticos, es posible proponer que la estructura de sus homólogos EF-2 y EF-G, está organizada en un módulo catalítico implicado en la hidrólisis del GTP y otro efector, cuyo cambio conformacional, producto de la ruptura de la molécula de alta energía, promueve el movimiento del ribosoma a lo largo del mRNA durante la síntesis de proteínas [99]. En las estructuras tridimensionales de ScEf-2 y TtEF-G, el dominio III presenta la misma conformación, y se localiza justo entre el módulo catalítico y la parte efectora de la proteína. De acuerdo con su mecanismo de acción, es posible que la localización y plegamiento de este dominio otorgue la flexibilidad suficiente a la proteína, para permitir que la energía obtenida durante la hidrólisis de GTP en el dominio G, sea acoplada a un cambio de conformación en los dominios IV y V de la translocasa, de manera tal, que al final promueva el rearreglo estructural del ribosoma durante el evento de translocación [99]. Debido a que existe la posibilidad de que las proteínas EFL1 presenten el mismo plegamiento que las translocasas ribosomales, es posible que el dominio III también ocupe la misma función en esta familia de proteínas. Lo anterior podría apoyar la hipótesis de que las translocasas y las proteínas EFL1, a pesar de participar en procesos completamente diferentes, llevan a cabo su función mediante un mecanismo de acción similar [95,96].

Debido a que los miembros de EFL1 se encuentran altamente conservados en el dominio Eukarya, es posible sugerir que su función se ha mantenido evolutivamente. Además, la presencia de ortólogos de Sdo1/SBDS y Tif6/eIF6 en estos organismos, permite suponer que la vía de maduración ribosomal en la que están implicados estos factores accesorios, está altamente conservada en este dominio de la vida. Para evaluar la conservación funcional de la familia EFL1, decidimos llevar a cabo ensayos de complementación con las proteínas SpEfl1, HsEFL1-1, HsEFL1-2 (con el fin de evaluar su posible rol en la maduración ribosomal) y MmEfl1 en células de levadura *efl1Δ*. Estos experimentos mostraron que ninguna de las proteínas ensayadas rescataban el defecto de crecimiento producido por la pérdida de la proteína endógena y por tanto, sugerían que a pesar de su relación evolutiva posiblemente no presentaban la misma función [6] o que son específicas de especie. [144]. Sandler *et al.* [41] sugieren que algunas proteínas optimizan sus superficies de contacto como resultado de un fenómeno de coevolución con sus compañeros de interacción. En consecuencia, la asociación entre estas proteínas sólo puede ocurrir de manera especie-específica y por lo tanto, no son capaces de funcionar en ensayos de complementación entre especies a menos que el resto de sus compañeros de interacción también estén presentes. Boockock *et al.* realizaron ensayos de complementación en levadura empleando diversos ortólogos de SBDS. Al igual que en nuestros experimentos, las únicas proteínas que rescataron el defecto de crecimiento provocado por la pérdida de *SDO1*, fueron SpSdo1 y las dos quimeras de ScSdo1/HsSBDS. Además, estos investigadores revelaron que los dominios 1 y 3, eran intercambiables entre especies, mientras que el dominio 2 otorgaba a estas proteínas especie especificidad [11]. Resultados por parte de nuestro laboratorio demostraron a partir de ensayos de doble híbrido en levadura, que el dominio 2 de ScSdo1 estaba implicado en la interacción con el dominio II de ScEfl1 [145]. Esto también se observó *in vitro* [87]. La especificidad de especie para este dominio en las proteínas SBDS, y las diferencias en composición observadas entre ortólogos de EFL1, apoyan fuertemente la hipótesis de que ambas proteínas han co-evolucionado y que su reconocimiento es específico de especie. En consecuencia, mediante el uso de ensayos de doble complementación en células de levadura

efl1Δ sdo1Δ decidimos demostrar que las familias EFL1 y SBDS han evolucionado a la par. Nuevamente, de todos los pares de proteínas ensayados, sólo el correspondiente a *S. cerevisiae* y ScEfl1 con proteínas quiméricas de ScSdo1/HsSBDS fueron capaces de reestablecer el crecimiento normal de las células de levadura mutantes (Figura 32 y Figura 33). Los ensayos con las proteínas quiméricas exponen información sumamente interesante. Por un lado, muestran claramente que la interacción entre ScEfl1 y ScSdo1 ocurre entre más de un sitio, siendo el dominio 2 del intercambiador de nucleótidos, la región de reconocimiento principal. Por otra parte, el hecho de que estas proteínas sólo complementen la función de los genes eliminados en levadura en presencia de ScEfl1 indica que hay un fenómeno de co-evolución entre ellas. Sin embargo, este comportamiento no se observa con el par de proteínas de origen murino, los pares de proteínas humanos, e inclusive con el par de *S. pombe* que es el organismo evolutivamente más cercano a *S. cerevisiae* evaluado en nuestra investigación. Lo anterior, probablemente, se debe a la dependencia que el organismo experimental muestra por la presencia de ScEfl1 quien acoplada a ScSdo1 o a una proteína quimérica ScSdo1/HsSBDS, es la única que puede reconocer un tercer componente molecular exclusivo de levadura, que ninguno de los otros complejos proteicos heterólogos ensayados es capaz de hacer. Esto sugiere que las proteínas EFL1, además de haber coevolucionado con las proteínas SBDS, lo han hecho con un componente adicional en la subunidad ribosomal 60S que sólo la GTPasa es capaz de reconocer específicamente.

Estudios *in vitro* han revelado que la activación de la hidrólisis de GTP en HsEFL1 es mediada por su interacción con la subunidad 60S [131]. Este mismo comportamiento ha sido observado también para EF-2 y EF-G, y se ha demostrado que el tallo ribosomal es el componente proteico implicado específicamente en la activación de estas translocasas [146–148]. En bacterias, esta estructura está compuesta por la proteína L10 y dos o tres copias de homodímeros de L12 [149,150]. En el caso de las arqueas, el tallo consiste en una proteína P0 unida a tres homodímeros de P1 [151,152]. Por último, en eucariontes, este componente ribosomal está formado por una proteína P0 unida a

dos copias de heterodímeros P1/P2 [153,154]. A pesar de que algunas de estas proteínas comparten características estructurales y funcionales entre sí, hay reportes que indican que los componentes individuales de estos complejos proteicos no son funcionalmente intercambiables entre miembros de los diferentes dominios de la vida [148,155,156].

La vía de maduración en donde participan las proteínas EFL1 está relacionada con el ensamblaje del tallo ribosomal. Se ha observado que los defectos sobre la proteína P0 en *S. cerevisiae* impiden que ScEfl1 y ScSdo1 liberen a Tif6 de la subunidad mayor. Esto ha llevado a suponer que la función de estos factores accesorios es comprobar que el tallo ribosomal se ha ensamblado adecuadamente [69,95]. Por otro lado, a pesar de que ScEf-2 se une al mismo sitio que ScEfl1 sobre el ribosoma [136], esta translocasa no es capaz de reemplazar la función de la proteína de maduración ribosomal [78], sugiriendo que estas GTPasas reconocen regiones específicas en su sitio de unión. En conjunto, lo anterior apoya nuestra hipótesis de que las proteínas EFL1 funcionan de manera especie-específica, y que coevolucionaron con un componente ribosomal que podría ser alguna proteína del tallo ribosomal.

Finalmente, como se ha mencionado a lo largo de este trabajo, las proteínas Sdo1/SBDS y Tif6/eIF6 presentan ortólogos en Archaea [11,117]. *Benelli et al.* demostraron que el ortólogo de Tif6/eIF6 en arquea, aIF6, también se une a la subunidad mayor (en este caso 50S) y se comporta como un factor de antiasociación en estos organismos. Sin embargo, aún no se ha identificado el mecanismo por el cual esta proteína es liberada de la subunidad 50S en arqueas. En nuestra búsqueda de miembros de EFL1 no identificamos ortólogos en estos procarotes, sin embargo, la relación filogenética entre estas proteínas de maduración ribosomal y aEF-2, sugiere que, además de su función en el proceso de síntesis de proteínas [148], estas últimas podrían representar el mismo papel que EFL1 en Archaea. Para demostrarlo, decidimos ensayar la complementación funcional de AfEf-2 y AfSbds, de manera individual y en conjunto. Al igual que para los miembros de EFL1 ensayados, las proteínas de *A. fulgidus* tampoco fueron capaces de rescatar el defecto de crecimiento en levadura generado por la pérdida

de los genes *SDO1* y *EFL1*. Sin embargo, la posibilidad de que estas proteínas desempeñen el mismo papel que su contraparte en eucariontes, aún no puede descartarse por completo debido a que es probable que el resultado negativo obtenido para estas proteínas en los ensayos de complementación, haya sido consecuencia del probable fenómeno de coevolución que circunda el reconocimiento entre las proteínas EFL1/SBDS y su sitio de unión sobre el ribosoma.

8 Conclusiones

- La familia de proteínas EFL1 es exclusiva del dominio Eukarya y presenta miembros en todos los subgrupos de los Phylum que conforman este dominio.
- El alineamiento múltiple entre secuencias de miembros de EFL1, muestra que los dominios G, III y V son las regiones de estas proteínas con la mayor cantidad de residuos conservados y sugiere que su plegamiento está conservado evolutivamente.
- El análisis filogenético entre miembros de las GTPasas EFL1, eEF-2, aEF-2 y EF-G muestra que todas estas proteínas presentan un origen evolutivo en común.
- La relación evolutiva entre las proteínas EFL1 y eEF-2 sugiere que ambas presentan un plegamiento similar. Además, el análisis de estructura secundaria acoplado al modelo tridimensional de ScEfl1 obtenido por homología, sugiere que la estructura terciaria de los miembros de la familia de GTPasas implicadas en el proceso de maduración ribosomal, también podría conservarse de manera evolutiva.
- Las proteínas SpEfl1, HsEFL1-1, HsEFL1-2, MmEFL1 y AfEF-2 no complementan la función de la proteína ScEfl1 en células de *S. cerevisiae efl1Δ*.
- Los pares de proteínas EFL1/SBDS de *H. sapiens*, *M. musculus*, *S. pombe* y EF-2/SBDS de *A. fulgidus* no son capaces de complementar la función de ScEfl1 y ScSdo1 en células de *S. cerevisiae efl1Δ/sdo1Δ*.
- Los ensayos de complementación empleando proteínas quiméricas de SBDS indican que existe co-evolución entre estas proteínas y la familia de GTPasas EFL1.
- Los ensayos de complementación empleando proteínas quiméricas de SBDS sugieren que hay más de un sitio de interacción entre estas proteínas y las GTPasas EFL1, y muestran claramente, que el sitio de interacción principal recae en el dominio 2 del intercambiador de nucleótidos.

9 Perspectivas

- Construir la secuencia codificante de la proteína quimérica EFL1 humana que contenga los dominios 1 y 2 de la proteína Efl1 de *S. cerevisiae* y clonarla en el vector de expresión pAM1.
- Construir la secuencia codificante de la proteína quimérica EFL1 de *S. cerevisiae* que contenga los dominios 1 y 2 de la proteína EFL1 humana y clonarla en el vector de expresión pAM1.
- Evaluar la complementación genética de las proteínas EFL1 quiméricas para demostrar que su función es especie específica y mapear los sitios implicados en el reconocimiento de estas GTPasas por el ribosoma.
- Evaluar la complementación genética de las proteínas EFL1 quiméricas con sus correspondientes proteínas SBDS nativas y las proteínas quiméricas obtenidas en este trabajo, para demostrar que su reconocimiento es especie-específico y ocurre entre el dominio 2 de ambas proteínas.
- Realizar ensayos de actividad de las proteínas EFL1 humana y de levadura, empleando subunidades ribosomales 60S, para demostrar que su activación por esta partícula es especie específica.
- Realizar ensayos de actividad de las proteínas EFL1 humana y de levadura, empleando sus correspondientes proteínas SBDS, para demostrar que la modulación de su actividad es un fenómeno especie específico.
- Evaluar la interacción mediante ensayos de doble híbrido en levadura, entre la proteína Efl1 de *S. cerevisiae* y las proteínas del tallo ribosomal, para identificar si alguno de estos componentes está implicado en la activación de GTPasa.
- Obtener las proteínas aEF2 y Sbds de *A. fulgidus* de manera recombinante, y evaluar *in vitro* si la proteína Sbds actúa como un intercambiador de nucleótidos para la GTPasa, de manera análoga a lo que ocurre con las proteínas EFL1 y SBDS de eucariontes.

10 Bibliografía

- [1] J. Alföldi, K. Lindblad-Toh, Comparative genomics as a tool to understand evolution and disease, *Genome Res.* 23 (2013) 1063–1068. doi:10.1101/gr.157503.113.
- [2] G.M. Rubin, Comparative Genomics of the Eukaryotes, *Science* (80-.). 287 (2000) 2204–2215. doi:10.1126/science.287.5461.2204.
- [3] A. Osterman, R. Overbeek, Missing genes in metabolic pathways: A comparative genomics approach, *Curr. Opin. Chem. Biol.* 7 (2003) 238–251. doi:10.1016/S1367-5931(03)00027-9.
- [4] R.T. van der Heijden, B. Snel, V. van Noort, M.A. Huynen, Orthology prediction at scalable resolution by phylogenetic tree analysis, *BMC Bioinformatics.* 8 (2007) 83. doi:10.1186/1471-2105-8-83.
- [5] S. Heinicke, M.S. Livstone, C. Lu, R. Oughtred, F. Kang, S. V. Angiuoli, O. White, D. Botstein, K. Dolinski, The Princeton Protein Orthology Database (P-POD): A comparative genomics analysis tool for biologists, *PLoS One.* 2 (2007). doi:10.1371/journal.pone.0000766.
- [6] T. Gabaldón, E. V Koonin, Functional and evolutionary implications of gene orthology, *Nat. Rev. Genet.* 14 (2013) 360–366. doi:10.1038/nrg3456.
- [7] F.S.L. Brinkman, D.D. Leipe, *Phylogenetic Analysis*, 2001. doi:10.1002/0471223921.ch14.
- [8] L.A. Mirny, M.S. Gelfand, Using orthologous and paralogous proteins to identify specificity-determining residues in bacterial transcription factors, *J. Mol. Biol.* 321 (2002) 7–20. doi:10.1016/S0022-2836(02)00587-9.
- [9] L. Sanchez-Pulido, C.P. Ponting, Cdc45: The missing recj ortholog in eukaryotes?, *Bioinformatics.* 27 (2011) 1885–1888. doi:10.1093/bioinformatics/btr332.
- [10] F. Blombach, K.S. Makarova, J. Marrero, B. Siebers, E. V Koonin, J. van der Oost, Identification of an ortholog of the eukaryotic RNA polymerase III subunit RPC34 in Crenarchaeota and Thaumarchaeota suggests specialization of RNA polymerases for coding and non-coding RNAs in Archaea, *Biol Direct.* 4 (2009) 39. doi:10.1186/1745-6150-4-39.
- [11] G.R.B. Boocock, M.R. Marit, J.M. Rommens, Phylogeny, sequence conservation, and functional complementation of the SBDS protein family., *Genomics.* 87 (2006) 758–771. doi:10.1016/j.ygeno.2006.01.010.
- [12] I. V. Gregoret, Y.M. Lee, H. V. Goodson, Molecular evolution of the histone deacetylase family: Functional implications of phylogenetic analysis, *J. Mol. Biol.* 338 (2004) 17–31. doi:10.1016/j.jmb.2004.02.006.
- [13] L.C. Hileman, J.F. Sundstrom, A. Litt, M. Chen, T. Shumba, V.F. Irish, Molecular and phylogenetic analyses of the MADS-box gene family in tomato, *Mol. Biol. Evol.* 23 (2006) 2245–2258. doi:10.1093/molbev/msl095.
- [14] G. Gutiérrez, M.D. Ganfornina, D. Sánchez, Evolution of the lipocalin family as inferred from a protein sequence phylogeny, *Biochim. Biophys. Acta - Protein Struct. Mol. Enzymol.* 1482 (2000) 35–45. doi:10.1016/S0167-4838(00)00151-5.
- [15] F. Burki, The eukaryotic tree of life from a global phylogenomic perspective,

- Cold Spring Harb. Perspect. Biol. 6 (2014).
doi:10.1101/cshperspect.a016147.
- [16] T.R. Gregory, Understanding Evolutionary Trees, *Evol. Educ. Outreach.* 1 (2008) 121–137. doi:10.1007/s12052-008-0035-x.
- [17] S.F. Altschul, Substitution Matrices, *Encycl. Life Sci.* (2013) 1–4. doi:10.1002/9780470015902.a0005265.pub3.
- [18] F. Abascal, R. Zardoya, D. Posada, ProtTest : Selection of best-fit models of protein evolution What can I use ProtTest for ? – Introduction The program : using ProtTest, *Bioinformatics.* 21 (2005) 1–17. doi:10.1093/bioinformatics/bti263.
- [19] W.-H. Li, W.H. Li, Molecular evolution, *Mol. Evol.* 15 (1997) xv+487p. <Go to ISI>://BCI:BCI199799668192.
- [20] J. Felsenstein, An Alternating Least Squares Approach to Inferring Phylogenies from Pairwise Distances, *Syst. Biol.* 46 (1997) 101. doi:10.2307/2413638.
- [21] S. Kumar, G. Stecher, K. Tamura, MEGA7: Molecular Evolutionary Genetics Analysis version 7.0 for bigger datasets., *Mol. Biol. Evol.* (2016) msw054. doi:10.1093/molbev/msw054.
- [22] J.C. Wilgenbusch, D. Swofford, Inferring Evolutionary Trees with PAUP*, in: *Curr. Protoc. Bioinforma.*, 2003: p. Unit 6.4. doi:10.1002/0471250953.bi0604s00.
- [23] D.L. Swofford, G.J. Olsen, P.J. Waddell, D.M. Hillis, Phylogenetic Inference, *Mol. Syst.* (1996) 407–514. doi:10.1093/sysbio/syr131.
- [24] C.A. Buerkle, Maximum-likelihood estimation of a hybrid index based on molecular markers, *Mol. Ecol. Notes.* 5 (2005) 684–687. doi:10.1111/j.1471-8286.2005.01011.x.
- [25] J. Tuimala, A primer to phylogenetic analysis using the PHYLIP package, 2006. <http://www.life.umd.edu/labs/delwiche/bsci348s/lab/phylip2.pdf>.
- [26] B. Alix, D.A. Boubacar, M. Vladimir, T-REX: A web server for inferring, validating and visualizing phylogenetic trees and networks, *Nucleic Acids Res.* 40 (2012). doi:10.1093/nar/gks485.
- [27] J. Felsenstein, Confidence limits on phylogenies: an approach using the bootstrap, *Evolution (N. Y).* 39 (1985) 783–791. doi:10.2307/2408678.
- [28] S. Holmes, Bootstrapping Phylogenetic Trees: Theory and Methods, *Stat. Sci.* 18 (2003) 241–255. doi:10.1214/ss/1063994979.
- [29] G. Moreno-Hagelsieb, K. Latimer, Choosing BLAST options for better detection of orthologs as reciprocal best hits, *Bioinformatics.* 24 (2008) 319–324. doi:10.1093/bioinformatics/btm585.
- [30] N. Ward, G. Moreno-Hagelsieb, Quickly finding orthologs as reciprocal best hits with BLAT, LAST, and UBLAST: How much do we miss?, *PLoS One.* 9 (2014). doi:10.1371/journal.pone.0101850.
- [31] T. Johnson, Reciprocal best hits are not a logically sufficient condition for orthology, *Most.* 5 (2007) 1–8. <http://arxiv.org/abs/0706.0117>.
- [32] R.T. van der Heijden, B. Snel, V. van Noort, M.A. Huynen, Orthology prediction at scalable resolution by phylogenetic tree analysis, *BMC Bioinformatics.* 8 (2007) 83. doi:10.1186/1471-2105-8-83.
- [33] S. Fischer, B.P. Brunk, F. Chen, X. Gao, O.S. Harb, J.B. Iodice, D. Shanmugam, D.S. Roos, C.J. Stoeckert, Using OrthoMCL to assign proteins

- to OrthoMCL-DB groups or to cluster proteomes into new ortholog groups, *Curr. Protoc. Bioinforma.* (2011). doi:10.1002/0471250953.bi0612s35.
- [34] M. Remm, C.E. V Storm, E.L.L. Sonnhammer, Automatic clustering of orthologs and in-paralogs from pairwise species comparisons., *J. Mol. Biol.* 314 (2001) 1041–52. doi:10.1006/jmbi.2000.5197.
- [35] G. Östlund, T. Schmitt, K. Forslund, T. Köstler, D.N. Messina, S. Roopra, O. Frings, E.L.L. Sonnhammer, Inparanoid 7: New algorithms and tools for eukaryotic orthology analysis, *Nucleic Acids Res.* 38 (2009). doi:10.1093/nar/gkp931.
- [36] F. Chen, A.J. Mackey, C.J. Stoeckert, D.S. Roos, OrthoMCL-DB: querying a comprehensive multi-species collection of ortholog groups., *Nucleic Acids Res.* 34 (2006) D363-8. doi:10.1093/nar/gkj123.
- [37] K. Dolinski, D. Botstein, Orthology and functional conservation in eukaryotes., 2007. doi:10.1146/annurev.genet.40.110405.090439.
- [38] J.E. Kranz, C. Holm, Cloning by function: an alternative approach for identifying yeast homologs of genes from other organisms., *Proc. Natl. Acad. Sci. U. S. A.* 87 (1990) 6629–6633. doi:10.1073/pnas.87.17.6629.
- [39] S.L. Forsburg, The art and design of genetic screens: yeast., *Nat. Rev. Genet.* 2 (2001) 659–668. doi:10.1038/35088500.
- [40] A.A. Duina, M.E. Miller, J.B. Keeney, Budding yeast for budding geneticists: A primer on the *Saccharomyces cerevisiae* model system, *Genetics.* 197 (2014) 33–48. doi:10.1534/genetics.114.163188.
- [41] I. Sandler, M. Abu-Qarn, A. Aharoni, Protein co-evolution: How do we combine bioinformatics and experimental approaches?, *Mol. Biosyst.* (2012) 175–181. doi:10.1039/c2mb25317h.
- [42] J.T. Pronk, Auxotrophic yeast strains in fundamental and applied research, *Appl. Environ. Microbiol.* 68 (2002) 2095–2100. doi:10.1128/AEM.68.5.2095-2100.2002.
- [43] A.H.Y. Tong, C. Boone, Synthetic genetic array analysis in *Saccharomyces cerevisiae*., *Methods Mol. Biol.* 313 (2006) 171–92. doi:10.1016/S0076-6879(10)70007-0.
- [44] N. Gunge, Yeast DNA plasmids, *Annu. Rev. Microbiol.* 37 (1983) 253–276. doi:doi:10.1146/annurev.mi.37.100183.001345.
- [45] D.L. Ludwig, C. V. Bruschi, The 2-??m plasmid as a nonselectable, stable, high copy number yeast vector, *Plasmid.* 25 (1991) 81–95. doi:10.1016/0147-619X(91)90019-S.
- [46] T. Stearns, H. Ma, D. Botstein, Manipulating yeast genome using plasmid vectors, *Methods Enzymol.* 185 (1990) 280–297. doi:10.1016/0076-6879(90)85025-J.
- [47] Z. Wang, L. Jin, Z. Yuan, G. Wegrzyn, A. Wegrzyn, Classification of plasmid vectors using replication origin, selection marker and promoter as criteria, *Plasmid.* 61 (2009) 47–51. doi:10.1016/j.plasmid.2008.09.003.
- [48] A. Narla, B.L. Ebert, Ribosomopathies: Human disorders of ribosome dysfunction, *Blood.* 115 (2010) 3196–3205. doi:10.1182/blood-2009-10-178129.
- [49] H. Nakhoul, J. Ke, X. Zhou, W. Liao, S.X. Zeng, H. Lu, Ribosomopathies : mechanisms of disease, *Clin. Med. Insights Blood Disord.* 7 (2014) 7–16. doi:10.4137/CMBD.S16952.RECEIVED.

- [50] T. Teng, G. Thomas, C.A. Mercer, Growth control and ribosomopathies, *Curr. Opin. Genet. Dev.* 23 (2013) 63–71. doi:10.1016/j.gde.2013.02.001.
- [51] J. Armistead, B. Triggs-Raine, Diverse diseases from a ubiquitous process: The ribosomopathy paradox, *FEBS Lett.* 588 (2014) 1491–1500. doi:10.1016/j.febslet.2014.03.024.
- [52] A. Khanna-Gupta, Bone Marrow Failure Syndromes: The Ribosomopathies, NIH Public Access, *J. Bone Marrow Res.* 1 (2013) 1–5. doi:10.4172/jbmr.1000106.Bone.
- [53] V.G. Panse, A.W. Johnson, Maturation of eukaryotic ribosomes: acquisition of functionality, *Trends Biochem. Sci.* 35 (2010) 260–266. doi:10.1016/j.tibs.2010.01.001.
- [54] J.L. Woolford, S.J. Baserga, Ribosome biogenesis in the yeast *Saccharomyces cerevisiae*, *Genetics.* 195 (2013) 643–681. doi:10.1534/genetics.113.153197.
- [55] S.L. French, Y.N. Osheim, F. Cioci, M. Nomura, A.L. Beyer, In exponentially growing *Saccharomyces cerevisiae* cells, rRNA synthesis is determined by the summed RNA polymerase I loading rate rather than by the number of active genes., *Mol. Cell. Biol.* 23 (2003) 1558–1568. doi:10.1128/MCB.23.5.1558-1568.2003.
- [56] M. Koš, D. Tollervey, Yeast Pre-rRNA Processing and Modification Occur Cotranscriptionally, *Mol. Cell.* 37 (2010) 809–820. doi:10.1016/j.molcel.2010.02.024.
- [57] W.A. Decatur, M.J. Fournier, rRNA modifications and ribosome function, *Trends Biochem. Sci.* 27 (2002) 344–351. doi:10.1016/S0968-0004(02)02109-6.
- [58] J.L. Baxter-Roshek, A.N. Petrov, J.D. Dinman, Optimization of ribosome structure and function by rRNA base modification, *PLoS One.* 2 (2007). doi:10.1371/journal.pone.0000174.
- [59] X. h. Liang, Q. Liu, M.J. Fournier, rRNA Modifications in an Intersubunit Bridge of the Ribosome Strongly Affect Both Ribosome Biogenesis and Activity, *Mol. Cell.* 28 (2007) 965–977. doi:10.1016/j.molcel.2007.10.012.
- [60] K. Jack, C. Bellodi, D.M. Landry, R.O. Niederer, A. Meskauskas, S. Musalgaonkar, N. Kopmar, O. Krasnykh, A.M. Dean, S.R. Thompson, D. Ruggero, J.D. Dinman, RRNA Pseudouridylation Defects Affect Ribosomal Ligand Binding and Translational Fidelity from Yeast to Human Cells, *Mol. Cell.* 44 (2011) 660–666. doi:10.1016/j.molcel.2011.09.017.
- [61] J.R. Warner, The economics of ribosome biosynthesis in yeast, *Trends Biochem. Sci.* 24 (1999) 437–440. doi:10.1016/S0968-0004(99)01460-7.
- [62] J. Trapman, J. Retèl, R.J. Planta, Ribosomal precursor particles from yeast., *Exp. Cell Res.* 90 (1975) 95–104. doi:10.1016/0014-4827(75)90361-4.
- [63] T.A. Nissan, J. Baßler, E. Petfalski, D. Tollervey, E. Hurt, 60S pre-ribosome formation viewed from assembly in the nucleolus until export to the cytoplasm, *EMBO J.* 21 (2002) 5539–5547. doi:10.1093/emboj/cdf547.
- [64] A. Fatica, A.D. Cronshaw, M. Dlaki, D. Tollervey, Ssf1p prevents premature processing of an early pre-60S ribosomal particle, *Mol. Cell.* 9 (2002) 341–351. doi:10.1016/S1097-2765(02)00458-6.
- [65] P. Harnpicharnchai, J. Jakovljevic, E. Horsey, T. Miles, J. Roman, M. Rout, D. Meagher, B. Imai, Y. Guo, C.J. Brame, J. Shabanowitz, D.F. Hunt, J.L.

- Woolford, Composition and functional characterization of yeast 66S ribosome assembly intermediates, *Mol. Cell.* 8 (2001) 505–515. doi:10.1016/S1097-2765(01)00344-6.
- [66] C. Saveanu, D. Bienvenu, A. Namane, P.E. Gleizes, N. Gas, A. Jacquier, M. Fromont-Racine, Nog2p, a putative GTPase associated with pre-60S subunits and required for late 60S maturation steps, *EMBO J.* 20 (2001) 6475–6484. doi:10.1093/emboj/20.22.6475.
- [67] A.K. Henras, J. Soudet, M. G erus, S. Lebaron, M. Caizergues-Ferrer, A. Mougin, Y. Henry, The post-transcriptional steps of eukaryotic ribosome biogenesis, *Cell. Mol. Life Sci.* 65 (2008) 2334–2359. doi:10.1007/s00018-008-8027-0.
- [68] P. Nerurkar, M. Altvater, S. Gerhardy, S. Sch utz, U. Fischer, C. Weirich, V.G. Panse, Eukaryotic Ribosome Assembly and Nuclear Export., *Int. Rev. Cell Mol. Biol.* 319 (2015) 107–40. doi:10.1016/bs.ircmb.2015.07.002.
- [69] K.Y. Lo, Z. Li, C. Bussiere, S. Bresson, E.M. Marcotte, A.W. Johnson, Defining the pathway of cytoplasmic maturation of the 60S ribosomal subunit, *Mol. Cell.* 39 (2010) 196–208. doi:10.1016/j.molcel.2010.06.018.
- [70] H. Tschochner, E. Hurt, Pre-ribosomes on the road from the nucleolus to the cytoplasm, *Trends Cell Biol.* 13 (2003) 255–263. doi:10.1016/S0962-8924(03)00054-0.
- [71] E. Hurt, S. Hannus, B. Schmelzl, D. Lau, D. Tollervey, G. Simos, A novel in vivo assay reveals inhibition of ribosomal nuclear export in Ran-cycle and nucleoporin mutants, *J. Cell Biol.* 144 (1999) 389–401. doi:10.1083/jcb.144.3.389.
- [72] J. de la Cruz, D. Kressler, M. Rojo, D. Tollervey, P. Linder, Spb4p, an essential putative RNA helicase, is required for a late step in the assembly of 60S ribosomal subunits in *Saccharomyces cerevisiae.*, *RNA.* 4 (1998) 1268–1281. doi:10.1017/S1355838298981158.
- [73]  . Vaň cov a, J. Wolf, G. Martin, D. Blank, S. Dettwiler, A. Friedlein, H. Langen, G. Keith, W. Keller, A new yeast poly(A) polymerase complex involved in RNA quality control, *PLoS Biol.* 3 (2005) 0986–0997. doi:10.1371/journal.pbio.0030189.
- [74] J. LaCava, J. Houseley, C. Saveanu, E. Petfalski, E. Thompson, A. Jacquier, D. Tollervey, RNA degradation by the exosome is promoted by a nuclear polyadenylation complex, *Cell.* 121 (2005) 713–724. doi:10.1016/j.cell.2005.04.029.
- [75] M. Rodr guez-Mateos, J.J. Garc a-G omez, R. Francisco-Velilla, M. Remacha, J. De la Cruz, J.P.G. Ballesta, Role and dynamics of the ribosomal protein P0 and its related trans-acting factor Mrt4 during ribosome assembly in *Saccharomyces cerevisiae*, *Nucleic Acids Res.* 37 (2009) 7519–7532. doi:10.1093/nar/gkp806.
- [76] M. Gartmann, M. Blau, J.P. Armache, T. Mielke, M. Topf, R. Beckmann, Mechanism of eIF6-mediated inhibition of ribosomal subunit joining, *J. Biol. Chem.* 285 (2010) 14848–14851. doi:10.1074/jbc.C109.096057.
- [77] K. Karbstein, Quality control mechanisms during ribosome maturation, *Trends Cell Biol.* 23 (2013) 242–250. doi:10.1016/j.tcb.2013.01.004.
- [78] a. M. B cam, F. Nasr, W. Racki, M. Zagulski, C. Herbert, Ria1p (Ynl163c), a protein similar to elongation factors 2, is involved in the biogenesis of the

- 60S subunit of the ribosome in *Saccharomyces cerevisiae*, *Mol. Genet. Genomics*. 266 (2001) 454–462. doi:10.1007/s004380100548.
- [79] B. Senger, D.L. Lafontaine, J.S. Graindorge, O. Gadal, a Camasses, a Sanni, J.M. Garnier, M. Breitenbach, E. Hurt, F. Fasiolo, The nucle(ol)ar Tif6p and Efl1p are required for a late cytoplasmic step of ribosome synthesis., *Mol. Cell*. 8 (2001) 1363–73. <http://www.ncbi.nlm.nih.gov/pubmed/11779510>.
- [80] T.F. Menne, B. Goyenechea, N. Sánchez-Puig, C.C. Wong, L.M. Tonkin, P.J. Ancliff, R.L. Brost, M. Costanzo, C. Boone, A.J. Warren, The Shwachman-Bodian-Diamond syndrome protein mediates translational activation of ribosomes in yeast., *Nat. Genet.* 39 (2007) 486–495. doi:10.1038/ng1994.
- [81] A. García-Márquez, A. Gijssbers, E. De La Mora, N. Sánchez-Puig, Defective guanine nucleotide exchange in the elongation factor-like 1 (EFL1) GTPase by mutations in the Shwachman-Diamond syndrome protein, *J. Biol. Chem.* 290 (2015) 17669–17678. doi:10.1074/jbc.M114.626275.
- [82] T.F. Menne, B. Goyenechea, N. Sánchez-Puig, C.C. Wong, L.M. Tonkin, P.J. Ancliff, R.L. Brost, M. Costanzo, C. Boone, A.J. Warren, The Shwachman-Bodian-Diamond syndrome protein mediates translational activation of ribosomes in yeast., *Nat. Genet.* 39 (2007) 486–95. doi:10.1038/ng1994.
- [83] C. Shamas, T.F. Menne, C. Hilcenko, S.R. Michell, B. Goyenechea, G.R.B. Boocock, P.R. Durie, J.M. Rommens, A.J. Warren, Structural and mutational analysis of the SBDS protein family: Insight into the leukemia-associated Shwachman-Diamond Syndrome, *J. Biol. Chem.* 280 (2005) 19221–19229. doi:10.1074/jbc.M414656200.
- [84] A. Savchenko, N. Krogan, J.R. Cort, E. Evdokimova, J.M. Lew, A. a Yee, L. Sánchez-Pulido, M. a Andrade, A. Bochkarev, J.D. Watson, M. a Kennedy, J. Greenblatt, T. Hughes, C.H. Arrowsmith, J.M. Rommens, A.M. Edwards, The Shwachman-Bodian-Diamond syndrome protein family is involved in RNA metabolism., *J. Biol. Chem.* 280 (2005) 19213–20. doi:10.1074/jbc.M414421200.
- [85] J.F. de Oliveira, M.L. Sforça, T.M.A. Blumenschein, M.B. Goldfeder, B.G. Guimarães, C.C. Oliveira, N.I.T. Zanchin, A.C. Zeri, Structure, dynamics, and rna interaction analysis of the human sbds protein, *J. Mol. Biol.* 396 (2010) 1053–1069. doi:10.1016/j.jmb.2009.12.039.
- [86] C. Ma, K. Yan, D. Tan, N. Li, Y. Zhang, Y. Yuan, Z. Li, M.-Q. Dong, J. Lei, N. Gao, Structural dynamics of the yeast Shwachman-Diamond syndrome protein (Sdo1) on the ribosome and its implication in the 60S subunit maturation., *Protein Cell*. 7 (2016) 187–200. doi:10.1007/s13238-015-0242-5.
- [87] N. Asano, H. Atsumi, A. Nakamura, Y. Tanaka, I. Tanaka, M. Yao, Direct interaction between EFL1 and SBDS is mediated by an intrinsically disordered insertion domain., *Biochem. Biophys. Res. Commun.* 443 (2014) 1251–6. doi:10.1016/j.bbrc.2013.12.143.
- [88] S. Zhang, M. Shi, C. Hui, J.M. Rommens, Loss of the mouse ortholog of the shwachman-diamond syndrome gene (Sbds) results in early embryonic lethality., *Mol. Cell. Biol.* 26 (2006) 6656–63. doi:10.1128/MCB.00091-06.
- [89] G.R.B. Boocock, J. a Morrison, M. Popovic, N. Richards, L. Ellis, P.R. Durie, J.M. Rommens, Mutations in SBDS are associated with Shwachman-

- Diamond syndrome., *Nat. Genet.* 33 (2003) 97–101. doi:10.1038/ng1062.
- [90] E. Costa, R. Santos, Hematologically important mutations: Shwachman-Diamond syndrome, *Blood Cells, Mol. Dis.* 40 (2008) 183–184. doi:10.1016/j.bcmed.2007.07.008.
- [91] Kathleen L. McCann and Susan J. Baserga, M. Ribosomopathies, *NIH Public Access*, 341 (2014) 849–850. doi:10.1126/science.1244156.Mysterious.
- [92] A. Savchenko, N. Krogatt, J.R. Cort, E. Evdokimova, J.M. Lew, A.A. Yee, L. Sánchez-Pulido, M.A. Andrade, A. Bochkarev, J.D. Watson, M.A. Kennedy, J. Greenblatt, T. Hughes, C.H. Arrowsmith, J.M. Rommens, A.M. Edwards, The Shwachman-Bodian-Diamond syndrome protein family is involved in RNA metabolism, *J. Biol. Chem.* 280 (2005) 19213–19220. doi:10.1074/jbc.M414421200.
- [93] F. Weis, E. Giudice, M. Churcher, L. Jin, C. Hilcenko, C.C. Wong, D. Traynor, R.R. Kay, A.J. Warren, Mechanism of eIF6 release from the nascent 60S ribosomal subunit., *Nat. Struct. Mol. Biol.* 22 (2015) 914–919. doi:10.1038/nsmb.3112.
- [94] H.L. Ball, B. Zhang, J.J. Riches, R. Gandhi, J. Li, J.M. Rommens, J.S. Myers, Shwachman-Bodian Diamond syndrome is a multi-functional protein implicated in cellular stress responses, *Hum. Mol. Genet.* 18 (2009) 3684–3695. doi:10.1093/hmg/ddp316.
- [95] C. Bussiere, Y. Hashem, S. Arora, J. Frank, A.W. Johnson, Integrity of the P-site is probed during maturation of the 60S ribosomal subunit., *J. Cell Biol.* 197 (2012) 747–59. doi:10.1083/jcb.201112131.
- [96] J.S. Graindorge, J.C. Rousselle, B. Senger, P. Lenormand, A. Namane, F. Lacroute, F. Fasiolo, Deletion of EFL1 results in heterogeneity of the 60 S GTPase-associated rRNA conformation, *J. Mol. Biol.* 352 (2005) 355–369. doi:10.1016/j.jmb.2005.07.037.
- [97] C.M. Spahn, M.G. Gomez-Lorenzo, R.A. Grassucci, R. Jorgensen, G.R. Andersen, R. Beckmann, P.A. Penczek, J.P. Ballesta, J. Frank, Domain movements of elongation factor eEF2 and the eukaryotic 80S ribosome facilitate tRNA translocation, *EMBO J.* 23 (2004) 1008–1019. doi:10.1038/sj.emboj.7600102.
- [98] A. Wittinghofer, I.R. Vetter, Structure-function relationships of the G domain, a canonical switch motif., *Annu. Rev. Biochem.* 80 (2011) 943–971. doi:10.1146/annurev-biochem-062708-134043.
- [99] R. Jørgensen, P. a Ortiz, A. Carr-Schmid, P. Nissen, T.G. Kinzy, G.R. Andersen, Two crystal structures demonstrate large conformational changes in the eukaryotic ribosomal translocase., *Nat. Struct. Biol.* 10 (2003) 379–85. doi:10.1038/nsb923.
- [100] K.W. Kaufmann, G.H. Lemmon, S.L. Deluca, J.H. Sheehan, J. Meiler, Practically useful: What the R osetta protein modeling suite can do for you, *Biochemistry.* 49 (2010) 2987–2998. doi:10.1021/bi902153g.
- [101] I. Ebersberger, S. Simm, M.S. Leisegang, P. Schmitzberger, O. Mirus, A. Von Haeseler, M.T. Bohnsack, E. Schleiff, The evolution of the ribosome biogenesis pathway from a yeast perspective, *Nucleic Acids Res.* 42 (2014) 1509–1523. doi:10.1093/nar/gkt1137.
- [102] Z. Zhang, W.I. Wood, A profile hidden Markov model for signal peptides generated by HMMER, *Bioinformatics.* 19 (2003) 307–308.

- doi:10.1093/bioinformatics/19.2.307.
- [103] R.D. Finn, J. Clements, S.R. Eddy, HMMER web server: Interactive sequence similarity searching, *Nucleic Acids Res.* 39 (2011). doi:10.1093/nar/gkr367.
- [104] R.D. Finn, J. Clements, W. Arndt, B.L. Miller, T.J. Wheeler, F. Schreiber, A. Bateman, S.R. Eddy, HMMER web server: 2015 update., *Nucleic Acids Res.* 43 (2015) W30-8. doi:10.1093/nar/gkv397.
- [105] E. Boutet, D. Lieberherr, M. Tognolli, M. Schneider, a Bairoch, UniProtKB/Swiss-Prot, *Methods Mol Biol.* 406 (2007) 89–112. doi:1-59745-535-0:89 [pii].
- [106] A.G.B. Simpson, A.J. Roger, The real “kingdoms” of eukaryotes., *Curr. Biol.* 14 (2004) R693–R696. doi:10.1016/j.cub.2004.08.038.
- [107] P. Di Tommaso, S. Moretti, I. Xenarios, M. Orobitg, A. Montanyola, J.M. Chang, J.F. Taly, C. Notredame, T-Coffee: A web server for the multiple sequence alignment of protein and RNA sequences using structural information and homology extension, *Nucleic Acids Res.* 39 (2011). doi:10.1093/nar/gkr245.
- [108] V.A. Simossis, J. Heringa, The PRALINE online server: Optimising progressive multiple alignment on the web, *Comput. Biol. Chem.* 27 (2003) 511–519. doi:10.1016/j.compbiolchem.2003.09.002.
- [109] M. Clamp, J. Cuff, S.M. Searle, G.J. Barton, The Jalview Java alignment editor, *Bioinformatics.* 20 (2004) 426–427. doi:10.1093/bioinformatics/btg430.
- [110] B. Webb, A. Sali, Comparative Protein Structure Modeling Using MODELLER., *Curr. Protoc. Bioinformatics.* 47 (2014) 5.6.1-5.6.32. doi:10.1002/0471250953.bi0506s47.
- [111] K. Katoh, D.M. Standley, MAFFT: Iterative refinement and additional methods, *Methods Mol. Biol.* 1079 (2014) 131–146. doi:10.1007/978-1-62703-646-7_8.
- [112] F. Abascal, R. Zardoya, D. Posada, ProtTest: Selection of best-fit models of protein evolution, *Bioinformatics.* 21 (2005) 2104–2105. doi:10.1093/bioinformatics/bti263.
- [113] A. Hin Yan Tong, C. Boone, High-Throughput Strain Construction and Systematic Synthetic Lethal Screening in *Saccharomyces cerevisiae*, *Methods Mol. Biol.* 36 (2005) 1–19. doi:10.1016/S0076-6879(10)70007-0.
- [114] a H. Tong, C. Boone, Synthetic genetic array analysis in *Saccharomyces cerevisiae*, *Methods Mol. Biol. (Clifton, NJ).* 313 (2005) 171–192. doi:10.1016/S0076-6879(10)70007-0.
- [115] O. Edelheit, A. Hanukoglu, I. Hanukoglu, Simple and efficient site-directed mutagenesis using two single-primer reactions in parallel to generate mutants for protein structure-function studies, *BMC Biotechnol.* 9 (2009) 61. doi:10.1186/1472-6750-9-61.
- [116] A.J. Roger, A.G.B. Simpson, Evolution: Revisiting the Root of the Eukaryote Tree, *Curr. Biol.* 19 (2009). doi:10.1016/j.cub.2008.12.032.
- [117] D. Benelli, S. Marzi, C. Mancone, T. Alonzi, A. la Teana, P. Londei, Function and ribosomal localization of aIF6, a translational regulator shared by archaea and eukarya., *Nucleic Acids Res.* 37 (2009) 256–67. doi:10.1093/nar/gkn959.

- [118] Y. Kato, M. Konishi, M. Shigyo, T. Yoneyama, S. Yanagisawa, Characterization of plant eukaryotic translation initiation factor 6 (eIF6) genes: The essential role in embryogenesis and their differential expression in Arabidopsis and rice, *Biochem. Biophys. Res. Commun.* 397 (2010) 673–678. doi:10.1016/j.bbrc.2010.06.001.
- [119] L. Ruohonen, M.K. Aalto, S. Keränen, Modifications to the ADH1 promoter of *Saccharomyces cerevisiae* for efficient production of heterologous proteins., *J. Biotechnol.* 39 (1995) 193–203. doi:10.1016/0168-1656(95)00024-K.
- [120] J. Sun, Z. Shao, H. Zhao, N. Nair, F. Wen, J.H. Xu, H. Zhao, Cloning and characterization of a panel of constitutive promoters for applications in pathway engineering in *Saccharomyces cerevisiae*, *Biotechnol. Bioeng.* 109 (2012) 2082–2092. doi:10.1002/bit.24481.
- [121] M. Kornprobst, M. Turk, N. Kellner, J. Cheng, D. Flemming, I. Koš-Braun, M. Koš, M. Thoms, O. Berninghausen, R. Beckmann, E. Hurt, Architecture of the 90S Pre-ribosome: A Structural View on the Birth of the Eukaryotic Ribosome, *Cell.* 166 (2016) 380–393. doi:10.1016/j.cell.2016.06.014.
- [122] S. Granneman, S.J. Baserga, Ribosome biogenesis: of knobs and RNA processing., *Exp. Cell Res.* 296 (2004) 43–50. doi:10.1016/j.yexcr.2004.03.016.
- [123] K. Karbstein, Quality control mechanisms during ribosome maturation., *Trends Cell Biol.* 23 (2013) 242–50. doi:10.1016/j.tcb.2013.01.004.
- [124] A. Gijbers, A. García-Márquez, A. Luviano, N. Sánchez-Puig, Guanine nucleotide exchange in the ribosomal GTPase EFL1 is modulated by the protein mutated in the Shwachman-Diamond Syndrome, *Biochem. Biophys. Res. Commun.* 437 (2013) 349–354. doi:10.1016/j.bbrc.2013.06.077.
- [125] J.F. Fahrni, I. Bolivar, C. Berney, E. Nassonova, A. Smirnov, J. Pawlowski, Phylogeny of Lobose Amoebae Based on Actin and Small-Subunit Ribosomal RNA Genes, *Mol. Biol. Evol.* 20 (2003) 1881–1886. doi:10.1093/molbev/msg201.
- [126] A.G.B. Simpson, Cytoskeletal organization, phylogenetic affinities and systematics in the contentious taxon Excavata (Eukaryota), *Int. J. Syst. Evol. Microbiol.* 53 (2003) 1759–1777. doi:10.1099/ijs.0.02578-0.
- [127] T. Cavalier-Smith, Chloroplast evolution: Secondary symbiogenesis and multiple losses, *Curr. Biol.* 12 (2002). doi:10.1016/S0960-9822(01)00675-3.
- [128] P.G. Falkowski, M.E. Katz, A.H. Knoll, A. Quigg, J.A. Raven, O. Schofield, F.J.R. Taylor, The Evolution of Modern Eukaryotic Phytoplankton, *Science* (80-.). 305 (2004) 354–360. doi:10.1126/science.1095964.
- [129] S.I. Nikolaev, C. Berney, J.F. Fahrni, I. Bolivar, S. Polet, A.P. Mylnikov, V. V Aleshin, N.B. Petrov, J. Pawlowski, The twilight of Heliozoa and rise of Rhizaria, an emerging supergroup of amoeboid eukaryotes., *Proc. Natl. Acad. Sci. U. S. A.* 101 (2004) 8066–8071. doi:10.1073/pnas.0308602101.
- [130] J.J. Harrington, B. Sherf, S. Rundlett, P.D. Jackson, R. Perry, S. Cain, C. Leventhal, M. Thornton, R. Ramachandran, J. Whittington, L. Lerner, D. Costanzo, K. McElligott, S. Boozer, R. Mays, E. Smith, N. Veloso, a Klika, J. Hess, K. Cothren, K. Lo, J. Offenbacher, J. Danzig, M. Ducar, Creation of genome-wide protein expression libraries using random activation of gene expression., *Nat. Biotechnol.* 19 (2001) 440–445. doi:10.1038/88107.
- [131] A.J. Finch, C. Hilcenko, N. Basse, L.F. Drynan, B. Goyenechea, T.F. Menne,

- A. González Fernández, P. Simpson, C.S. D'Santos, M.J. Arends, J. Donadieu, C. Bellanné-Chantelot, M. Costanzo, C. Boone, A.N. McKenzie, S.M. V Freund, A.J. Warren, Uncoupling of GTP hydrolysis from eIF6 release on the ribosome causes Shwachman-Diamond syndrome., *Genes Dev.* 25 (2011) 917–29. doi:10.1101/gad.623011.
- [132] W. V. Gilbert, Functional specialization of ribosomes?, *Trends Biochem. Sci.* 36 (2011) 127–132. doi:10.1016/j.tibs.2010.12.002.
- [133] A. Filipovska, O. Rackham, Specialization from synthesis: How ribosome diversity can customize protein function, *FEBS Lett.* 587 (2013) 1189–1197. doi:10.1016/j.febslet.2013.02.032.
- [134] S. Xue, M. Barna, Specialized ribosomes: a new frontier in gene regulation and organismal biology, *Nat. Rev. Mol. Cell Biol.* 13 (2012) 355–369. doi:10.1038/nrm3359.
- [135] T. Preiss, All Ribosomes Are Created Equal. Really?, *Trends Biochem. Sci.* 41 (2016) 121–123. doi:10.1016/j.tibs.2015.11.009.
- [136] J.S. Graindorge, J.C. Rousselle, B. Senger, P. Lenormand, A. Namane, F. Lacroute, F. Fasiolo, Deletion of EFL1 results in heterogeneity of the 60 S GTPase-associated rRNA conformation, *J. Mol. Biol.* 352 (2005) 355–369. doi:10.1016/j.jmb.2005.07.037.
- [137] A.G. Murzin, How far divergent evolution goes in proteins, *Curr. Opin. Struct. Biol.* 8 (1998) 380–387. doi:10.1016/S0959-440X(98)80073-0.
- [138] D.R. Flower, The lipocalin protein family: structure and function., *Biochem. J.* 318 (Pt 1 (1996) 1–14. doi:10.1042/bj3180001.
- [139] M.D. Ganfornina, G. Gutiérrez, M. Bastiani, D. S, A Phylogenetic Analysis of the Lipocalin Protein Family, *Mol. Biol. Evol.* 17 (2000) 114–126. doi:10.1093/oxfordjournals.molbev.a026224.
- [140] D.R. Flower, A.C.T. North, C.E. Sansom, The lipocalin protein family: Structural and sequence overview, *Biochim. Biophys. Acta - Protein Struct. Mol. Enzymol.* 1482 (2000) 9–24. doi:10.1016/S0167-4838(00)00148-5.
- [141] L.C. James, D.S. Tawfik, Conformational diversity and protein evolution - A 60-year-old hypothesis revisited, *Trends Biochem. Sci.* 28 (2003) 361–368. doi:10.1016/S0968-0004(03)00135-X.
- [142] M.G. Gomez-Lorenzo, C.M. Spahn, R.K. Agrawal, R.A. Grassucci, P. Penczek, K. Chakraborty, J.P. Ballesta, J.L. Lavandera, J.F. Garcia-Bustos, J. Frank, Three-dimensional cryo-electron microscopy localization of EF2 in the *Saccharomyces cerevisiae* 80S ribosome at 17.5 Å resolution., *EMBO J.* 19 (2000) 2710–8. doi:10.1093/emboj/19.11.2710.
- [143] S.R. Connell, C. Takemoto, D.N. Wilson, H. Wang, K. Murayama, T. Terada, M. Shirouzu, M. Rost, M. Schüler, J. Giesebrecht, M. Dabrowski, T. Mielke, P. Fucini, S. Yokoyama, C.M.T. Spahn, Structural Basis for Interaction of the Ribosome with the Switch Regions of GTP-Bound Elongation Factors, *Mol. Cell.* 25 (2007) 751–764. doi:10.1016/j.molcel.2007.01.027.
- [144] E.S. Haag, S. Wang, J. Kimble, Rapid coevolution of the nematode sex-determining genes *fem-3* and *tra-2*, *Curr. Biol.* 12 (2002) 2035–2041. doi:10.1016/S0960-9822(02)01333-7.
- [145] M.G. Alfonso, Estudios de interacción entre las proteínas Efl1 y Sdo1 mediante doble híbrido en levadura, Universidad Nacional Autónoma de México, 2015.

- [146] A.K.H. Choi, E.C.K. Wong, K.M. Lee, K.B. Wong, Structures of eukaryotic ribosomal stalk proteins and its complex with trichosanthin, and their implications in recruiting ribosome-inactivating proteins to the ribosomes, *Toxins (Basel)*. 7 (2015) 638–647. doi:10.3390/toxins7030638.
- [147] M. Helgstrand, C.S. Mandava, F.A.A. Mulder, A. Liljas, S. Sanyal, M. Akke, The Ribosomal Stalk Binds to Translation Factors IF2, EF-Tu, EF-G and RF3 via a Conserved Region of the L12 C-terminal Domain, *J. Mol. Biol.* 365 (2007) 468–479. doi:10.1016/j.jmb.2006.10.025.
- [148] H. Imai, T. Miyoshi, R. Murakami, K. Ito, Y. Ishino, T. Uchiumi, Functional role of the C-terminal tail of the archaeal ribosomal stalk in recruitment of two elongation factors to the sarcin/ricin loop of 23S rRNA, *Genes to Cells*. 20 (2015) 613–624. doi:10.1111/gtc.12256.
- [149] M. Diaconu, U. Kothe, F. Schlünzen, N. Fischer, J.M. Harms, A.G. Tonevitsky, H. Stark, M. V. Rodnina, M.C. Wahl, Structural basis for the function of the ribosomal L7/12 stalk in factor binding and GTPase activation, *Cell*. 121 (2005) 991–1004. doi:10.1016/j.cell.2005.04.015.
- [150] A.R. Subramanian, Copies of proteins L7 and L12 and heterogeneity of the large subunit of *Escherichia coli* ribosome, *J. Mol. Biol.* 95 (1975). doi:10.1016/0022-2836(75)90330-7.
- [151] Y. Maki, T. Hashimoto, M. Zhou, T. Naganuma, J. Ohta, T. Nomura, C. V. Robinson, T. Uchiumi, Three binding sites for stalk protein dimers are generally present in ribosomes from archaeal organism, *J. Biol. Chem.* 282 (2007) 32827–32833. doi:10.1074/jbc.M705412200.
- [152] T. Naganuma, N. Nomura, M. Yao, M. Mochizuki, T. Uchiumi, I. Tanaka, Structural basis for translation factor recruitment to the eukaryotic/archaeal ribosomes, *J. Biol. Chem.* 285 (2010) 4747–4756. doi:10.1074/jbc.M109.068098.
- [153] K.M. Lee, C.W.H. Yu, D.S.B. Chan, T.Y.H. Chiu, G. Zhu, K.H. Sze, P.C. Shaw, K.B. Wong, Solution structure of the dimerization domain of ribosomal protein P2 provides insights for the structural organization of eukaryotic stalk, *Nucleic Acids Res.* 38 (2010) 5206–5216. doi:10.1093/nar/gkq231.
- [154] T. Shimizu, M. Nakagaki, Y. Nishi, Y. Kobayashi, A. Hachimori, T. Uchiumi, Interaction among silkworm ribosomal proteins P1, P2 and P0 required for functional protein binding to the GTPase-associated domain of 28S rRNA., *Nucleic Acids Res.* 30 (2002) 2620–2627.
- [155] T. Uchiumi, K. Hori, T. Nomura, A. Hachimori, Replacement of L7/L12.L10 protein complex in *Escherichia coli* ribosomes with the eukaryotic counterpart changes the specificity of elongation factor binding, *J. Biol. Chem.* 274 (1999) 27578–27582. doi:10.1074/jbc.274.39.27578.
- [156] T. Uchiumi, S. Honma, T. Nomura, E.R. Dabbs, A. Hachimori, Translation elongation by a hybrid ribosome in which proteins at the GTPase center of the *Escherichia coli* ribosome are replaced with rat counterparts, *J. Biol. Chem.* 277 (2002) 3857–3862. doi:10.1074/jbc.M107730200.
- [157] S. Harju, H. Fedosyuk, K.R. Peterson, Rapid isolation of yeast genomic DNA: Bust n' Grab, *BMC Biotechnol.* 4 (2004) 1. doi:10.1186/1472-6750-4-8.

11 Información suplementaria

Tabla Suplementaria 1. Información de las proteínas empleadas para realizar el árbol filogenético de las proteínas EFL1, EF-2, EF-G y EF-4.

Nombre	Organismo	ID Uniprot
eEF-2_ARATH	<i>Arabidopsis thaliana</i>	Q9SGT4_ARATH
eEF-2_DICDI	<i>Dictyostelium discoideum</i>	EF2_DICDI
eEF-2_CAEEL	<i>Caenorhabditis elegans</i>	EF2_CAEEL
eEF-2_DROME	<i>Drosophila melanogaster</i>	EF2_DROME
eEF-2_CHICK	<i>Gallus gallus</i>	EF2_CHICK
eEF-2_HUMAN	<i>Homo sapiens</i>	EF2_HUMAN
eEF-2_MOUSE	<i>Mus musculus</i>	EF2_MOUSE
eEF-2_RAT	<i>Rattus norvegicus</i>	EF2_RAT
eEF-2_YEAST	<i>Saccharomyces cerevisiae</i>	EF2_YEAST
eEF-2_SCHPO	<i>Schizosaccharomyces pombe</i>	EF2_SCHPO
aEF-2_HALS3	<i>Halobacterium salinarum</i>	EF2_HALS3
aEF-2_METVS	<i>Methanococcus vannielii</i>	EF2_METVS
aEF-2_ARCFU	<i>Archaeoglobus fulgidus</i>	EF2_ARCFU
aEF-2_THEKO	<i>Thermococcus kodakaraensis</i>	EF2_THEKO
aEF-2_PYRHO	<i>Pyrococcus horikoshii</i>	EF2_PYRHO
aEF-2_METB6	<i>Methanoregula boonei</i>	EF2_METB6
aEF-2_DESK1	<i>Desulfurococcus kamchatkensis</i>	EF2_DESK1
EFL1_MOUSE	<i>Mus musculus</i>	ETUD1_MOUSE
EFL1_HUMAN	<i>Homo sapiens</i>	ETUD1_HUMAN
EFL1_RAT	<i>Rattus norvegicus</i>	D3ZXJ5_RAT
EFL1_MELGA	<i>Meleagris gallopavo</i>	G1N7H9_MELGA
EFL1_XENTR	<i>Xenopus tropicalis</i>	F6Z2L8_XENTR
EFL1_DANRE	<i>Danio rerio</i>	E7FAW9_DANRE
EFL1_YEAST	<i>Saccharomyces cerevisiae</i>	RIA1_YEAST
EFL1_DROME	<i>Drosophila melanogaster</i>	Q9VV61_DROME
EFL1_SCHPO	<i>Schizosaccharomyces pombe</i>	RIA1_SCHPO
EFL1_ARATH	<i>Arabidopsis thaliana</i>	F4J2T9_ARATH
EFL1_ARALL	<i>Arabidopsis lyrata</i>	D7L2H7_ARALL
EFL1_DICDI	<i>Dictyostelium discoideum</i>	Q54WF2_DICDI
EF-G_LAWIP	<i>Lawsonia intracellularis</i>	EFG_LAWIP
EF-G_STRAW	<i>Streptomyces avermitilis</i>	EFG_STRAW
EF-G_STRCO	<i>Streptomyces coelicolor</i>	EFG1_STRCO
EF-G_BACFN	<i>Bacteroides fragilis</i>	EFG_BACFN
EF-G_DESVM	<i>Desulfovibrio vulgaris</i>	EFG_DESVM
EF-G_THEFY	<i>Thermobifida fusca</i>	EFG_THEFY
EF-G_NOSP7	<i>Nostoc punctiforme</i>	EFG_NOSP7
EF-G_NOCSJ	<i>Nocardioides sp.</i>	EFG_NOCSJ
EF-G_PETMO	<i>Petrotoga mobilis</i>	EFG_PETMO
EF-G_BACTN	<i>Bacteroides thetaiotaomicron</i>	EFG_BACTN
EF-G_DESAG	<i>Desulfovibrio alaskensis</i>	EFG_DESAG
EF-G_THEM4	<i>Thermosipho melanesiensis</i>	EFG_THEM4

EF-G_STRGG	<i>Streptomyces griseus subsp. griseus</i>	EFG_STRGG
EF-G_ANAVT	<i>Anabaena variabilis</i>	EFG_ANAVT
EF-4_HELPS	<i>Helicobacter pylori</i>	LEPA_HELPS
EF-4_BIFLD	<i>Bifidobacterium longum</i>	LEPA_BIFLD
EF-4_SYNS3	<i>Synechococcus sp.</i>	LEPA_SYNS3

Oligonucleótidos

Tabla Suplementaria 2. Oligonucleótidos empleados para ensamblar el vector de expresión pAM2 por PCR de traslape.

Nombre	Secuencia 5'-3'
F-Cmyc	GAACAAAAATTGATATCTGAAGAAGATTTGTAAGCGAATTTCTTATGATTTATGATT
R-Cmyc	TTACAAATCTTCTTCAGATATCAATTTTTGTTTCGGTATCGATAAGCTTGAATATCGAATTC
M13-Forward	TGTAAAACGACGGCCAGT
M13-Reverse	CAGGAAACAGCTATGACCATG

Tabla Suplementaria 3. Oligonucleótidos empleados para amplificar las secuencias codificantes de los miembros de la familia EFL1.

Nombre	Secuencia 5'-3'
ScFEfl1	GCGCgtcgacAATGCCTAGAGTGAATCGGAGACTTAC
ScREfl1	CCactagtATTCTTTTTCAAAGTACGTTGTTTTTC
HsFEfl1-1	GCGCgtcgacAATGGTGCTCAACAGTTTGGATAAGATG
HsREfl1-1	CGCggatccCTTATTTTTGCTGAGTGTCTCTGCTGC
SpFEfl1	CAACTATCTCATATACgtcgacACAATGCCTGTTATTCCTCCAG
SpREfl1	ATTCCTGCAGCCCGGGactagtATGCTTCAGTGTCTCTGCT
MmFEfl1	CAACTATCTCATATACgtcgacACAATGGTGCTCAGTGGTGTGAT
MmREfl1	ATTCCTGCAGCCCGGGactagtCTTGTTTTTGCTGAGGGTTCT
AtFEfl1	CAACTATCTCATATACgtcgacACAATGGATGAATCTGAAGGTAG
AtREfl1	ATTCCTGCAGCCCGGGactagtAACTTTACGAGCCAAGGTTCT

Tabla Suplementaria 4. Oligonucleótidos empleados para amplificar las secuencias codificantes de los miembros de la familia SDO1.

Nombre	Secuencia 5'-3'
ScFSDO1	acgcgtcgacACAATGCCTATCAATCAACCGTC
ScRSDO1	ccactagtGTTATGCGTTGTATTATCTATG
SpFSDO1	ACGCgtcgacACAATGGTTATTAGTCAACCTGTT
SpRSOD1	CCactagtAGTCCTAGCCTCGCTCATG
Rex1 SpSDO1	GATTTGTAAAACCTTCATCCAGATCTGTTTCAATTTGTTTCTCCATTCTGTAACTTT
Fex2 SpSDO1	TGAAACAGATCTGGATGAAG
HsFSBDS	ACGCgtcgacACAATGTCGATCTTCACCCCCAC
HsRSBDS	CCactagtTTCAAATTTCTCATCTCCTTC
MmFSBDS	ACGCgtcgacACAATGTCGATCTTCACCCCCAC
MmRSBDS	CCactagtTCAAATTTCTCATCGCCTTCC

AtFSBDS	ACGCgtcgacACAATGTCTGAAGACGCTGGTGCAAC
AtRSBDS	CCactagtGAAAGAGTAGTCTTTCAAATCTG
AfFSBDS	ACGCgtcgacacaATGGTATCCCTCGATAAGGC
AfRSBDS	CCactagtTCCTATCCTCCTCAAACCTT

Tabla Suplementaria 5. Oligonucleótidos empleados para amplificar el módulo de resistencia *NatMX4*.

Nombre	Secuencia 5'-3'
FEFL1::Nat	GTTTATAAAAATAACCAAACCTCAGTTATTTTTCTAAGCACGATAAGCAAATACA TGGAGGCCCAGAATACCCT
REFL1::Nat	CTAGTGAAAACCTATCTATATTAGTCTATGATTCATATAAACCTAAAAACCC AGTATAGCGACCAGCATTAC

Tabla Suplementaria 6. Oligonucleótidos empleados para amplificar el módulo de resistencia *KanMX6*.

Nombre	Secuencia 5'-3'
FSDO1::Kan	TGAGAAAGAAAACACGCAATGGGTAAGGAAAAGACTCAC
RSDO::1Kan	TGTATGCAAAGTGATTCTATATATAAATTAGAAAACTCATCGAGCATC

Tabla Suplementaria 7. Oligonucleótidos empleados para agregar el codón de termino a las construcciones para evaluar la conservación funcional de los miembros de la familia EFL1.

Nombre	Secuencia 5'-3'
FHsEFL1stop	GACTCTCAGCAAAAATAAGGGTAAGGATCCTGGCGGCCGCTC
RHsEFL1stop	GAGCGGCCGCCAGGATCCTTACCCTTATTTTTGCTGAGTGTC
FScEFL1stop	CGTACTTTGAAAAAGAATTAACCTAGTCCCGGGCTGCAGG
RScEFL1stop	CCTGCAGCCCCGGGACTAGTTTAAATTCTTTTTCAAAGTACG
FSpEFL1stop	GCAGAGGACACTGAAGCATTAACTAGTCCCGGGCTGCAGG
RSpEFL1stop	CCTGCAGCCCCGGGACTAGTTTAAATGCTTCAGTGTCTCTGC
FMmEFL1stop	CCCTCAGCAAAAACAAGTAACTAGTCCCGGGCTGCAGG
RMmEFL1stop	CCTGCAGCCCCGGGACTAGTTTACTTGTTTTTGCTGAGGG

Medios de cultivo

Tabla Suplementaria 8. Composición de los medios de cultivo para bacteria.

Medio de cultivo	Componentes
Medio Luria Bertani (LB)	NaCl 1%, Extracto de levadura 0.5%, Triptona 1%, pH 7.5, Agar 1.5% (solo para cajas)
Medio Luria Bertani + ampicilina	NaCl 1%, Extracto de levadura 0.5%, Triptona 1%, ampicilina 100 µg/L, pH 7.5. Agar 1.5% (solo para placas)

Tabla Suplementaria 9. Composición de los medios de cultivo para levadura.

Medio de cultivo	Componentes
Yeast Peptone Dextrose (YPD)	Extracto de levadura 1.1%, Peptona 2.2%, Sulfato de adenina 0.012%, Glucosa 2%
Yeast Peptone Dextrose (YPD) + cloNAT	Extracto de levadura 1.1%, Peptona 2.2%, Sulfato de adenina 0.012%, Glucosa 2%, cloNAT 0.1 mg/L
Yeast Peptone Dextrose (YPD) + cloNAT + G418	Extracto de levadura 1.1%, Peptona 2.2%, Sulfato de adenina 0.012%, Glucosa 2%, cloNAT 0.1 mg/mL, G418 0.2 mg/L
Medio SD –Ura	Base nitrogenada 0.8%, Casaminoácidos 0.11%, Sulfato de adenina 0.006%, L-tirosina 0.006%, Glucosa 2%, Dropout +Leu +Trp 1X, pH 5.8, Agar 2% (solo para cajas)
Medio SD –Met	Base nitrogenada 0.2%, Sulfato de adenina 0.006%, L-tirosina 0.006%, Uracilo 0.006%, Dropout –Met 1X, pH 5.8, Agar 2% (solo para cajas)
Medio SD –Met-Ura	Base nitrogenada 0.8%, Sulfato de adenina 0.006%, Tirosina 0.006%, Dropout –Met 1X, pH 5.8, Agar 2% (solo para cajas)
Medio de esporulación	Acetato de Potasio 1%, Extracto de levadura 0.1%, Dextrosa 0.05%, Polvo de aminoácidos para esporulación 0.01%, Agar 2%

Tabla Suplementaria 10. Composición de los medios de cultivo para los ensayos de complementación genética.

Medio de cultivo	Composición
Medio de crecimiento sintético -His/Lis/Arg + canavanina + tialisina	Base nitrogenada de levadura sin aminoácidos y sulfato de amonio 0.17%, glutamato monosódico 0.1%, suplemento de aminoácidos 0.2%, agar 2%, glucosa 2%, canavanina 250 µg/L, tialisina 250 µg/L
Medio de crecimiento sintético – His/Lis/Arg + canavanina + tialisina + cloNAT	Base nitrogenada de levadura sin aminoácidos y sulfato de amonio 0.17%, glutamato monosódico 0.1%, suplemento de aminoácidos (sin histidina, lisina y arginina) 0.2%, agar 2%, glucosa 2%, canavanina 250 µg/L, tialisina 250 µg/L, cloNAT 0.1 mg/L
Medio de crecimiento sintético -His/Lis/Arg + canavanina + tialisina + G418	Base nitrogenada de levadura sin aminoácidos y sulfato de amonio 0.17%, glutamato monosódico 0.1%, suplemento de aminoácidos (sin histidina, lisina y arginina) 0.2%, agar 2%, glucosa 2%, canavanina 250 µg/L, tialisina 250 mg/L, G418 0.2 mg/L

Medio de crecimiento sintético -His/Lis/Arg + canavanina + tialisina + cloNAT + G418	base nitrogenada de levadura sin aminoácidos y sulfato de amonio 0.17%, glutamato monosódico 0.1%, suplemento de aminoácidos (sin histidina, lisina y arginina) 0.2%, agar 2%, glucosa 2%, canavanina 250 µg/L, tialisina 250 µg/L, cloNAT 0.1 mg/L y G418 0.2 mg/L
---	---

Tabla Suplementaria 11. Composición de los suplementos de aminoácidos para los medios de cultivo de levadura.

Suplemento	Composición
Dropout + Leu + Trp 100X	Leucina 0.6%, Triptófano 0.4%.
Dropout – Met 100X	Arginina 0.2%, Histidina 0.1%, Isoleucina 0.6%, Leucina 0.6%, Lisina 0.4%, Fenilalanina 0.6%, Treonina 0.5% y Triptófano 0.4%
Suplemento de aminoácidos para medio de esporulación	Histidina 2 g, Leucina 2 g, lisina 2 g, uracilo 2 g. Se utiliza 0.1 g de la mezcla de polvo por cada L de medio
Suplemento de aminoácidos en polvo para medio de crecimiento sintético	Adenina 3 g, Uracilo 2 g, Inositol 2 g, Ácido p-aminobenzoico 0.2 g, Alanina 2 g, Arginina 2 g, Asparagina 2 g, Ácido aspártico 2 g, Cisteína 2 g, Ácido glutámico 2 g, Glutamina 2 g, Glicina 2 g, Histidina 2 g, Isoleucina 2 g, Leucina 10 g, Lisina 2 g, Metionina 2 g, Fenilalanina 2 g, Prolina 2 g, Serina 2 g, Treonina 2 g, Triptófano 2 g, Tirosina 2 g, Valina 2 g. Se utiliza 2 g de la mezcla de polvo por L de medio

Apéndice 1. Extracción de DNA genómico de células de *Saccharomyces cerevisiae*.

Se transfirieron 1.5 mL de cada uno de los cultivos de levadura a un tubo de microcentrífuga y se centrifugaron (2 min, 10,000 rpm) para obtener el botón celular. En seguida, las células se resuspendieron en 200 µL de solución amortiguadora Harju [157] (Tabla Suplementaria 12) y se provocó la lisis celular mediante tres ciclos de congelación-descongelación que consistieron en 2 min en nitrógeno líquido seguidos de 1 min en un baño de agua a 94 °C. Posteriormente, se agregaron 200 µL de cloroformo a cada uno de los tubos y se mezclaron por 2 min en vórtex. Después se centrifugó cada tubo 3 min a 14,500 rpm, la fase acuosa se transfirió a un tubo de microcentrífuga nuevo con 400 µL de etanol absoluto frío. Los tubos se mezclaron por inversión y para promover la precipitación del DNA se incubaron a –20 °C durante 30 min. Terminado el tiempo de incubación, el sobrenadante se eliminó por centrifugación (5 min, 14,500 rpm) y la pastilla formada se lavó con 500 µL de una solución de etanol al 70%. Una vez más, el sobrenadante fue eliminado por centrifugación (5 min, 14500 rpm) y las trazas del mismo se evaporaron mediante incubación de los tubos en una estufa a

37 °C. Finalmente, el DNA se resuspendió en 100 µL de agua libre de nucleasas y se trató con RNAsa (100 µg/mL) para eliminar las trazas de RNA.

Tabla Suplementaria 12. Composición de la solución amortiguadora Harju.

Solución	Composición
Amortiguador Harju	Triton X-100 2%, SDS 1%, NaCl 100 mM, Tris-HCl 10 mM, pH 8.0, EDTA 1 mM, pH 8.0

Apéndice 2. Aislamiento y purificación de RNA total empleando RNAzol.

Se inocularon 3 mL de medio YPD con una colonia de *S. pombe* y se incubó a 30 °C durante 18 h con agitación constante (250 rpm). Al día siguiente, se agregó la cantidad necesaria del cultivo anterior a 3 mL de medio YPD fresco para alcanzar una DO_{600nm} de 0.2. Este cultivo se mantuvo a 30 °C con agitación constante (250 rpm) hasta alcanzar una DO_{600nm} de 0.7 – 0.8. Una vez alcanzada la fase de crecimiento exponencial, se transfirieron 1.5 mL del cultivo a un tubo de microcentrifuga y se centrifugó por 2 min a 10,000 rpm. Posteriormente, el sobrenadante se eliminó por decantación y el botón celular se resuspendió en 1 mL de RNAzol (Sigma). La suspensión celular se mantuvo en hielo durante 15 min y después se centrifugó durante 10 min a 14,000 rpm y 4 °C. La fase acuosa fue transferida a un tubo nuevo con 500 µL de isopropanol (Sigma). La mezcla se mantuvo a temperatura ambiente durante 10 min para promover la precipitación del RNA y enseguida se centrifugó por 20 min a 14,000 rpm y 4 °C. El sobrenadante se eliminó por decantación y la pastilla se lavó con 500 µL de una solución de etanol al 70% v/v. El tubo se centrifugó una vez más por 5 min a 14,000 y 4 °C. El sobrenadante se eliminó por decantación y las trazas por evaporación. Finalmente, la pastilla se resuspendió en 25 µL de agua libre de nucleasas y se almacenó a -80 °C.

Apéndice 3. Extracción de DNA plasmídico por el método de EasyPrep.

Se inocularon 5 mL de medio LB + ampicilina (Tabla Suplementaria 8) con una clona transformante y se incubó a 37 °C durante toda la noche. Al día siguiente, el paquete celular fue obtenido mediante centrifugación (2 min, 10,000 rpm) y se resuspendió en 100 µL de amortiguador EasyPrep (Tabla Suplementaria 13) usando vórtex. Cada una de las mezclas fueron incubadas a temperatura ambiente por 15 minutos y posteriormente se dió un choque térmico a 94 °C durante 1 min y enseguida se colocaron en hielo durante 4 min.

Tabla Suplementaria 13. Composición de la solución amortiguadora EasyPrep.

Solución	Composición
Amortiguador EasyPrep	Tris-HCl 10 mM pH 8, EDTA 1 mM, BSA 100 mg/L, RNAsa A 200 µg/L, Sacarosa 15%, Lizosima 1 mg/L

Apéndice 4. Precipitación de DNA con etanol absoluto.

De acuerdo con el volumen de muestra, se adicionó al tubo un décimo de solución de acetato de sodio 3 M pH 5.2 y se mezcló por vórtex. En seguida, se agregaron 2.5 volúmenes de etanol absoluto frío (Sigma) y se incubó a -20 °C durante 30 min para promover la precipitación del DNA. Posteriormente, la mezcla fue centrifugada por 15 min a 14,000 rpm. El sobrenadante se eliminó por decantación y la pastilla se lavó con una solución de etanol al 70% v/v. Una vez más, la mezcla se centrifugó por 5 min a 14,000 rpm, el sobrenadante se eliminó por decantación y sus trazas por evaporación. Finalmente, el DNA se resuspendió en 25 µL de agua libre de nucleasas y se almacenó a -20 °C.

Apéndice 5. Transformación bacteriana por choque térmico.

Se mezclaron 5 µL de DNA plasmídico o 10 µL de reacción de ligación con 50 µL de células quimiocompetentes de *Escherichia coli* DH10B y se mantuvieron en hielo durante 30 min. Posteriormente, se les dió un choque térmico a 42 °C durante 1 min seguido de 2 min en hielo. Después, se agregaron 500 µL medio LB a cada tubo de células y se incubaron a 37 °C durante 1 h con agitación constante (200 rpm). Finalmente, cada reacción de transformación fue plaqueda en medio LB sólido con ampicilina (Tabla Suplementaria 8) y se incubó durante 16 h a 37 °C.

Apéndice 6. Preparación de células competentes de levadura

Se inocularon 10 mL de medio YPD (Tabla Suplementaria 9) con la cepa de interés de *Saccharomyces cerevisiae* y se incubó a 30 °C durante toda la noche con agitación constante (250 rpm). Posteriormente, se inocularon 200 mL de medio YPD fresco con la cantidad de pre-inóculo necesaria para alcanzar un DO₆₀₀ de 0.1. Este cultivo se incubó a 30 °C con agitación hasta alcanzar la fase de crecimiento logarítmica (DO₆₀₀ entre 0.6 y 0.8). El paquete celular se obtuvo por centrifugación (5000 rpm, 5 min, 21 °C) y se resuspendió en 2 mL de la solución CY (Tabla Suplementaria 14) para después ser repartido en alícuotas de 100 µL.

Apéndice 7. Transformación de células de *Saccharomyces cerevisiae* por el método de polietilenglicol/acarreador de DNA/DMSO

Se mezclaron 200 ng de DNA plasmídico con 20 µL (100 µg) de acarreador de DNA (Invitrogen) y se agregaron a 100 µL de células competentes de levadura (Apéndice 6). A cada tubo se le adicionó 600 µL de la solución TY (Tabla Suplementaria 14), los tubos se mezclaron por inversión y se mantuvieron a 30 °C durante 2 h con agitación constante (250 rpm). Posteriormente, se agregó 70 µL de DMSO (Sigma) a cada tubo y se les dió un choque térmico a 42 °C por 15 min. Para eliminar el sobrenadante, cada tubo se centrifugó a 5,000 rpm durante 1 min. Cada botón celular se resuspendió en 100 µL de amortiguador TE y se plaqueó en medio de selección. Los cultivos fueron incubados a 30 °C durante 3 días.

Tabla Suplementaria 14. Composición de las soluciones utilizadas para la preparación de células competentes de *Saccharomyces cerevisiae* y su transformación

Solución	Composición
TE 10X	Tris-HCl 100 mM pH 7.4, EDTA 10 mM pH 8.0
PEG 4000 50 % v/v	25 g de PEG 4000 y aforar a 50 mL con H ₂ O
CY	TE 1X, LiCl 100 mM, en agua
TY	TE 1X, LiCl 100 mM, en PEG 4000 50 % p/v

Apéndice 8. Genotipo relevante de la cepa de *E. coli* DH10B.

F⁻ *endA1 deoR⁺ recA1 galE15 galK16 nupG rpsL Δ(lac)X74 φ80lacZΔM15 araD139 Δ(ara,leu)7697 mcrA Δ(mrr-hsdRMS-mcrBC) Str^R λ⁻*