



UNIVERSIDAD AUTÓNOMA DEL ESTADO DE MORELOS
INSTITUTO DE INVESTIGACIÓN EN CIENCIAS BÁSICAS Y APLICADAS
CENTRO DE INVESTIGACIÓN EN DINÁMICA CELULAR

**“Análisis in silico de las secuencias de unión a ribosoma que
definen el inicio de la traducción en procariontes”**

TESIS PROFESIONAL PARA OBTENER EL GRADO DE:

DOCTOR EN CIENCIAS

PRESENTA:
KAREL JOHAN ESTRADA GUERRA

DIRECTOR: DR. ENRIQUE MERINO PÉREZ

CUERNAVACA, MORELOS
2023

SEPTIEMBRE,



Siempre, a mis hijos.

AGRADECIMIENTOS

A Enrique Merino: por acompañarme y guiarme en este viaje, por toda la paciencia y por esos jalones de mano que me ayudaron a mantener la inercia y seguir adelante. ¡Llegué, tío Lucas, lo logramos!

A mi comité tutorial: Sonia Dávila, Armando Hernández y Cinthia Núñez, por su disposición constante para firmar las actas en la última noche antes del cierre del sistema, porque el estudiante dejó el tutorial para el último día. Mil gracias por toda la ayuda, las buenas ideas y el aliento en el camino.

A mis sinodales: los doctores Ramón González, Ramón Batista, Guillermo Gosset, Miguel Ángel Cevallos, Verónica Narváez, María Eugenia Núñez y Edgar Dantán, muchas gracias por las pláticas edificantes, las correcciones y los consejos.

A Alejandro Sánchez y Lorenzo Segovia, por todo el apoyo.

A Félix Santana; los gráficos vectoriales escalables y los scripts de R no serán lo mismo sin las sesiones de “having fun with aserin”.

A Alejandro Garcarrubio: Silvio tenía razón, qué difícil es contar una historia cuando no se habla de un viejo, de un niño o de sí mismo... gracias, mi querido Ale.

A mi familia:

A los mejores abuelitos del mundo, Nilo y Ana.

A mis padres y hermano, por el amor y el ejemplo enorme de nunca detenerse.

Al amor de mi vida, por creer que la belleza no se rinde ante el poder, por preferir amar.

NOMENCLATURA.....	6
RESUMEN.....	7
ABSTRACT.....	8
1. INTRODUCCIÓN.....	9
1.1 Mecanismos alternativos de traducción en procariontes.....	14
1.1.1 Mecanismo mediado por una proteína ribosomal S1.....	14
1.1.2 Mecanismo utilizado para la traducción de mRNA sin líder.....	15
1.1.3 Mecanismo dependiente de la estructura secundaria del mRNA.....	16
2. OBJETIVOS.....	17
2.1 Objetivo General.....	17
2.2 Objetivos particulares.....	17
3. METODOLOGÍA.....	18
3.1 Selección de genomas.....	18
3.2 Obtención de los genes rRNA 16S y las secuencias 5' intergénicas.....	18
3.3 Identificación de los terminales 3' del rRNA 16S.....	19
3.4 Cálculo de energía libre de interacción entre moléculas de RNA.....	19
3.5 Definición del consenso de las secuencias antiSD y SD de un organismo o de una clase filogenética.....	22
3.6 Análisis de la distribución filogenética de organismos con diferentes tipos de secuencias antiSD y SD.....	26
4. RESULTADOS Y DISCUSIÓN.....	27
4.1 Análisis del número de copias del gen 16S por organismo.....	28
4.2 Análisis del SD y antiSD de organismos modelo.....	29
4.3 Análisis de los tipos de antiSD por clases filogenéticas.....	30
4.4 Clasificación de los organismos según la presencia/ausencia de la secuencia AntiSD y la capacidad de sus mRNAs para interactuar con el rRNA 16S.....	39
i) Organismos con alta ocurrencia de SD y el núcleo CCUCC conservado.....	42
ii) Organismos con un CCUCC canónico conservado, pero deficientes en las interacciones rRNA 16S/mRNA.....	42
iii) Organismos que carecen del CCUCC canónico, cuyos extremos 3' de los rRNA 16S pueden interactuar de manera estable con la región líder de los mRNAs.....	45
iv) Organismos que carecen del núcleo canónico CCUCC y no muestran una interacción estable entre el 3'-terminal del rRNA 16S y el líder de los mRNAs.....	47
5. CONCLUSIONES.....	50
6. PERSPECTIVAS.....	53
7. APÉNDICE.....	54
7.1 Algunos de los scripts de perl utilizados durante el estudio:.....	54
1) Extraer los datos de las bases que superaron el límite del 40 por ciento.....	54

2) Generar histogramas por especie.....	55
3) Generar histograma para las clases.....	57
4) Analizar la salida de RNAhybrid y convertirla en una tabla de frecuencias por posición...	59
8. BIBLIOGRAFÍA.....	61

NOMENCLATURA

SD - Shine-Dalgarno

antiSD - Anti Shine-Dalgarno

RNA - Ácido ribonucleico

mRNA - RNA mensajero

rRNA - RNA ribosomal

tRNA - RNA de transferencia

cDNA - DNA complementario

siRNAs - RNA pequeño de interferencia

miRNA - MicroRNA

pb - Pares de bases

RBS - (Ribosome Binding Site) Sitio de unión al ribosoma

MAST - Motif Alignment and Search Tool

MEME - Multiple Expectation Maximization for Motif Elicitation

CDS - Secuencia codificante

UTR - Región no traducida del RNA mensajero

ΔG - Cambio en la energía libre de Gibbs

RPS1 - Proteína ribosomal S1

EM - Esperanza-maximización

IF - Factores de iniciación

CCUCC - Core canónico del Anti Shine-Dalgarno

nota: Todas las secuencias que aparecen en el texto tienen dirección 5' -> 3'. Por simplicidad se omite esta nomenclatura

RESUMEN

El inicio de la traducción en procariotas está principalmente definido, aunque no exclusivamente, por la interacción entre la secuencia anti-Shine-Dalgarno (antiSD), ubicada en el extremo 3' del RNA ribosomal 16S, y una secuencia complementaria, el sitio de unión del ribosoma o Shine-Dalgarno (SD), ubicado río arriba del codón de inicio en los mRNA. El antiSD tiene un núcleo altamente conservado 5'-CCUCC-3', pero se han encontrado variaciones entre especies con respecto a la participación de las bases adyacentes en la unión. Estas variaciones se han descrito en ciertas bacterias y, en menor medida, en algunas arqueas. Para analizar estas variaciones en nuestro estudio realizamos un análisis estadístico de las interacciones termodinámicas entre la región 5' UTR del mRNA y las regiones del extremo 3' del rRNA 16S de más de 6400 organismos, con el objetivo de identificar las secuencias antiSD y SD más probables en una amplia gama de procariontes, incluyendo un gran número de arqueas.

En nuestro estudio identificamos 15 grupos de variantes de antiSD que se asocian significativamente con el origen filogenético de los organismos. Además, nuestros hallazgos revelaron que ciertos organismos exhiben variaciones en el propio núcleo del antiSD, destacando que un núcleo inalterado no es necesariamente requerido para la interacción entre el extremo 3' del rRNA y la región que precede al AUG del mRNA. En nuestro estudio también clasificamos a los organismos en cuatro categorías distintas dependiendo de la conservación del núcleo y la evidencia de interacción : i) aquellos que poseen un núcleo conservado y muestran evidencia de unión; ii) aquellos con un núcleo conservado pero sin evidencia de unión; iii) aquellos que exhiben unión en ausencia de un núcleo conservado; y iv) aquellos que carecen tanto de un núcleo conservado como de evidencia de unión. Nuestros resultados indican que, independientemente de la conservación de la secuencia central 5'-CCUCC-3' (CCUCC) presente en el extremo 3' del rRNA 16S, la posición y secuencia de la antiSD pueden variar dependiendo de los orígenes filogenéticos de los organismos. Estos hallazgos se discuten en términos de la evolución de la iniciación de la traducción en organismos procariotas.

ABSTRACT

The initiation of translation in prokaryotes is primarily defined, although not exclusively, by the interaction between the anti-Shine-Dalgarno sequence (antiSD), located at the 3'-terminus of the 16S ribosomal RNA, and a complementary sequence, the ribosome binding site, or Shine-Dalgarno (SD), located upstream of the start codon in mRNA. The antiSD has a highly conserved core 5'-CCUCC-3', but variations have been found between species regarding the involvement of adjacent bases in binding. These variations have been described in certain bacteria and, to a lesser extent, in some archaea. To analyze these variations in our study, we conducted a statistical analysis of the thermodynamic interactions between the 5' UTR region of mRNA and the 3' end regions of 16S rRNA from 6400 organisms, with the aim of identifying the most probable antiSD and SD sequences across a wide range of prokaryotes, including a large number of archaea.

In our study, we identified 15 groups of antiSD variants that are significantly associated with the phylogenetic origin of organisms. Furthermore, our findings revealed that certain organisms exhibit variations in the core of antiSD itself. Importantly, an unaltered core is not necessarily required for the interaction between the 3' end of rRNA and the region preceding the AUG of mRNA. In our study, we also classified organisms into four distinct categories depending on the conservation of the core and evidence of interaction: i) those possessing a conserved core and showing evidence of binding; ii) those with a conserved core but no evidence of binding; iii) those exhibiting binding in the absence of a conserved core; and iv) those lacking both a conserved core and evidence of binding. Our results indicate that, regardless of the conservation of the central sequence 5'-CCUCC-3' (CCUCC) present at the 3' end of 16S rRNA, the position and sequence of antiSD can vary depending on the phylogenetic origins of organisms. These findings are discussed in terms of the evolution of translation initiation in prokaryotic organisms.

1. INTRODUCCIÓN

La traducción es el proceso biológico que permite la síntesis de proteínas. A diferencia de las células eucariontas, donde la transcripción ocurre en el núcleo y la traducción tiene lugar en el citoplasma, en los procariotes ambos procesos ocurren acoplados en el citoplasma. La traducción se lleva a cabo en los ribosomas (Green & Noller, 1997), que se clasifican según su velocidad de sedimentación, en la medida de unidades Svedbergs. Los ribosomas bacterianos tienen una sedimentación de aproximadamente 70S y están compuestos por dos subunidades: una pequeña de 30S y una subunidad grande, 50S.

La síntesis de una proteína ocurre de manera secuencial, agregando los aminoácidos en dirección N-terminal a C-terminal. Este proceso consta de tres etapas: iniciación, elongación y terminación. En la etapa de iniciación, se forma un complejo que incluye la subunidad menor del ribosoma, un RNA de transferencia (tRNA) con el anticodón UAC, el aminoácido metionina químicamente modificado y la ayuda de tres factores de iniciación (IF1, IF2 y IF3). Esta etapa de inicio de la traducción es crucial para regular la síntesis de proteínas.

Casi cinco décadas después de la publicación del artículo de J. Shine y L. Dalgarno "Determinant of cistron specificity in bacterial ribosomes" (Shine & Dalgarno, 1975), se reconoce que en la mayoría de las bacterias, el proceso de iniciación de la traducción implica la estabilización del complejo de preiniciación mRNA-30S mediante el reconocimiento específico de una secuencia en el mRNA ubicada de 5 a 8 nucleótidos del codón de inicio del gen a traducir, con una secuencia complementaria en el 3'-terminal del rRNA 16S de la subunidad 30S del ribosoma (Fig. 1). Esta interacción es crucial para definir de manera inequívoca el sitio de inicio de la traducción en la mayoría de los organismos procarióticos. En reconocimiento a la contribución científica de estos autores, estas secuencias se conocen como Shine-Dalgarno (SD) y anti-Shine-Dalgarno (antiSD), respectivamente. La similitud de estas secuencias observadas inicialmente en un pequeño conjunto de organismos modelo llevó a la idea de que las secuencias antiSD y SD eran, en términos generales, universales (Fig. 2). Estudios más recientes han revelado la existencia de vías alternativas independientes de las

interacciones SD/antiSD, sin embargo la interacción entre la secuencia SD en el extremo 5' UTR del mRNA y la secuencia antiSD en el extremo 3' del RNA ribosomal 16S, se conserva de manera significativa en más del 80% de los procariontes.

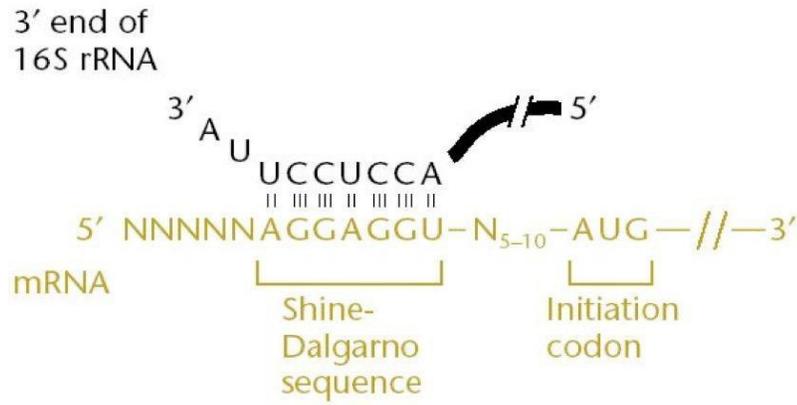


Fig. 1. Extremo 3' de rRNA 16S hibridando con la secuencia Shine-Dalgarno del mRNA.

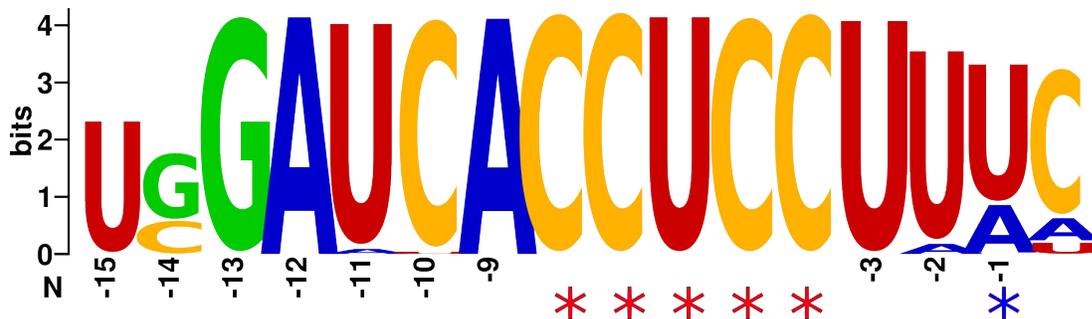


Fig. 2. Logo de la conservación de la secuencia del extremo 3' del 16S en organismos procariontes. El contenido informacional de la conservación está determinado en bits, siendo el valor de 4, el máximo posible de una posición 100% conservada. La posición -1 indica el final del motivo más conservado.

El inicio de la traducción de un mRNA depende de varios elementos de la estructura primaria y secundaria del RNA. Estos incluyen el tipo de codón de inicio (Srivastava et al., 2016), la ausencia de estructuras secundarias de RNA (Scharff et al., 2011), la presencia de una proteína ribosomal S1 (Boni et al., 1991)(Nakagawa et al., 2010), la presencia de una región líder 5' no traducida en el mRNA (Moll et al., 2002) y un sitio de unión del ribosoma o la secuencia SD (Shine & Dalgarno, 1975). Cuando está

presente, el SD generalmente se encuentra aproximadamente 8 bases antes del codón de inicio (AUG). Aunque las secuencias SD suelen ser variables, son parcialmente complementarias a la antiSD en el rRNA 16S de la especie. Por otro lado, la secuencia antiSD es muy similar entre las especies. Por lo general, se encuentran algunos nucleótidos antes del extremo 3' del rRNA, superponiéndose con un motivo CCUCC muy conservado (el "núcleo" del antiSD). La interacción generalmente implica de 5 a 8 bases, con 3 o más pertenecientes al núcleo. Se sabe que las interacciones más fuertes permiten tasas más altas de inicio de la traducción. La interacción SD/ antiSD es vital en la mayoría de los organismos procariontes para establecer el sitio correcto de inicio de la traducción (Wen et al., 2021).

Sin embargo, el filo Bacteroidota, generalmente no utiliza este proceso (Nakagawa et al., 2010; Wen et al., 2021)(Accetto & Avguštin, 2011)(Nakagawa et al., 2017). En su lugar, utilizan otros elementos del mRNA para la iniciación, incluyendo un conjunto específico de adeninas río arriba del AUG (Baez et al., 2019). A pesar de que los Bacteroidotas tienen un antiSD conservado, sus ribosomas no reconocen secuencias SD (Jha et al., 2021).

Un estudio de McNutt et al. (McNutt et al., 2021) mostró que el papel del antiSD en *Flavobacterium johnsoniae* (*F. johnsoniae*), un miembro de los Bacteroidotas, es significativamente menos relevante que en *Escherichia coli* (*E. coli*), de acuerdo con el uso de SD en ambos organismos. En los ribosomas de *F. johnsoniae*, la secuencia antiSD queda secuestrada por la interacción del extremo 3' del 16S rRNA con las proteínas (bS21, bS18 y bS6). Esto sugiere un mecanismo común en todo el filo Bacteroidota. Curiosamente, si se elimina la proteína bS21, la traducción de genes reporteros con un SD aumenta en *F. johnsoniae* (McNutt et al., 2023). El gen *rpsU* en *F. johnsoniae* y bacterias relacionadas, que codifica bS21, contiene una secuencia SD fuerte. Esto indica que la secuencia SD podría utilizarse como un elemento regulador, permitiendo a estas proteínas controlar su propia traducción. Un modelo propuesto sugiere que a medida que aumentan los niveles de bS21, también aumentarán los ribosomas completos, reduciendo en consecuencia la producción de bS21, creando un mecanismo de

autorregulación (Jha et al., 2021). Esto se correlaciona con investigaciones que muestran que la síntesis de proteínas bS21 depende de una subpoblación de ribosomas que carecen de bS21 (McNutt et al., 2023).

Al igual que las bacterias, muchos genes arqueales utilizan el mecanismo SD/antiSD. Sin embargo, también existen mRNA arqueales que no tienen región 5' UTR, es decir son mRNA "sin líder", con una región no traducida en el extremo 5' de menos de 8 nucleótidos (Betlach et al., 1984). Estos mRNA no tienen una secuencia SD, por lo que el ribosoma reconoce directamente el codón de inicio (AUG) (Brock et al., 2008). Anteriormente se pensaba que los arqueas dependían de la interacción RNA/RNA para la traducción, pero estudios recientes sugieren que las arqueas pueden utilizar dos mecanismos de inicio diferentes: uno similar al bacteriano basado en motivos SD y otro para la traducción de mRNA sin líder (Tolstrup et al., 2000).

En las arqueas, los motivos SD son esenciales para el inicio de la traducción, y su eliminación podría detener la producción de proteínas. Sin embargo, incluso con la eliminación de la región 5' UTR, el mRNA aún puede traducirse, lo que indica un mecanismo diferente (Condò et al., 1999). Se sugiere que este mecanismo podría implicar subunidades ribosomales pre-cargadas con tRNA iniciador, similar a la iniciación eucariota. Las subunidades ribosomales pequeñas de las arqueas podrían interactuar fuertemente con mRNAs con líder si están presentes motivos SD, pero no con los "sin líder", lo que requeriría la presencia de met-tRNA_i para el reconocimiento del sitio de inicio.

En el proceso de traducción sin líder, el factor de inicio 2 (IF2) desempeña un papel crucial. En las bacterias, IF2 promueve la entrada del tRNA iniciador (f-met-tRNA) en el sitio P del ribosoma. Para la traducción sin líder, se requieren subunidades ribosomales 30S pre-cargadas con f-met-tRNA iniciador (Benelli et al., 2003). Los mRNA sin líder reconocen un codón AUG terminal 5' mediante la interacción con el anticodón del f-met-tRNA unido al ribosoma sin ninguna interacción previa mRNA/ribosoma. IF2 es fundamental para este proceso, al igual que en la iniciación eucariota. La abundancia de IF2, especialmente la relación IF2/IF3, favorece la

traducción "sin líder" sobre la basada en SD (Grill et al., 2000). Por lo tanto, el factor IF2 es vital para el proceso de iniciación en el mecanismo de traducción sin líder.

Teniendo en cuenta las diferentes vías para el proceso de inicio de la traducción en organismos procariotas y la mayor representatividad actual de organismos secuenciados, nuestro estudio tuvo como objetivo analizar las bases en el rRNA 16S y los mRNA involucrados para definir el sitio de inicio de la traducción.

La secuencia SD, que generalmente se encuentra 8 bases río arriba del codón de inicio de una secuencia codificante (CDS), se complementa con el extremo 3' del RNA ribosomal 16S. La interacción entre el SD y las secuencias antiSD facilita la formación del pre-complejo de iniciación alrededor del codón de inicio del mRNA, el cual comprende una subunidad 30S y tres factores de iniciación (IF1, IF2 e IF3). Se ha verificado experimentalmente la importancia de esta interacción para una iniciación eficiente de la traducción tanto en bacterias como en arqueas. Además, se sabe que existe una relación significativa entre la similitud de las secuencias SD de un gen con el antiSD y la eficiencia de inicio de la traducción del mismo.

Aunque se asume que la secuencia SD está conservada entre organismos bacterianos, diferentes estudios genómicos han demostrado que la presencia de la secuencia SD consenso en los mRNAs de un organismo varía significativamente. Se ha reportado que una secuencia SD de mayor longitud no necesariamente produce un aumento en la traducción, ya que esto puede deberse a la presencia de estructuras secundarias adicionales o a una complementariedad muy fuerte que prolonga el tiempo necesario para el despeje del sitio de unión, lo que evita la formación de polisomas (Komarova et al., 2001).

Para predecir la hibridación entre dos moléculas de RNA, se suele utilizar el cálculo del cambio en la energía libre de Gibbs (ΔG , en kilocalorías por mol). Esto permite identificar y cuantificar diversos aspectos, como los posibles sitios de unión de estructuras secundarias para una subsecuencia dentro de un mRNA, los blancos para secuencias siRNAs y miRNAs, o la secuencia SD favorable al evaluar el ΔG entre la

secuencia antiSD y las bases antes del codón de inicio de un gen, que es donde comúnmente se encuentra la secuencia SD. Los valores de ΔG para un SD funcional puede variar de acuerdo al programa utilizado, encontrándose entre -3.5 kcal/mol y -9 kcal/mol. Un SD se clasifica como débil si está por encima de este rango, y como fuerte si su valor está por debajo.

1.1 Mecanismos alternativos de traducción en procariontes.

A pesar de que el mecanismo de traducción en organismos procariontes y arqueas se basa principalmente en la interacción entre el rRNA 16S y el SD, se conocen mecanismos alternativos que no requieren de la secuencia SD. Estos mecanismos incluyen:

1- Mecanismo mediado por la proteína ribosomal S1 (RPS1). La proteína ribosomal S1 permite el inicio de la traducción de manera independiente de la secuencia SD.

2- Mecanismo de mRNAs que no poseen una región líder en el extremo 5', donde la interacción entre el rRNA 16S y el SD no se puede llevar a cabo.

3- Mecanismo dependiente de la estructura secundaria del mRNA.

1.1.1 Mecanismo mediado por una proteína ribosomal S1

La proteína ribosomal S1 (RPS1) es componente de la subunidad 30S ribosomal que se conoce por su alta afinidad por RNAs de cadena-sencilla y zonas ricas en pirimidinas (A/U) (Boni et al., 1991). Por ejemplo, en *E. coli*, la proteína RPS1 y el factor de iniciación IF3 interactúan con la región 5' UTR del mRNA, iniciando eficientemente la traducción a pesar de la ausencia de secuencias SD en alguno de estos (Tzareva et al., 1994). IF3, en este contexto, está potencialmente involucrada en la selección de los codones de inicio correctos en el sitio P (Hussain et al., 2016).

La proteína RPS1 le confiere a la subunidad 30S ribosomal la actividad de chaperona de RNA que es esencial para el acoplamiento (“docking”) y desnaturalización de mRNA estructurados, y para posicionar correctamente al codón

de inicio dentro del canal de traducción (Duval et al., 2013), confiriéndole al ribosoma las condiciones necesarias para iniciar la traducción de los mRNA a pesar de sus estructura secundaria.

1.1.2 Mecanismo utilizado para la traducción de mRNA sin líder

Se conoce que algunos mRNA no poseen secuencia líder. En estos casos el mRNA puede ser traducido mediante su unión directa con la subunidad pequeña del ribosoma 70S que incluirá un tRNA con el residuo de N-formilmetionina (fMet). En algunas arqueas, los mRNAs sin líderes tienen 15 veces mayor actividad en la traducción de mensajeros en relación a los mRNAs con presencia de secuencia SD.

Como hemos dicho, el inicio de la traducción en bacterias generalmente ocurre mediante la interacción del rRNA 16S y el extremo 5' UTR del mRNA (Fig. 3A) que junto con los factores de iniciación (IF1, IF2 y IF3) y el aminoacil-tRNA fMet-tRNA_f forman el complejo de iniciación 30S de tal forma que el codón de iniciación queda ubicado dentro de la parte del sitio P de la subunidad menor del ribosoma. Se ha observado que un aumento en la concentración del factor IF2 estimula la capacidad de reclutamiento del fMet-tRNA_f (tRNA-iniciador) y por tanto el pegado de este al sitio P (Grill et al., 2000). Estos resultados sugieren que el reconocimiento 30S-mRNA (sin líder) se lleva a cabo por un complejo 30S-IF2-tRNA-iniciador (Fig. 3B) (Moll et al., 2002).

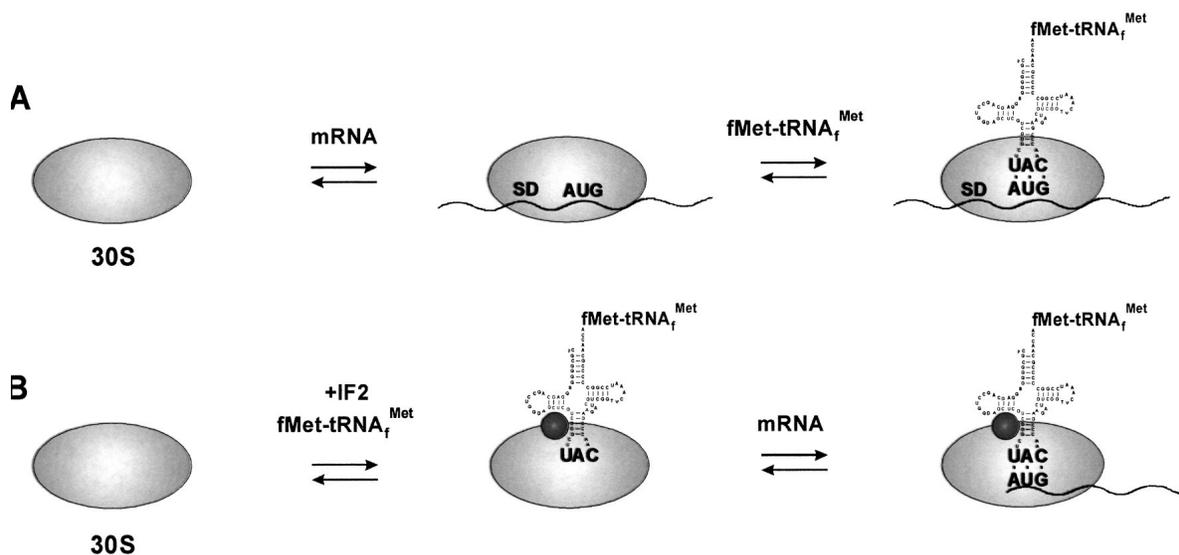


Fig. 3. Vías de inicio de la traducción en procariotes. (A) Reclutamiento de un ribosoma procariótico por un mRNA que contiene una región de inicio de traducción canónica a través la interacción SD/antiSD. (B) Reconocimiento del codón de inicio de un mRNA sin líder por un complejo del 30S y el tRNA iniciador. Figura tomada de (Moll et al., 2002).

1.1.3 Mecanismo dependiente de la estructura secundaria del mRNA

Se conoce que la secuencia que rodea el codón de inicio posee una estructura secundaria significativamente menos estable en los mRNAs sin SD en comparación con la estructura presente en aquellos con SD (Nakagawa et al., 2017; Scharff et al., 2011), mostrando además esta región un sesgo en la frecuencia de nucleótidos, lo que potencialmente desfavorece el emparejamiento de bases y el plegamiento local del mRNA, sugiriendo que la falta de estructura secundaria es una característica universal para guiar la iniciación de la traducción del mRNA sin SD en el sitio de inicio correcto. Nuestro conocimiento actual es insuficiente para explicar cómo los ribosomas distinguen eficientemente el codón de inicio correcto de los codones AUG internos en los mRNA sin SD, más allá de la pista proporcionada por la estructura secundaria (Scharff et al., 2011).

En términos generales, las observaciones iniciales de los elementos que intervienen en el inicio de la traducción en procariotes, se describieron a partir de observaciones en organismos modelo, como *E. coli* y *Bacillus subtilis* (*B. subtilis*). Gracias a los avances significativos en las nuevas tecnologías de secuenciación, actualmente existen miles de genomas secuenciados en su totalidad, por lo que es posible realizar un estudio global que los incluya.

2. OBJETIVOS

2.1 Objetivo General

Identificar los tipos de secuencias utilizados en procariontes como sitio de unión al ribosoma (RBS) y su secuencia complementaria en el rRNA 16S conocida como antiSD.

2.2 Objetivos particulares

1. Identificación y análisis de las secuencias desestructuradas o libres en el extremo 3' de los genes rRNA 16S.
2. Análisis de la energía libre (ΔG) de interacción entre las secuencias SD de los mRNAs de un organismo con la región antiSD de sus correspondientes RNAs ribosomales 16S y predecir el número de genes con posible uso de esta interacción para el inicio de la traducción.
3. Identificar organismos con ausencia del núcleo o firma canónica CCUCC.
4. Definir la distribución filogenética de las características de la secuencia SD y antiSD tomando en cuenta el tamaño de la región de interacción, el grado de conservación de esta y su posición en relación al extremo 3' del gen 16S ribosomal y la región 5' UTR de los mRNAs.

3. METODOLOGÍA

3.1 Selección de genomas

De los 7,272 genomas de organismos procariotas en la base de datos de KEGG hasta mayo de 2022, se seleccionaron 6,457 organismos que, según sus anotaciones genómicas, tenían el mayor número de secuencias codificadoras de proteínas (CDS, por sus siglas en inglés) y al menos un rRNA 16S con una región terminal 3' completa. Estos organismos se eligieron como representantes de una especie y representan un total de 87 clases filogenéticas que se consideraron en nuestro estudio.

3.2 Obtención de los genes rRNA 16S y las secuencias 5' intergénicas.

Las secuencias nucleotídicas de los rRNA 16S se obtuvieron siguiendo los siguientes pasos: a) Generamos motivos de secuencia representativo para los rRNA 16S procariotas, que incluye la región antiSD y secuencias conservadas adyacentes. Para ello, seleccionamos aleatoriamente 280 secuencias de rRNA 16S de organismos bacterianos representativos y 33 secuencias de rRNA 16S de organismos arqueas representativos utilizados en nuestro estudio. b) Estas secuencias, de aproximadamente 1,500 pares de bases de longitud, se analizaron utilizando el programa MEME (Bailey et al., 2006, 2009) para identificar motivos de secuencia conservados y sus correspondientes matrices MEME, utilizando los parámetros "-motifs" establecido en 50, "-evt" establecido en 1e-50 y "-w" establecido en 60 (Fig 4). c) La matriz MEME correspondiente a todas las regiones no variables del rRNA 16S, que comúnmente incluye el CCUCC de la antiSD en el último motivo, se utilizó con el programa MAST (Bailey et al., 2009) para diferenciar qué genes ribosomales 16S de los organismos estudiados contenían la región antiSD y cuáles no. Los parámetros utilizados en nuestro programa MAST fueron "-ev" establecido en 1e-100, "-mt" establecido en 1e-08 y "-best". Mediante esta estrategia, se desarrolló Ribbon (Estrada & Merino, 2020), un programa de acceso libre que se encuentra disponible para su acceso a través del repositorio de GitHub. Las regiones intergénicas 5' de los mRNAs se obtuvieron considerando únicamente los genes codificadores de proteínas, según lo anotado en la base de datos KEGG.

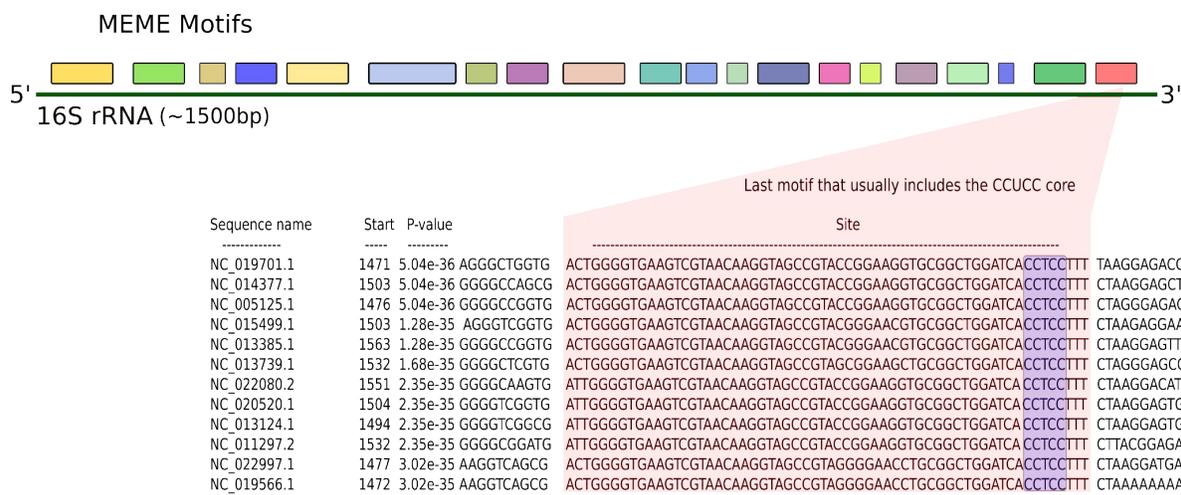


Fig. 4. Motivos de secuencia conservados en los 16S rRNA. El último motivo incluye el núcleo CCUCC del antiSD. Nótese que se utilizaron las secuencias del cDNA de los rRNA para la búsqueda de motivos, lo que es equivalente para este análisis.

3.3 Identificación de los terminales 3' del rRNA 16S

Con base en la identificación de los rRNA 16S utilizando nuestro análisis MEME/MAST previamente descrito, se identificó la ausencia de secuencias antiSD canónico en algunos genes rRNA 16S al examinar cuidadosamente el alineamiento múltiple de las últimas 25 bases de secuencias generadas. Además, para los organismos que poseen más de un gen de rRNA 16S, verificamos la presencia del CCUCC en todas sus copias parálogas.

3.4 Cálculo de energía libre de interacción entre moléculas de RNA.

Para evaluar la energía libre de hibridación entre las secuencias antiSD y SD, primero realizamos un análisis de la estructura secundaria del 3'-terminal de los rRNA 16S para identificar la región no estructurada capaz de interactuar con la secuencia SD del mRNA. Según nuestro estudio *in silico*, esta región no estructurada del 3'-terminal de los rRNA 16S suele tener una longitud de 13 a 16 bases (Fig 5).

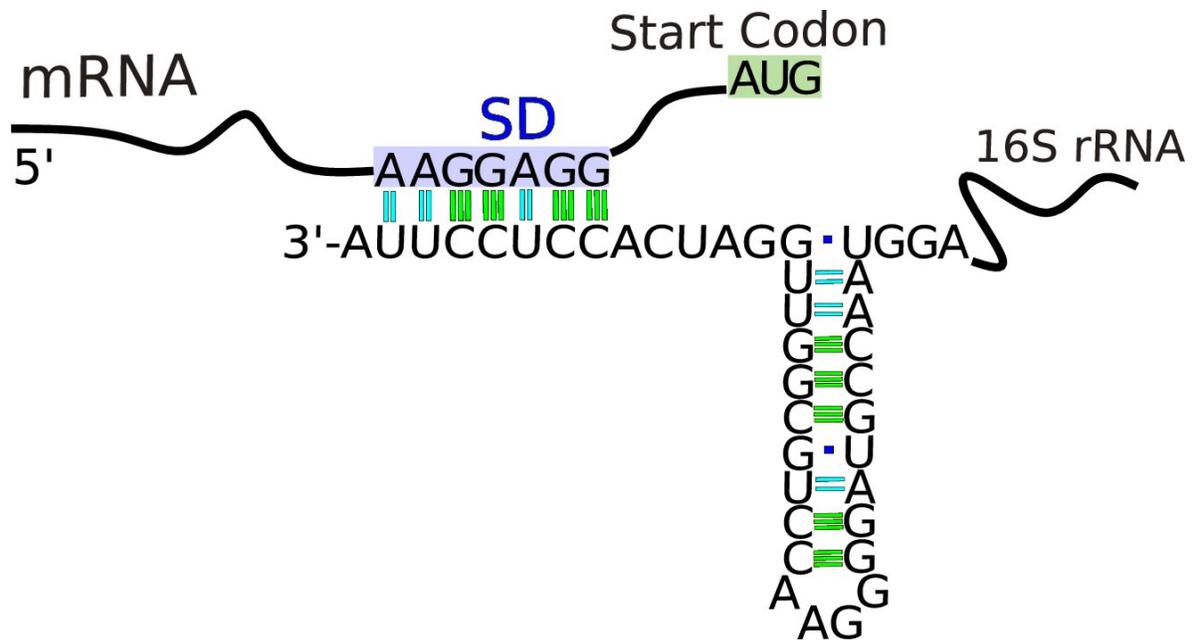


Fig. 5. Se presenta la estructura secundaria de los últimos 40 nucleótidos del rRNA 16S de *E. coli*. La secuencia SD y el codón de inicio están coloreados en morado y verde, respectivamente. Los últimos 13 nucleótidos del rRNA 16S son no estructurados y corresponden a la región más probable para interactuar con la secuencia SD en los mRNAs.

Por otro lado, para analizar las interacciones antiSD/SD, consideramos la región de 18 nucleótidos ubicada entre las bases del intervalo [-20, -2) río arriba del codón de inicio de los mRNAs de nuestros organismos. El cálculo de la energía libre de hibridación se realizó utilizando el software RNAhybrid (Rehmsmeier et al., 2004) con los parámetros por defecto. RNAhybrid ignora las estructuras intramoleculares y busca todas las estructuras de múltiples bucles para encontrar regiones de interacción termodinámicamente favorables, ofreciendo un detallado resultado de los nucleótidos que interactúan (Fig 6). Para minimizar la probabilidad de obtener resultados incorrectos, RNAhybrid se corrió en todas las subsecuencias de 10 nucleótidos del intervalo de 18 bases del UTR, utilizando una ventana deslizante, quedándonos con la interacción más favorable (menor valor ΔG) de las 9 ventanas. Para cada mRNA analizado, se registraron las posiciones y las bases que interactúan en las secuencias antiSD y SD para evaluar sus frecuencias relativas. Estas frecuencias se estimaron considerando: a) todos los mRNAs de un organismo; y b) todos los mRNAs de todos los organismos de una clase filogenética particular. Para una mejor visualización y análisis,

las frecuencias de los nucleótidos que interactúan en una posición dada en las secuencias antiSD y SD se representaron como histogramas.

mfe: **-13.2** kcal/mol

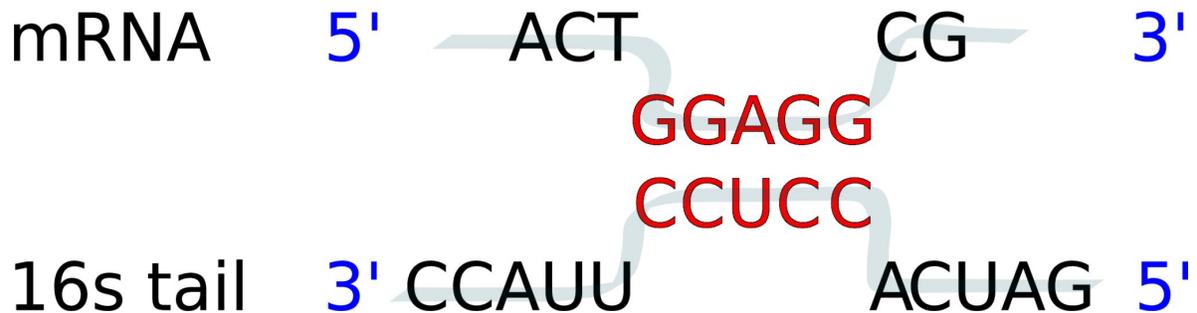


Fig. 6. Ejemplo de la salida del programa RNAhybrid. Las bases que interactúan en las secuencias SD y antiSD están resaltadas en rojo. Observa que -13.2 kcal/mol corresponde al valor más significativo para una hibridación perfecta del núcleo canónico de las secuencias SD y antiSD.

Adicional a la longitud de las secuencias que hibridan, otro factor que puede influir en los valores de la energía libre de Gibbs es el contenido de GC genómico, esto se debe a la formación de puentes de hidrógeno, siendo esta mayor para las bases GC. Entre las bases guanina (G) y citosina (C) se forman tres puentes de hidrógeno, mientras que para las bases adenina (A) y uracilo (U) solo dos, así es que en una secuencia con un alto contenido de GC, puede haber más oportunidades para que se formen puentes de hidrógeno, lo que resultará en un ΔG más negativo. Esto sugiere que genomas con alto contenido de GC pueden tener una mayor afinidad o estabilidad en sus interacciones de ARN sin embargo, nuestro estudio observó una correlación muy débil entre el contenido de GC y el ΔG en todos los genomas analizados (Fig 7), lo que sugiere que ajustar el umbral de SD basado en el contenido de GC podría tener un impacto insignificante.

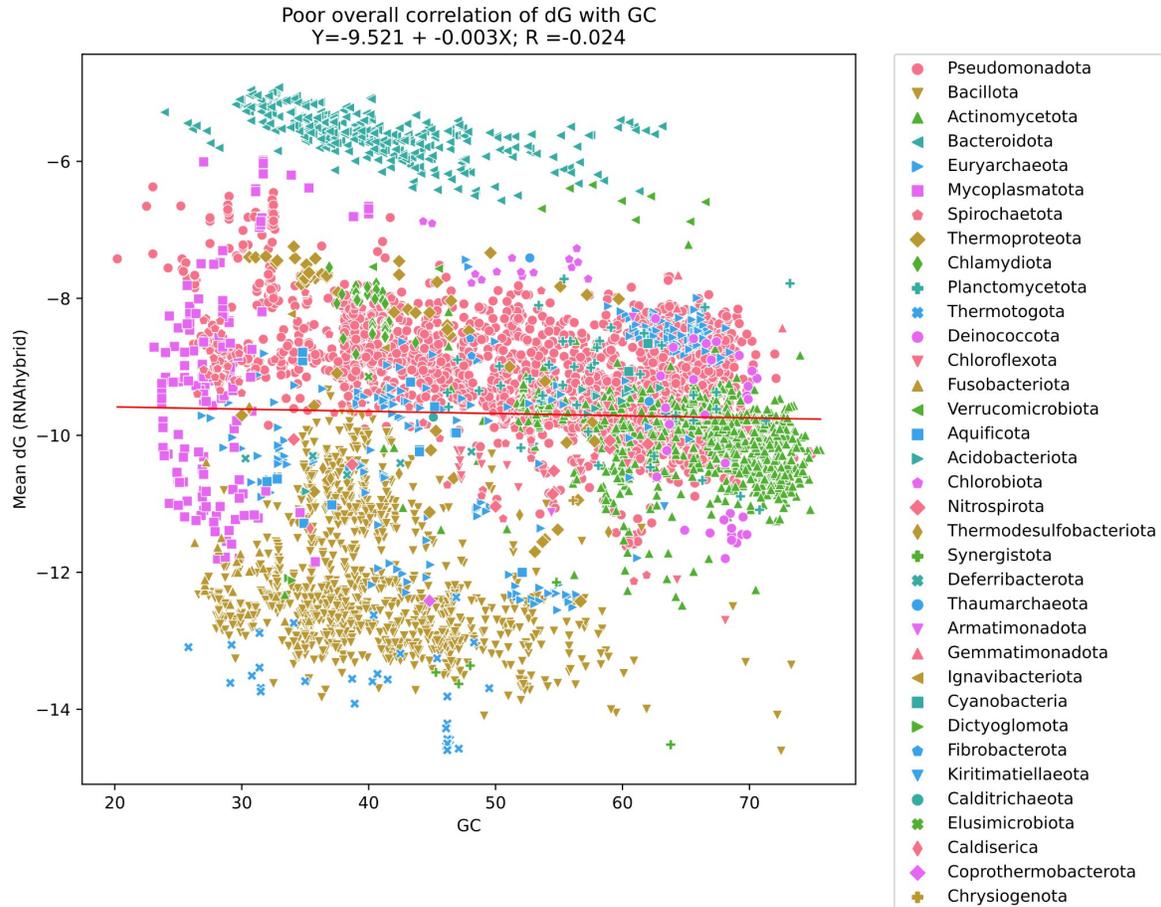


Fig 7. Correlación entre la ΔG media y el contenido de GC genómico.

Cuando examinamos cada filo individualmente, a veces observamos una correlación más alta. Por ejemplo, en el caso de Pseudomonadota, nuestro filo más representado, hubo una correlación moderadamente negativa. Sin embargo, este patrón no fue consistente en todos los filios. Muchos filios en nuestro conjunto de datos estaban representados por un número limitado de genomas y mostraban un rango estrecho de contenido de GC. En consecuencia, cualquier corrección basada en el contenido de GC solo se aplicaría a un número pequeño, los filios más grandes, y dichos ajustes selectivos podrían potencialmente introducir sesgos en nuestro análisis.

3.5 Definición del consenso de las secuencias antiSD y SD de un organismo o de una clase filogenética

Para determinar las bases de nucleótidos de la secuencia que define el sitio de unión del ribosoma, o secuencia SD de un organismo, para cada CDS del genoma

secuenciado, identificamos la región más probable del 3'-terminal no estructurado del rRNA 16S que puede interactuar con la región 5' inmediatamente antes del codón de inicio de los mRNAs. Luego, registramos solo las bases que interactúan de manera termodinámicamente favorable entre el rRNA 16S y el mRNA para evaluar sus frecuencias relativas y representarlas como histogramas. Consideramos que este procedimiento es más preciso que aquellos que solo se basan en las frecuencias relativas de las bases antes del codón de inicio de la traducción.

Para identificar las interacciones estables entre el rRNA 16S y el mRNA que probablemente actúen como secuencias antiSD y SD funcionales, aplicamos un valor de corte de -8.4 kcal/mol como umbral de energía de unión válido. Para obtener nuestra línea de corte ($\Delta G < -8.4$ kcal/mol) seguimos dos análisis. Primero realizamos un procedimiento similar al descrito en (Starmer et al., 2006), este umbral de ΔG se determinó como el ΔG promedio de los 4 tetrámeros que pueden coincidir con el núcleo conservado CCUCCUU del 16S: AAGG (-7.1 kcal/mol), AGGA (-8.0 kcal/mol), GGAG (-9.3 kcal/mol) y GAGG (-9.1 kcal/mol), según los valores obtenidos por RNAhybrid. Segundo, utilizando UTRs "falsos" para medir la frecuencia de interacciones verdaderas y falsas, este análisis mostró que la línea de corte concuerda perfectamente con el valor de ΔG en el que la frecuencia de interacciones "verdaderas" de SD es igual a la frecuencia de interacciones aleatorias (Fig 8), donde los valores a la derecha incluirían más falsos positivos que verdaderos positivos y los valores a la izquierda excluirían más verdaderos positivos que falsos positivos.

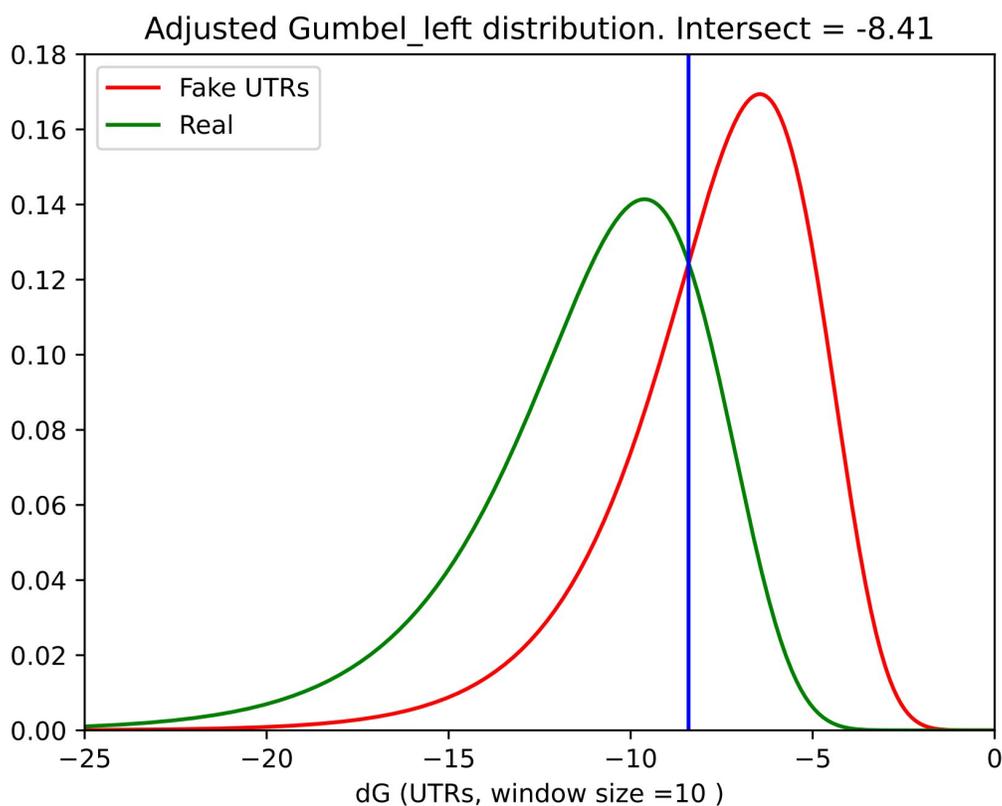


Fig 8. Densidad de los valores de ΔG de UTRs Reales y UTRs "Falsos". Ambas colecciones contenían 10,000 muestras obtenidas de todos los genomas. Para enriquecer los verdaderos SDs, la colección Real se restringió a coincidencias con al menos 3 nucleótidos en el núcleo antiSD y una distancia desde el CCUCC hasta el codón de inicio de entre 8 y 11 nucleótidos. Ambas curvas se intersectan en $\Delta G = -8.4$.

Todas las bases en el 3'-terminal del rRNA 16S que se presentaron en al menos el 40% de las interacciones rRNA/mRNA predichas con una energía libre inferior a nuestro valor de corte de -8.4 kcal/mol se consideraron parte del consenso de la secuencia antiSD. Por otro lado, la secuencia complementaria a este consenso antiSD se consideró la secuencia SD de consenso del organismo. Nuestro estudio contempla que la interacción entre una secuencia líder de un mRNA y el 3'-terminal del rRNA 16S de un organismo contribuye a definir el sitio de inicio de la traducción si dicha interacción tiene una energía libre inferior a nuestro valor de corte de -8.4 kcal/mol y si la base más frecuente del consenso antiSD/SD definida explícitamente para el organismo mencionado se incluye en la interacción. Además, solo creamos perfiles de antiSD para aquellos organismos cuya frecuencia de SD estaba por encima del 5.5% en relación con

el número total de mRNAs. Los organismos con un porcentaje de SD más bajo se consideraron "Escasos en SD": organismos deficientes de interacción SD/antiSD, es decir que probablemente no utilizan este mecanismo como forma de inicio de traducción. Evaluando el impacto de varios umbrales de porcentaje en la clasificación de los organismos que utilizan el mecanismo SD/antiSD. Descubrimos que, exclusivamente dentro del rango del 4-7% (Fig 9), cualquier umbral elegido resultaba en clasificaciones casi idénticas, lo que indica un grado de estabilidad, en consecuencia, consideramos que el umbral del 5.5% era un valor razonable para nuestro estudio.

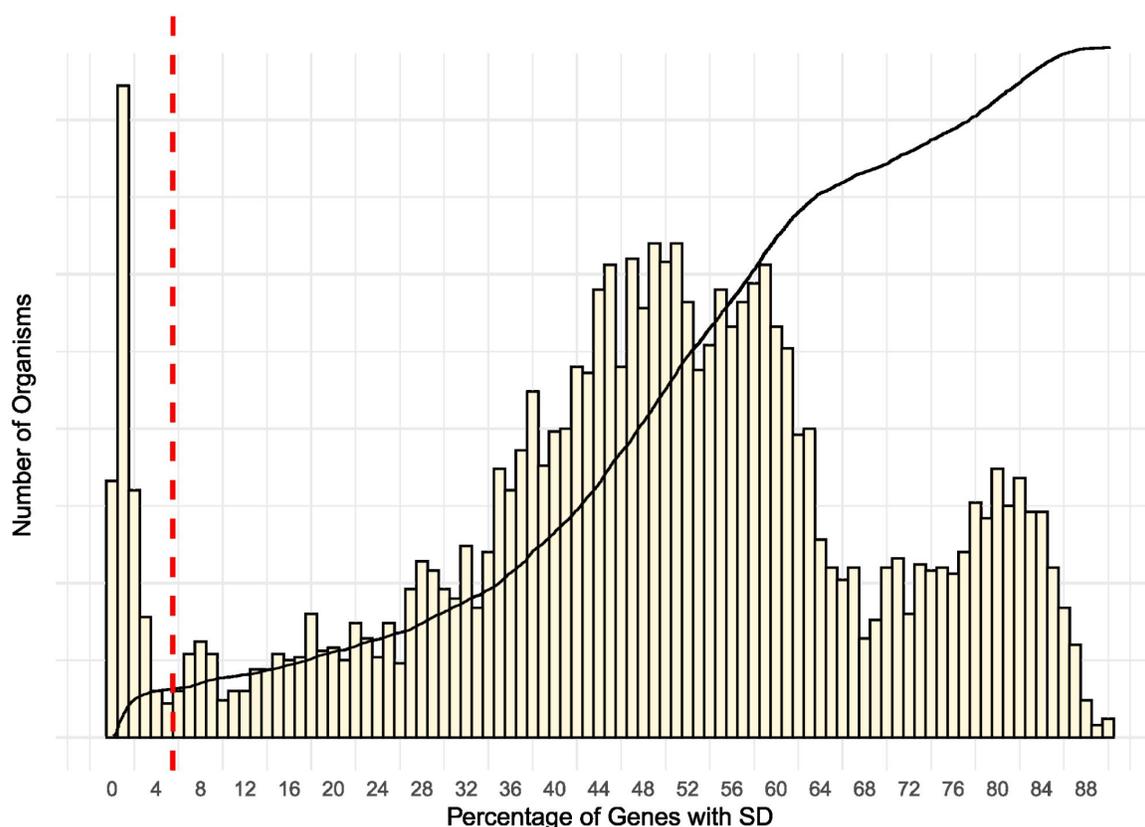


Fig 9. Distribución del promedio de genes que superan el umbral de -8.4 kcal/mol para la interacción SD/antiSD para los organismos estudiados. La línea discontinua en rojo indica el corte de 5.5 seleccionado.

Para determinar las secuencias consenso correspondientes a las secuencias antiSD y SD de una clase filogenética en particular, realizamos el mismo procedimiento descrito para los organismos, considerando todas las interacciones funcionales predichas entre el rRNA 16S y el mRNA de todos los organismos dentro de la clase

estudiada.

3.6 Análisis de la distribución filogenética de organismos con diferentes tipos de secuencias antiSD y SD

Para analizar la distribución filogenética de los organismos con diferentes características en sus secuencias SD y antiSD, construimos un árbol filogenético que incluye (1) una muestra aleatoria de 252 organismos de la lista de organismos representativos considerados en nuestro estudio, y (2) 71 organismos representativos que carecen de la secuencia antiSD en su gen rRNA 16S. Posteriormente, se utilizó el programa Mafft y SSU-align (v0.1.1) (Nawrocki, 2009) para alinear las secuencias de rRNA 16S, utilizando el parámetro "--key-out default" que permite utilizar máscaras precalculadas en el rRNA 16S de bacterias y arqueas para descartar las posiciones no informativas de los alineamientos múltiples.

El árbol filogenético se realizó utilizando el programa Jmodeltest (v2.1.10) (Posada, 2008), con los parámetros "-i -f -g 4 -t ML -AIC -BIC" y con el modelo sugerido por él según los criterios de Akaike y Bayesiano, utilizamos el programa Phyml (v20120412) con los siguientes parámetros "-d nt -n 1 -b 0 --run_id GTR + I + G -m 012345 -fm -ve -c 4 -ae - -no_memory_check -o tlr -s BEST". Finalmente, para la visualización y edición del árbol se utilizó iTOL (Letunic & Bork, 2016).

4. RESULTADOS Y DISCUSIÓN

Teniendo en cuenta las diferentes vías para el proceso de iniciación de la traducción en procariontes y gran representatividad actual de organismos secuenciados, nuestro estudio tuvo como objetivo analizar los posibles elementos de secuencia en el rRNA 16S y los mRNA, involucrados en la definición del sitio de inicio de la traducción. Aunque la comparación de las secuencias SD y antiSD entre un conjunto de organismos modelo se ha documentado (Abolbaghaei et al., 2017; Accetto & Avguštin, 2011; Wei et al., 2017), en la muchos de estos estudios, las secuencias SD de los organismos se han definido en función de la frecuencia de nucleótidos río arriba del codón de inicio AUG del mRNA y, posteriormente, se ha verificado la existencia de una secuencia complementaria en el 3'-terminal del rRNA 16S sin considerar las interacciones termodinámicas reales de cada una de las bases que definen tales interacciones. Teniendo en cuenta lo anterior, el objetivo del presente estudio fue realizar un análisis estadístico de las interacciones termodinámicas entre el 3'-terminal del rRNA 16S y el 5' UTR de los mRNA para definir de manera más precisa las secuencias antiSD y SD más probables en el conjunto de secuencias genómicas actualmente disponibles y observar las características de estas secuencias específicas en las 87 clases filogenéticas que incluye nuestro trabajo.

Para identificar las bases que definen la interacción entre el rRNA 16S y sus contrapartes en la región líder de los mRNA de un organismo, inicialmente determinamos la región no estructurada del extremo 3' de los genes del rRNA 16S mediante el análisis de posibles estructuras secundarias utilizando el programa RNAfold (Gruber et al., 2008) con los parámetros predeterminados. En este análisis, consideramos los últimos 40 nucleótidos de los genes del rRNA 16S, ya que contienen la estructura secundaria más a la derecha de bucle-tallo del rRNA 16S y el extremo 3' no estructurado, que tiene el potencial de interactuar con la región no codificante 5' de los mRNA (Fig 5).

4.1 Análisis del número de copias del gen 16S por organismo

Se ha documentado que el número de copias parálogas de los genes de rRNA en los organismos bacterianos puede variar significativamente (Espejo & Plaza, 2018; Klappenbach et al., 2000). Aunque las razones de esta variación no están del todo claras, se ha propuesto que el número de copias de genes de rRNA está asociado con una respuesta adaptativa de las bacterias a sus correspondientes nichos ecológicos, disponibilidad de nutrientes y tamaños de genoma. En nuestro caso, del conjunto de 6,457 organismos, se encontró que 5,396 de ellos tenían más de una copia del gen del rRNA 16S (Fig 10); sin embargo, en más del 98% de los casos, se identificó que los últimos veinte nucleótidos de las copias parálogas del gen del rRNA 16S eran idénticos. Además, en casi todos los organismos con copias del gen del rRNA 16S, la mayoría de las copias conservan el CCUCC en su extremo 3'. Siempre que fue posible, los análisis de hibridación termodinámica se realizaron con dichos rRNA 16S.

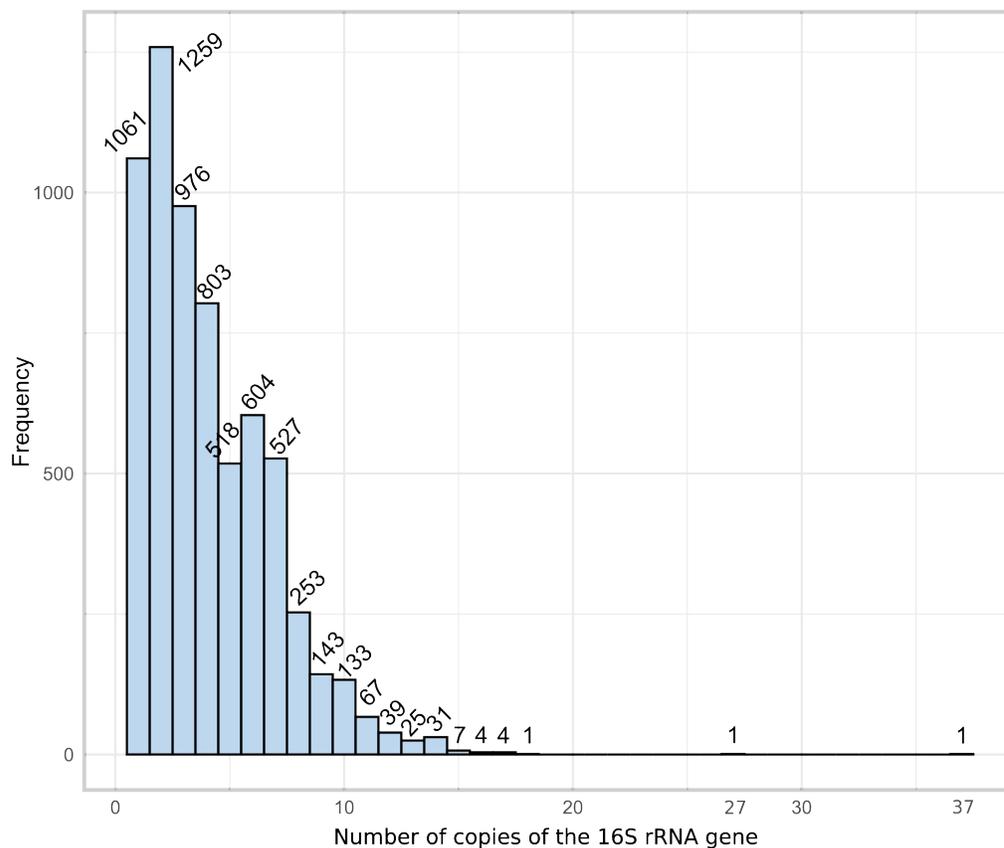


Fig 10. Distribución del número de copias del rRNA 16S en los organismos estudiados. Los genes del rRNA 16S fueron identificados en nuestro conjunto de 6,457 organismos seleccionados utilizando el programa MAST con matrices de probabilidad de posición MEME construidas considerando un conjunto de secuencias representativas de genes del rRNA 16S de bacterias y arqueas, según se describe en la sección de Materiales y Métodos.

4.2 Análisis del SD y antiSD de organismos modelo

En primera instancia, utilizamos nuestro protocolo bioinformático en los organismos modelo *E. coli* y *B. subtilis*, ya que el SD y el antiSD de estos organismos se han analizados recientemente. Nuestro estudio encontró que el 71% de los mRNA de *E. coli* tienen un SD que hibrida con el antiSD correspondiente con una energía de unión de -8.4 kcal/mol o menos. Esta frecuencia es considerablemente menor que la de *B. subtilis*, donde el 85% de los genes tienen dicha interacción. Además, observamos que las secuencias del antiSD de estos organismos, mostradas en la Fig. 10, presentaron diferencias significativas en cuanto a su longitud (7 vs. 10), secuencia (5'-CCUCCUU-3' vs. 5'-UCACCUCCU-3'), y su posición precisa a partir del extremo 3' de los rRNA 16S. En *E. coli*, la interacción ocurre entre los nucleótidos -8 a -2, mientras que en *B. subtilis* ocurre entre los nucleótidos -12 a -3.

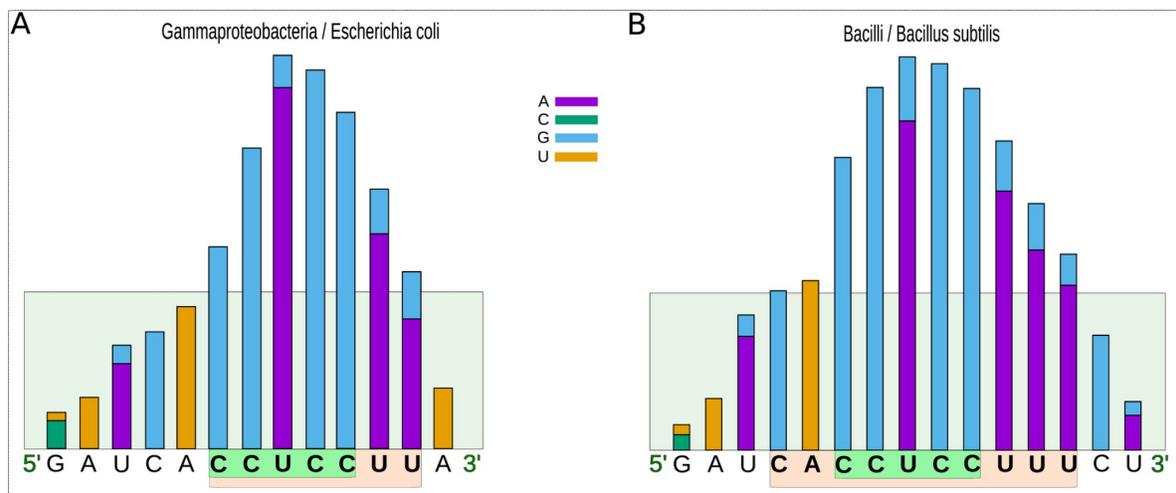


Fig 11. Histogramas de las frecuencias relativas en las que interactúan las bases de las secuencias antiSD/SD en organismos modelo. Con base en nuestros análisis termodinámicos, predijimos las bases de las secuencias líder de los mRNA en dos organismos

modelo que tienen más probabilidades de interactuar con el extremo 3' de sus rRNA 16S. Luego, representamos las frecuencias de las bases que interactúan en forma de histogramas. Las posiciones con una frecuencia relativa mayor al 40% (área verde hacia arriba) se consideraron parte de la secuencia SD, y su contraparte en el rRNA 16S se consideró una secuencia antiSD. En cada posición, se indica la frecuencia de la base que interactúa en el mRNA. Los histogramas ilustran las interacciones antiSD/SD para los organismos *E. coli* (A) y *B. subtilis* (B). Se muestra en color **nude** la secuencia antiSD de *E. coli* (5'-CCUCCUU-3') y la secuencia antiSD extendida de *B. subtilis* (5'-UCACCUCCU-3'). El CCUCC canónico se muestra en verde.

4.3 Análisis de los tipos de antiSD por clases filogenéticas

Para obtener resultados mejor respaldados y estadísticamente significativos sobre las definiciones de las secuencias de consenso antiSD y SD, repetimos el procedimiento analítico descrito anteriormente para el conjunto completo de genomas de secuencias representativas agrupados por clases filogenéticas. Como resultado, de las 87 clases consideradas inicialmente en nuestra investigación, 78 utilizan la interacción rRNA 16S/mRNA para definir el sitio de inicio de la traducción y conservan el CCUCC en al menos el 5.5% de los genes, que fue nuestro criterio de corte. El análisis de los histogramas revela la diversidad de los elementos antiSD en términos de su posición relativa al core CCUCC, lo cual puede estar relacionado con el origen filogenético de los microorganismos. Como ejemplos de esta diversidad, la Fig 12 muestra los histogramas de frecuencia de tres clases representativas: Epsilonproteobacteria, Thermatogae y Deinococci, con diferentes tendencias a sesgarse hacia la derecha, izquierda o al centro, respectivamente. Resultados preliminares indican que la posición promedio del perfil de antiSD en un organismo tiene una correlación baja con la posición promedio del SD en el mRNA (Fig. 12).

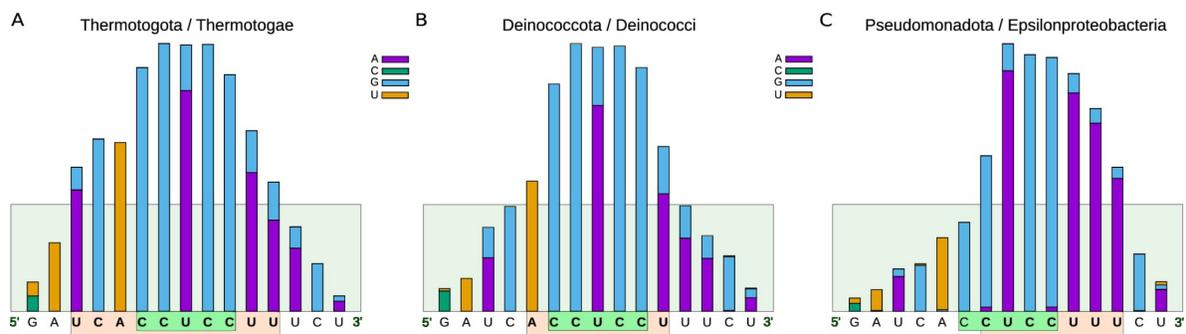


Fig 12. Histogramas de las frecuencias relativas en las que los nucleótidos del antiSD y SD interactúan en clases filogenéticas modelo-representativas. (A) Thermatogae, (B)

Deinococci y (C) Epsilonproteobacteria, como ejemplos de clases filogenéticas en las que los histogramas de las frecuencias relativas de las bases que interactúan en antiSD/SD están sesgados hacia la izquierda, centrados o sesgados hacia la derecha. Sus secuencias de antiSD y el CCUCC canónico se muestran en color naranja y verde, respectivamente.

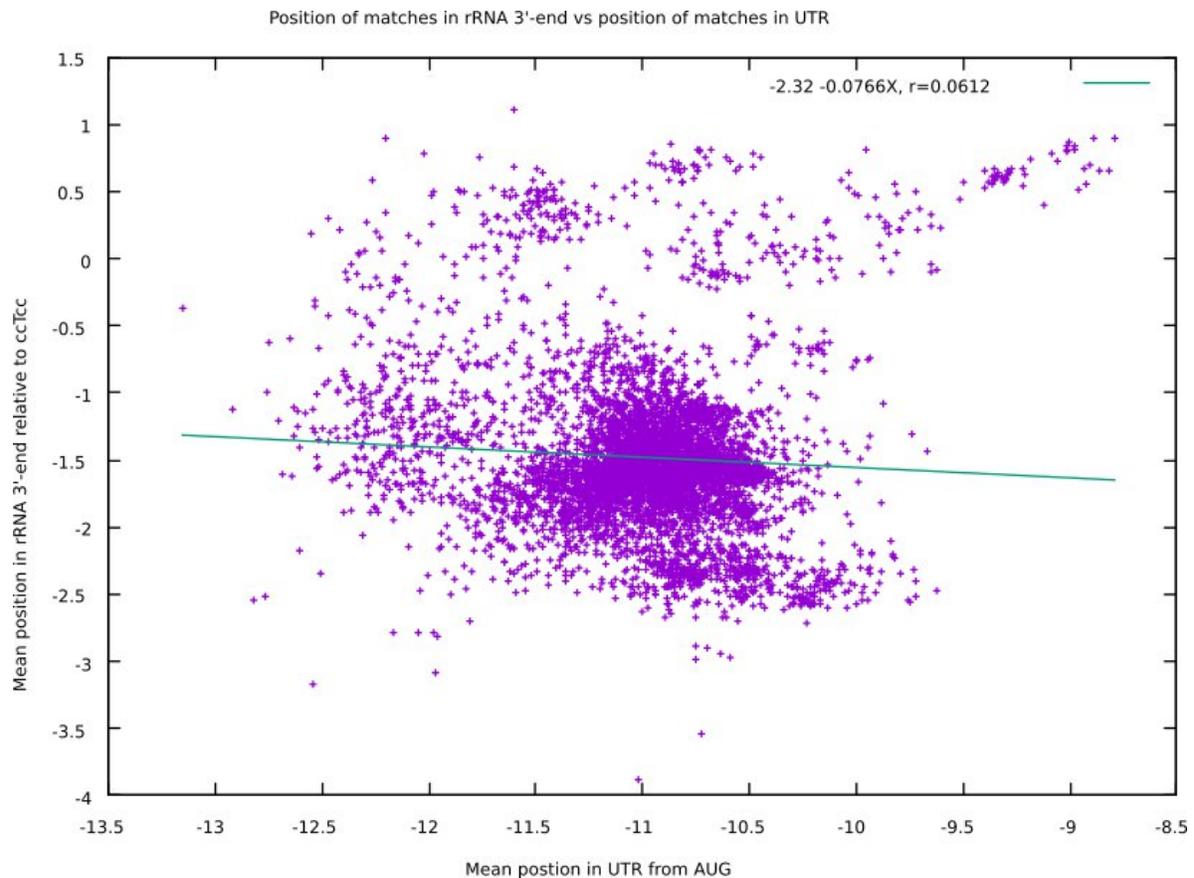


Fig 13. Comparación del promedio de la posición del antiSD con la posición promedio de los SDs en los organismos de estudio. La regresión lineal indica que la correlación es baja.

Además de las diferencias generales observadas en los histogramas de estar sesgados hacia uno de sus extremos, izquierda, centrado o derecha, llevamos a cabo un estudio más detallado en el que consideramos tanto la posición relativa del antiSD en términos del elemento central CCUCC como las diferencias en las secuencias adyacentes. Con base en estas variaciones, dividimos las clases filogenéticas analizadas en quince grupos (Fig 14). Es importante destacar que el antiSD y SD del organismo modelo *E. coli*, donde se realizó un trabajo pionero para definir las interacciones entre

el rRNA 16S y los mRNAs, corresponden al segundo grupo más grande en nuestra lista de tipos de antiSD/SD (Grupo 8).

Group	#Classes		
1	3	CCUCCUUUC 
2	1 CUCCUUU. 
3	23	CCUCCUUU. 
4	9	ACCUCCUUU. 
5	2	UCACCUCCUUU. 
6	4	UCACCUCCUU. . 
7	2	CACCUCCUU. . 
8	11	CCUCCUU. . 
9	4	ACCUCCUU. . 
10	3	ACCUCCU. . . 
11	2	CACCUCCU. . . 
12	8	UCACCUCCU. . . 
13	1	. . .	AUCACCUCCU. . . 
14	4	UCACCUCC. . . . 
15	1	. . .	AUCACCUC. . . . 

 Bacteria
  Archaea

Fig 14. Clasificación de las secuencias de antiSD según su secuencia y posición en el 3'-terminal del rRNA 16S. Con base en la frecuencia de interacciones entre los nucleótidos del 3'-terminal del rRNA 16S y las bases de las regiones 5' río arriba de los mRNAs, definimos y

clasificamos las secuencias de antiSD en nuestro conjunto de organismos de referencia según su posición relativa al CCUCC, que está presente en casi todas las moléculas de rRNA 16S. Las letras en la figura representan el consenso de las bases que interactúan en el rRNA de cada grupo, mientras que los puntos indican bases que no interactúan. Las secuencias de consenso se pueden encontrar en la Fig 11. El CCUCC está encerrado en un rectángulo rojo. Los gráficos circulares muestran las frecuencias relativas de bacterias y arqueas en cada grupo.

El análisis de las variaciones en la secuencia del antiSD, incluyendo su secuencia y longitud, indica una correlación con el origen filogenético de los microorganismos. La Fig 15 presenta histogramas representativos de cada uno de nuestros quince grupos. Una de las variaciones más evidentes ha sido reportada previamente en clases filogenéticas pertenecientes al dominio Arquea (grupos 12 al 15); el grupo 12 incluye clases de los filos Crenarchaeota, Thaumarchaeota y Euryarchaeota, mientras que los grupos del 13 al 15 solo incluyen clases del filo Euryarchaeota, en los que los histogramas se desplazan hacia la izquierda en comparación con el histograma del grupo 3, que agrupa a la mayoría de los organismos con secuencias "clásicas" de antiSD/SD.

Es importante destacar que los valores de estos porcentajes varían significativamente, desde organismos que carecen de secuencias SD en sus mRNAs hasta organismos con hasta el 90% de sus mensajeros que tienen secuencias SD. Nuestros resultados concuerdan con valores previamente reportados para clados filogenéticos en los cuales los organismos poseen secuencias SD en sus mRNAs (Omotajo et al., 2015) y para algunos organismos arqueales en los cuales el proceso de iniciación de la traducción es independiente de una secuencia SD (Torarinsson et al., 2005).

A
C
G
U

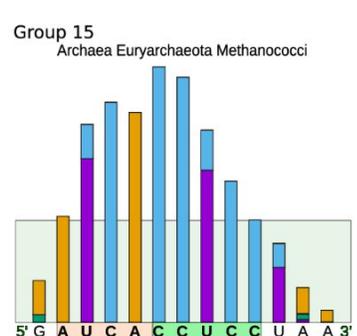
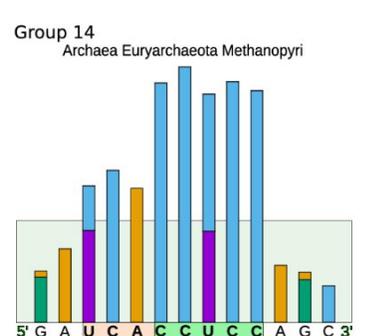
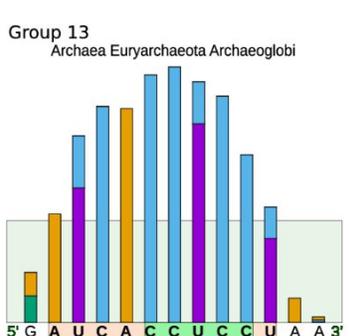
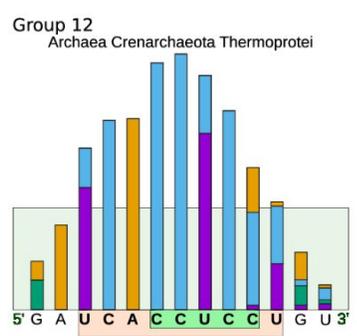
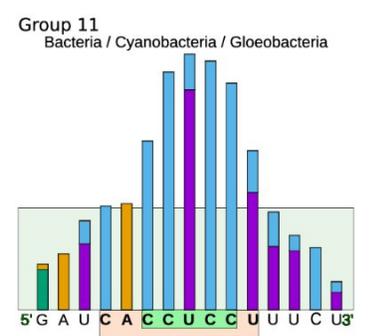
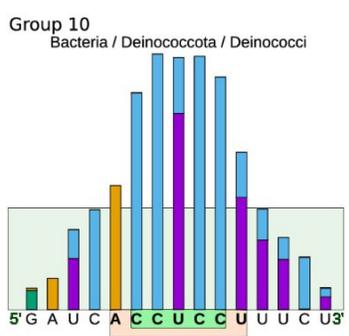
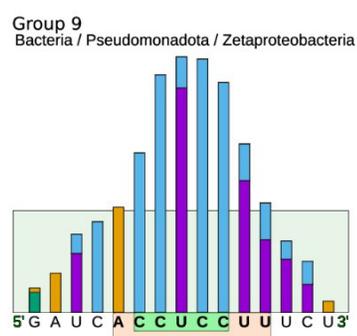
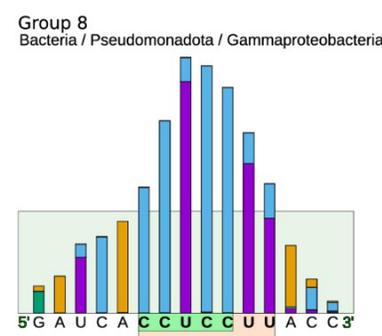
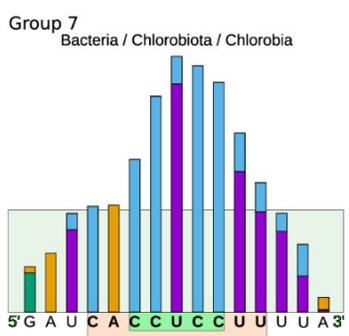
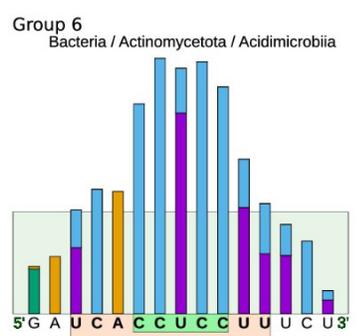
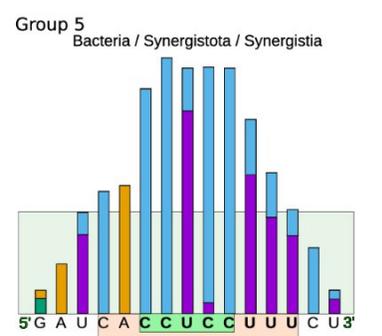
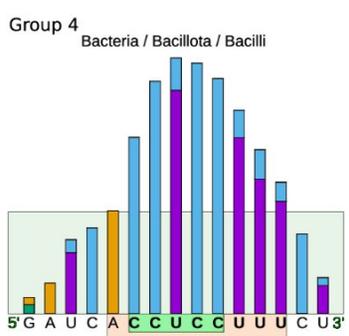
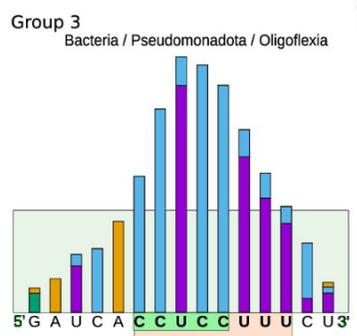
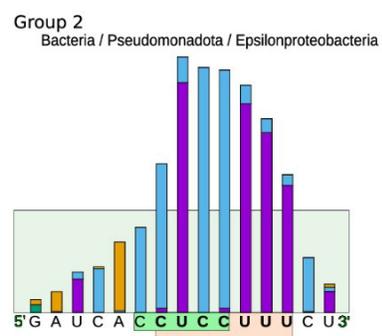
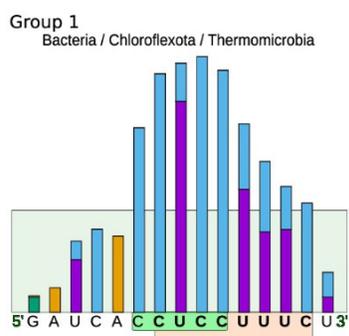


Fig 15. Histogramas de clases representativas de los quince grupos de antiSDs identificados en nuestro análisis. De las 78 clases filogenéticas, se selecciona una de cada uno de los quince grupos de antiSD que definimos en base a la similitud de sus secuencias, longitudes y posiciones dentro de su 3'-terminal.

Tabla 1. Secuencias de SD y AntiSD para procariotas agrupados por clases filogenéticas.

G	D	Phylum	Class	% SD	anti SD Pos.	anti SD Len gth	antiSD (5'-3')	SD (5'-3')
1	B	Actinomycetota	Coriobacteriia	74	-2+6	9	CCUCCUUUC	GAAAGGAGG
		Chloroflexota	Ktedonobacteria	64				
			Thermomicrobia	78				
2	C	Pseudomonadota	Epsilonproteobacteria	51	-1+5	7	CUCCUUU	AAAGGAG
3	t	Acidobacteriota	Acidobacteriia	45	-2+5	8	CCUCCUUU	AAAGGAGG
		Actinomycetota	Rubrobacteria	58				
			Actinomycetia	51				
		Armatimonadota	Chthonomonadetes	70				
		Calditrichaeota	Calditrichae	51				
		Chloroflexota	Chloroflexia	61				
			Anaerolineae	58				
			Caldilineae	42				
		Elusimicrobiota	Elusimicrobia	46				
			Endomicrobia	43				
		Fibrobacterota	Fibrobacteria	37				
		Fusobacteriota	Fusobacteriia	79				
		Ignavibacteriota	Ignavibacteria	32				
		Kiritimatiellaeota	Kiritimatiellae	63				
		Lentisphaerota	Lentisphaeria	60				
Mycoplasmata	Mollicutes	41						
Nitrospina	Nitrospina	37						
Nitrospirota	Nitrospira	57						
Planctomycetota	Phycisphaerae	46						
Pseudomonadota	Hydrogenophilalia	58						
	Oligoflexia	47						

		Alphaproteobacteria	47				
	Verrucomicrobiota	Methylacidiphilae	15				
4	Acidobacteriota	Holophagae	61	-3+5	9	ACCUCUUU	AAAGGAGGU
	Bacillota	Negativicutes	81				
		Tissierellia	78				
		Clostridia	78				
		Bacilli	74				
		Erysipelotrichia	74				
		Dictyoglomota	Dictyoglomia				
	Nitrospirota	Thermodesulfovibrionia	45				
	Thermodesulfobacteriota	Thermodesulfobacteria	50				
5	Synergistota	Synergistia	83	-5+5	11	UCACCUCCU UU	AAAGGAGGU GA
	Verrucomicrobiota	Spartobacteria	10				
6	Actinomycetota	Acidimicrobiia	45	-5+4	10	UCACCUCCU U	AAGGAGGUG A
		Thermoleophilia	45				
	Coprothermobacterota	Coprothermobacteria	63				
	Thermotogota	Thermotogae	80				
7	Acidobacteriota	Blastocatellia	32	-4+4	9	CACCUCCUU	AAGGAGGUG
	Chlorobiota	Chlorobia	13				
8	Actinomycetota	Nitriliruptoria	51	-2+4	7	CCUCCUU	AAGGAGG
	Bacillota	Limnochordia	78				
	Chlamydiota	Chlamydiia	34				
	Chloroflexota	Dehalococcoidia	59				
	Chrysiogenota	Chrysiogenetes	74				
	Planctomycetota	Planctomycetia	48				
	Pseudomonadota	Deltaproteobacteria	59				
		Acidithiobacillia	57				
		Gammaproteobacteria	47				
		Betaproteobacteria	37				
Spirochaetota	Spirochaetia	46					

9		Aquificota	Aquificae	48	-3+4	8	ACCUCCU	AAGGAGGU					
		Caldiserica	Caldisericia	53									
		Deferribacterota	Deferribacteres	61									
		Pseudomonadota	Zetaproteobacteria	56									
10		Chloroflexota	Tepidiformia	50	-3+3	7	ACCUCCU	AGGAGGU					
			Ardenticatenia	44									
		Deinococcota	Deinococci	45									
11		Acidobacteriota	Vicinamibacteria	41	-4+3	8	CACCUCCU	AGGAGGUG					
		Cyanobacteria	Gloeobacteria	37									
12		Armatimonadota	Fimbriimonadia	25	-5+3	9	UCACCUCCU	AGGAGGUGA					
		Gemmatimonadota	Gemmatimonadetes	22									
		Thaumarchaeota	Cyanophyceae	36									
	A	Euryarchaeota	Methanobacteria	56									
			Methanomicrobia	46									
	r	Thaumarchaeota	Conexivisphaeria	36									
			Nitrososphaeria	11									
	u	Thermoproteota	Thermoprotei	31									
	13	e	Euryarchaeota	Archaeoglobi					34	-6+3	10	AUCACCUC U	AGGAGGUGA U
	14	a	Euryarchaeota	Thermococci					73	-5+2	8	UCACCUC	GGAGGUGA
Methanopyri				75									
Nanohaloarchaea				22									
Halobacteria				23									
15		Euryarchaeota	Methanococci	69	-6+1	8	AUCACCUC	GAGGUGAU					

Las frecuencias de interacción entre las bases del 3'-terminal del rRNA 16S y la región inmediata río arriba del AUG de los mRNA de los organismos incluidos en nuestro estudio se determinaron y se utilizaron para definir la secuencia de consenso de la AntiSD de los organismos agrupados por clases filogenéticas. El primer y último nucleótido de la secuencia de AntiSD se miden desde el nucleótido central U del CCUCC, que se encuentra en la posición 0.

Cabe destacar que, aunque el grupo 12 incluye clases filogenéticas de bacterias y arqueas, comparten tipos de secuencias consenso similares para sus elementos de SD y AntiSD, que, en comparación con otros grupos, son más largos y desplazados hacia la izquierda con respecto al CCUCC. Las clases más representativas de este grupo son Fimbriimonadia, Gemmatimonadetes y Cyanophyceae del dominio Bacteria, y Methanobacteria, Methanomicrobia, Conexivisphaeria, Thermoprotei y Nitrososphaeria del dominio Arquea (Fig 16). **Estos resultados indican que en lugar de haber una secuencia consenso para bacterias y otra para arqueas, como se informó anteriormente (Torarinsson et al., 2005), existe un conjunto más amplio de secuencias de AntiSD/SD que los organismos han seleccionado para definir los sitios de inicio de la traducción y cuyas diferencias están asociadas a las clases filogenéticas a las que pertenecen.**

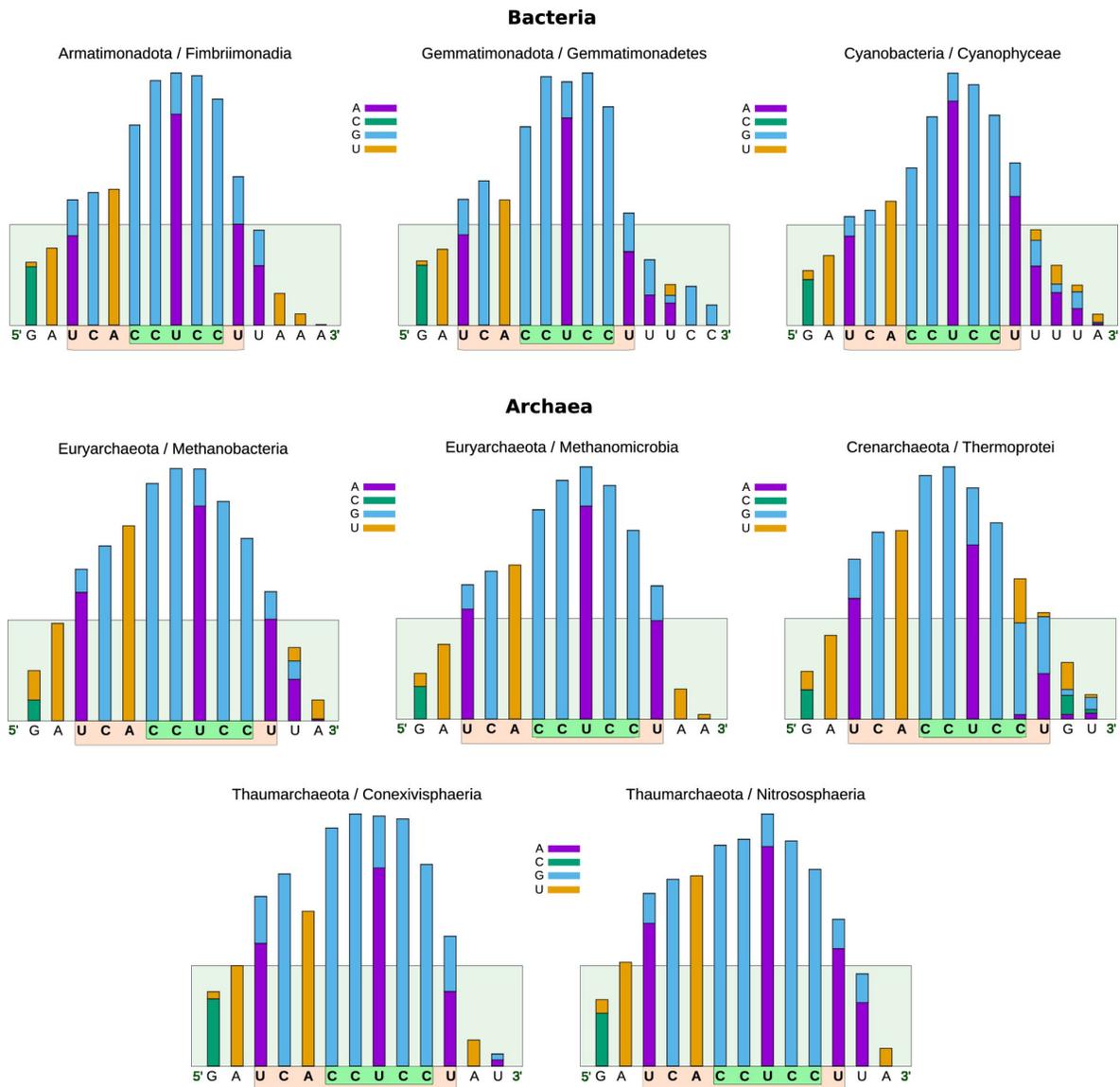


Fig 16. Desglose de las clases filogenéticas en el grupo 12 (de la Figura 13). Los histogramas incluyen tres clases filogenéticas de arqueas y cinco de bacterias. El grupo 12 es el único que incluye miembros de ambos dominios. Este grupo se caracteriza por tener una secuencia de AntiSD relativamente larga que comienza más a la izquierda del CCUCC, en comparación con la mayoría de las secuencias de AntiSD bacterianas.

4.4 Clasificación de los organismos según la presencia/ausencia de la secuencia AntiSD y la capacidad de sus mRNAs para interactuar con el rRNA 16S.

Los 15 grupos mencionados anteriormente representan clases taxonómicas que cumplen dos criterios:

1- Exhiben un CCUCC canónico reconocible.

2- El 3'-terminal de su rRNA 16S está involucrado en la unión estable de los mRNAs.

En contraste, otras clases taxonómicas no cumplen estas condiciones. Se considera que un organismo carece del CCUCC si alguna de las cinco bases de consenso difiere. De manera similar, se considera que un organismo carece de una unión estable si menos del 5.5% de los mRNAs demuestran una unión.

Al combinar la presencia o ausencia de estas dos características, podemos clasificar a todos los organismos en cuatro categorías, como se describe a continuación. La distribución filogenética de estas categorías se muestra en la Fig 17.

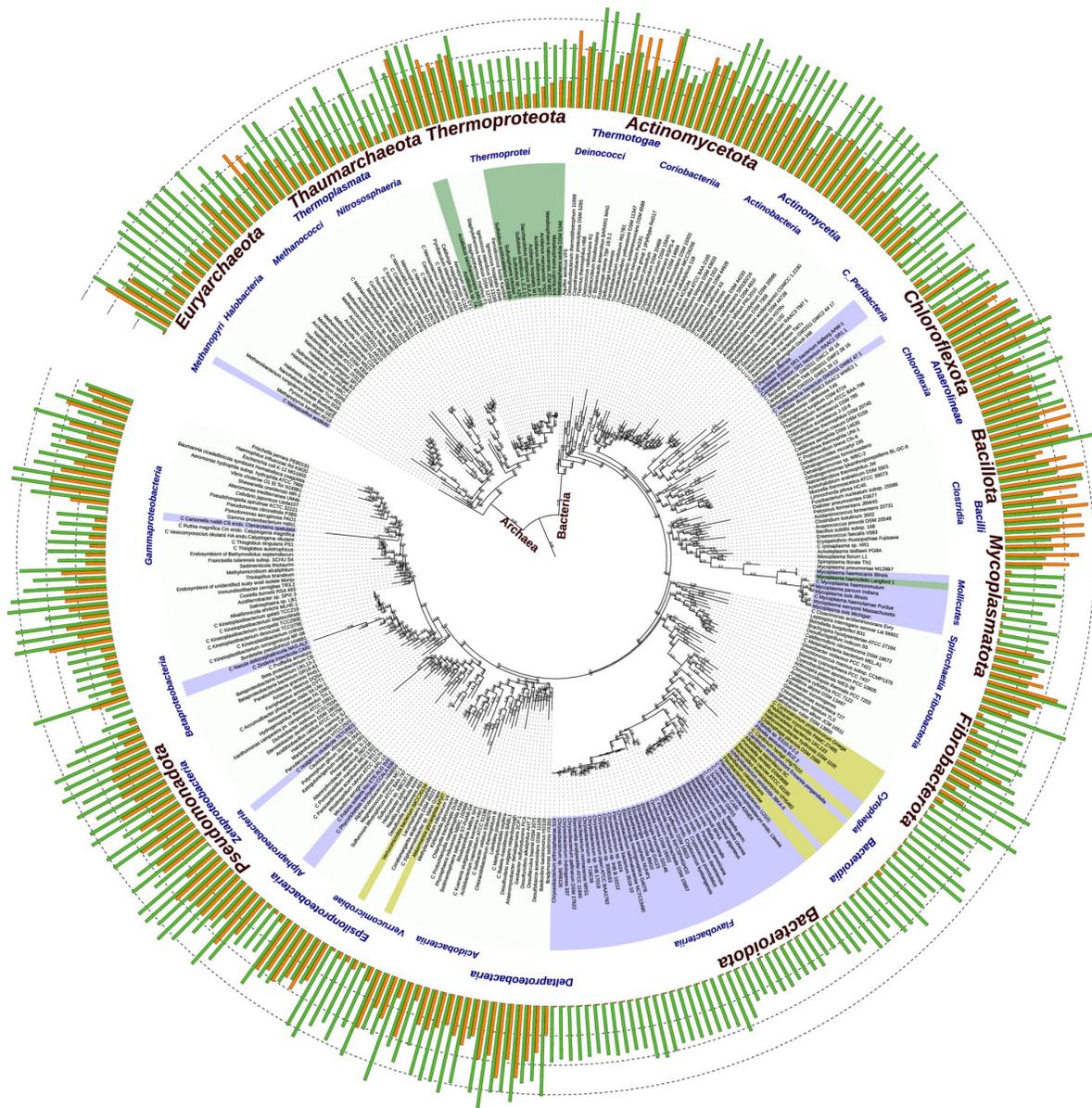


Fig 17. Árbol filogenético de organismos representativos con diferentes clases de secuencias AntiSD/SD. Para representar la distribución filogenética de los organismos pertenecientes a los diferentes tipos de categorías mencionadas anteriormente, se seleccionaron 323 miembros representativos de nuestra lista de organismos estudiados, y se utilizaron sus secuencias de genes rRNA 16S para construir el árbol filogenético. Las barras verdes y naranjas muestran el contenido de GC y el porcentaje de mRNAs capaces de establecer interacciones estables entre rRNA 16S y mRNA, respectivamente. Los nombres de los organismos están coloreados según uno de los cuatro grupos a los que pertenecen: sin color, amarillo, verde y morado, para representar los grupos i, ii, iii y iv, respectivamente.

Imagen en alta resolución en:

https://github.com/kjestrada/antiSD_paper_scripts/blob/master/Figures/Fig9.png

i) Organismos con alta ocurrencia de SD y el núcleo CCUCC conservado

Esta categoría es la más grande. De nuestro conjunto de 6,457 organismos considerados en nuestro estudio, el 92.6% pertenece a esta categoría (Figura 16, organismos no coloreados). Esta categoría comprende las 78 clases con perfiles de antiSD asignados (Figura 13). Aunque el criterio para considerar que hay alta ocurrencia de SD es del 5.5% de genes con SD, la mayoría de los organismos en esta categoría tienen una frecuencia mucho más alta. Las bacterias *E. coli* y *B. subtilis*, que fueron nuestras especies de referencia, pertenecen a esta categoría. Tienen el 64% y el 84% de los genes con SD, respectivamente. La categoría incluye a la mayoría de las especies bacterianas y arqueas. Curiosamente, las arqueas halófilas (*Halobacteria*), que inician la traducción de muchos de sus genes con un mecanismo independiente de SD, también pertenecen a esta categoría. La interacción SD/antiSD en esta categoría se considera canónica ya que abarca la mayoría de los organismos con la distribución filogenética más amplia. Hace casi 50 años, la secuencia SD, la complementariedad con el extremo 16S del rRNA y su papel en la iniciación de la traducción fueron propuestos por primera vez por J. Shine y L. Dalgarno (Shine & Dalgarno, 1975) en un estudio que involucraba a 4 miembros de esta categoría: *E. coli*, *P. aeruginosa*, *B. stearothermophilus* y *C. crescentus*.

ii) Organismos con un CCUCC canónico conservado, pero deficientes en las interacciones rRNA 16S/mRNA

A pesar de poseer el CCUCC canónico en sus rRNAs, los organismos de este grupo no pueden formar interacciones estables con la región 5' que precede al codón de inicio AUG. Estos organismos incluyen principalmente miembros de las clases Bacteroidia, Flavobacteriia, Chitinophagia, Cytophagia y Sphingobacteriia del filo Bacteroidota, pero también Mollicutes del filo Mycoplasmatota. En menor medida, también identificamos en este segundo grupo algunos miembros del filo Verrucomicrobia, como *Akkermansia muciniphila* y *Verrucomicrobia bacterium IMCC26134*, una Gammaproteobacteria y una Arquea, *Nanoarchaeum equitans* Kin4-M de la clase Nanoarchaeales. Estos organismos se muestran como nodos de color amarillo en la Fig 17. El bajo porcentaje de sus interacciones antiSD/SD se indica por la altura pequeña de las barras naranjas en la

figura confirmando un inicio de la traducción independiente de estas. Nuestro estudio confirma y amplía estos hallazgos indicando que la frecuencia de secuencias SD en Bacteroidota es menor que la frecuencia en comparación con lo que se esperaría por azar (Fig 18). La desviación observada sugiere la posibilidad de que la selección actúe sobre las UTR para evitar la interacción con el extremo 3' del ARN ribosomal 16S. Teniendo en cuenta que la iniciación de la traducción en los organismos pertenecientes a este segundo grupo es independiente de la interacción directa entre el rRNA 16S y la región líder del mRNA, así como su cercanía evolutiva a muchas otras especies que carecen de antiSD, es posible pensar que estas especies están en transición hacia la pérdida final del núcleo antiSD y eventualmente formarán parte de nuestra cuarta categoría.

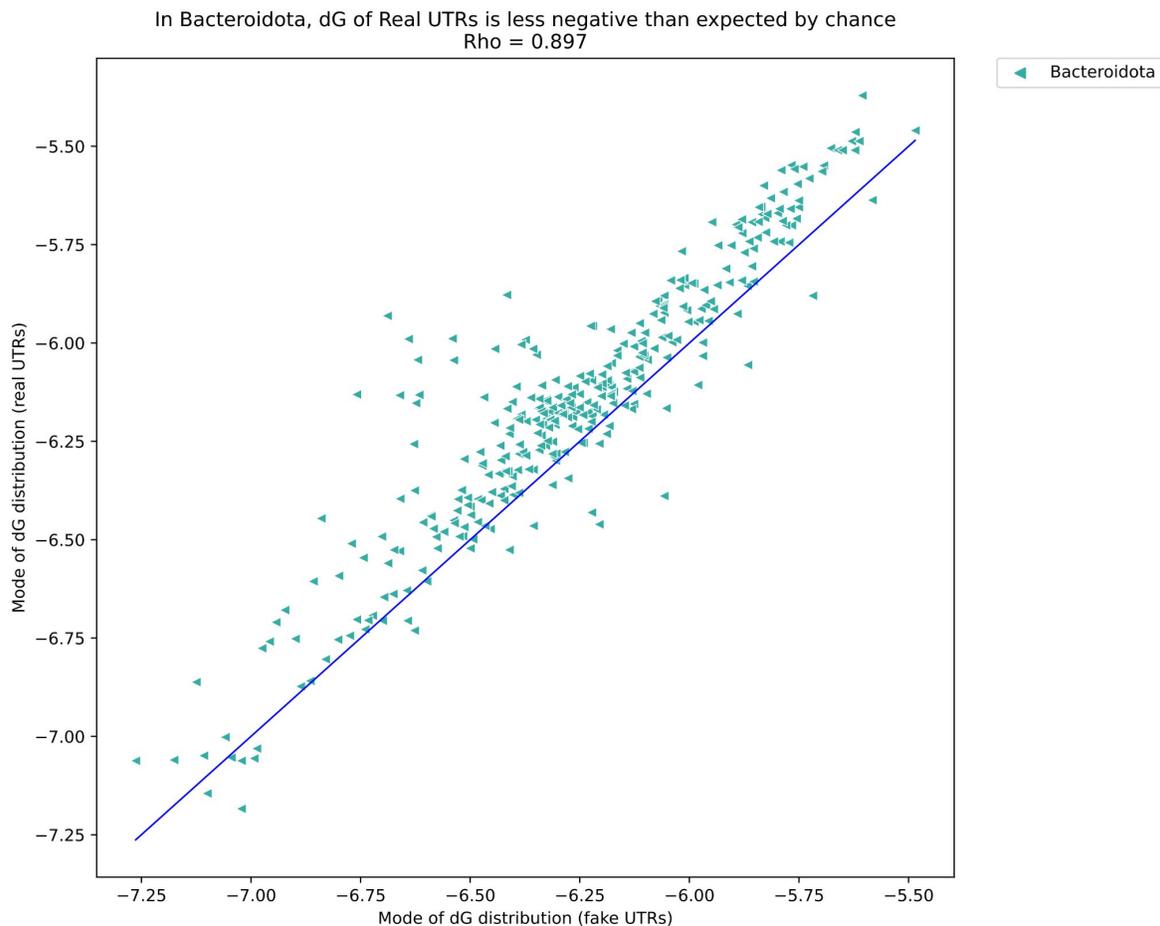


Fig 18. En el filo Bacteroidota, los valores de ΔG de las UTR tienden a ser menos negativos (indicando interacciones más débiles) que en secuencias aleatorias. Las secuencias aleatorias o UTR "falsas", son generadas aleatoriamente con el mismo contenido de GC de los

organismos.

Tabla 2. Organismos identificados en nuestra investigación por tener el CCUCC canónico en su rRNA 16S, sin embargo, menos del 5.5% de sus mRNAs pueden tener interacciones estables con los terminales 3' de sus rRNA 16Ss.

D	Phylum	Class	Species	% SD	Total
Bacteria	Bacteroidota	Bacteroidia	<i>Alistipes finegoldii</i>	2	66
			<i>Bacteroides fragilis</i>	2	
			<i>Prevotella oris</i>	2	
			...		
		Chitinophagia	<i>Arachidicoccus B3-10</i>	1	19
			<i>Chitinophaga caeni</i>	2	
	<i>Niastella koreensis</i>		1		
...					
Cytophagia	<i>Bernardetia litoralis</i>	0.1	54		
	<i>Hymenobacter nivis</i>	2			
	<i>Spirosoma aureum</i>	1			
	...				
Flavobacteriia	<i>Aquimarina sp. AD1</i>	1	130		
	<i>Flavobacterium album</i>	1			
	<i>Gramella fulva</i>	1			
	...				
Saprospira	<i>Saprospira grandis</i>	2	2		
	<i>Haliscomenobacter hydrossis</i>	1			
Sphingobacteriia	<i>Pedobacter cryoconitis</i>	1	31		
	<i>Sphingobacterium mizutaii</i>	2			
	...				
Cyanobacteria	Cyanophyceae	<i>Prochlorococcus marinus</i>	3	4	

	Pseudomonadota	Gammaproteobacteria	<i>Endosymbiont of S. gigas</i>	1	1
	Mycoplasmata	Mollicutes	<i>Mycoplasma genitalium</i> ...	2	6
	Verrucomicrobiota	Verrucomicrobiae	<i>Akkermansia glycaniphila</i> ...	4	2
Arquea	Nanoarchaeota	Nanoarchaeales	<i>Nanoarchaeum equitans</i>	3	1

La identificación del CCUCC en los terminales 3' de los rRNA 16S de los organismos estudiados se realizó utilizando el programa MAST y las matrices MEME, como se describe en las secciones de Materiales y Métodos. El número de mRNAs capaces de establecer interacciones estables con los rRNA 16S se indica como un porcentaje. Solo se incluyen organismos representativos por clase filogenética en la tabla, pero el número total de organismos identificados en cada categoría se muestra en la última columna.

iii) Organismos que carecen del CCUCC canónico, cuyos extremos 3' de los rRNA 16S pueden interactuar de manera estable con la región líder de los mRNAs.

Este grupo de organismos se caracteriza por una modificación en el elemento CCUCC de sus rRNA 16S; sin embargo, los extremos 3' de sus rRNA 16S aún pueden formar interacciones estables con una fracción significativa de los mRNAs totales del organismo, lo cual se representa mediante barras naranjas en la Fig 17 y puede variar del 5.5% a más del 70% (Tabla 3). Los miembros de este grupo están representados por el color verde en la Fig 17. Principalmente pertenecen al dominio Arquea, como es el caso de *Acidilobus* sp.7A y *Sulfolobus islandicus*, cuyos genes de rRNA 16S presentan cambios en su 3'-terminal, desde el CCUCC canónico al CCUCU y CCUCA, respectivamente (Fig 19). En casos excepcionales, también encontramos organismos bacterianos que pertenecen a este grupo, como el Alphaproteobacteria *Erythrobacter litoralis*. En este caso, el CCUCC canónico ha cambiado a ACUCC (Fig 19). Se esperaría que los cambios en el núcleo antiSD se acompañen de cambios correspondientes en el SD de muchos genes. Ese es el caso de C → U en la posición -4 en el rRNA 16S de *Acidilobus* sp.7A, que parece tener un cambio correspondiente G → A en las secuencias

de SD de sus mRNAs, como puede indicar nuestro análisis termodinámico. Curiosamente, ni el C – A en la posición -4 ni el C – A en la posición -10 en *Sulfolobus islandicus* y *Erythrobacter litoralis*, respectivamente, parecen tener un contraparte interactiva en las regiones líder de sus mRNAs. Además de los organismos mencionados anteriormente, nuestro estudio identificó 37 organismos que pertenecen a esta clasificación. La Tabla 3 incluye ejemplos representativos de los miembros de este grupo.

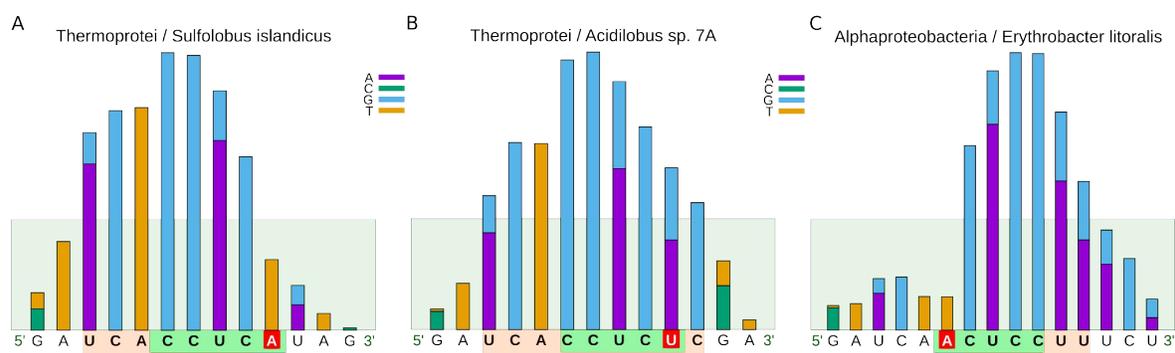


Fig 19. Histogramas de las frecuencias en las que las bases del 3'-terminal del rRNA 16S interactúan con la región 5' upstream de los mRNAs con frecuencias relativamente altas para organismos que carecen del CCUCC canónico. A) y B) corresponden a los histogramas de organismos filogenéticos arqueales, mientras que C) corresponde a la clase Alphaproteobacteria del dominio Bacteria. Según las frecuencias de interacción con las regiones líder de los mRNAs, las bases en el extremo 3' del rRNA 16S se definieron como parte de la secuencia antiSD y se indican en cajas marrones. Las secuencias canónicas CCUCC del rRNA 16S están enmarcadas en cajas verdes, y los cambios que las alejan del consenso se indican en cajas rojas.

Tabla 3. Organismos cuyos 3'-terminal del rRNA 16S carecen del núcleo canónico CCUCC, sin embargo, establecen interacciones estables con las regiones 5' upstream de sus mRNAs.

D	Phylum	Class	Species	% SD	Total
Arquea	Thermoproteota	Thermoprotei	<i>Acidianus sulfidivorans</i>	14	38
			<i>Acidilobus saccharovorans</i>	54	
			<i>Acidilobus sp. 7A</i>	53	

			<i>Metallosphaera prunae</i>	26	
			<i>Sulfolobus islandicus</i>	18	
			...		
Bacteria	Pseudomonadota	Alphaproteobacteria	<i>Erythrobacter litoralis</i>	43	1

iv) Organismos que carecen del núcleo canónico CCUCC y no muestran una interacción estable entre el 3'-terminal del rRNA 16S y el líder de los mRNAs.

Estos organismos se han identificado en el dominio de Arquea, particularmente en la clase Nanoarchaeia, donde muchos microorganismos se caracterizan por la ausencia de una región líder en sus mRNAs; consecuentemente, la presión selectiva para preservar un CCUCC funcional en sus genes de rRNA 16S se ha perdido. Además, también hemos identificado organismos pertenecientes a esta clasificación en el dominio de las bacterias, principalmente en la clase Flavobacteriia, lo cual es consistente con informes previos. Las 75 flavobacteriia presentes en esta categoría representan el 67% de sus miembros totales y constituyen el 55% de todos los Flavobacteriia en nuestro estudio. Nuestro estudio identificó 111 organismos en esta clasificación, algunos de los cuales se muestran como nodos de color púrpura en la Fig 17. Estos resultados proporcionan evidencia de que la pérdida de secuencias antiSD funcionales no se limita a las arqueobacterias, sino que podría haber ocurrido en diferentes momentos durante la evolución de los organismos procariotas. La Tabla 4 incluye ejemplos representativos de miembros de este grupo.

Tabla 4. Organismos procariotas que carecen del núcleo de secuencia antiSD CCUCC en sus rRNA 16S tampoco tienen interacciones estables con la región líder de sus mRNAs.

D	Phylum	Class	Species	% SD	Total
Arquea	Nanoarchaeota	C.Nanoarchaeia	<i>C.Nanopusillus acidilobi</i>	7	1
B a c t e r i a	Bacteroidota	Flavobacteriia	<i>Weeksella virosa</i>	2	75
			<i>Chryseobacterium glaciei</i>	3	
			<i>Riemerella anatipestifer</i>	3	
			...		
		Bacteroidia	<i>Alistipes_sp._dk3624</i>	7	1
		Cytophagia	<i>Fibrella sp.ES10-3-2-2</i>	3	3
	<i>Fibrella aestuarina</i>		5		
	<i>Cardinium sp.cEper1</i>		5		
			...		
		Mycoplasmata	Mollicutes	<i>Mycoplasma haemocanis</i>	8
	<i>Mycoplasma ovis</i>			5	
		...			
	Pseudomonadota	Alphaproteobacteria	<i>C.Hodgkinia cicadicola</i>	3	6
			<i>C. Fokinia solitaria</i>	7	
			...		
		Betaproteobacteria	<i>C.Vidania fulgoroideae</i>	3	3
	<i>C.Nasuia deltocephalinicola</i>		2		
	...				
		Gammaproteobacteria	<i>C.Carsonella ruddii CE</i>	2	7
	<i>C.Carsonella ruddii PC</i>		3		
	...				

	Verrucomicrobiota	No rank	<i>C. Pinguicoccus supinus</i>	3	1
	Candidatus Absconditabacteria	No rank	<i>C. division SR1 RAAC1</i>	2	2
			<i>C. division SR1 Aalbo</i>	7	
	Candidatus Campbellbacteria	No rank	<i>C. Campbellbacteria</i>	5	1
	Candidatus Peregrinibacteria	No rank	<i>C. Peribacter riflensis</i>	7	1
Candidatus Wolfbacteria	No rank	<i>C. Wolfbacteria GW2</i>	2	1	

Se muestran ejemplos representativos de organismos pertenecientes a la clase iv. La "C." en los nombres de las especies significa "Candidatus".

En nuestros análisis prestamos especial atención a la identificación de organismos procariontes que no poseen el CCUCC en los extremos 3' de sus genes ribosomales 16S. Nuestro estudio ha clasificado a estos organismos en nuestras categorías tercera y cuarta, basándonos en análisis termodinámicos que utilizan las interacciones de pares de bases entre sus rRNA 16S/mRNA para determinar el inicio de la traducción. En nuestro estudio, realizamos alineamientos múltiples rigurosos de secuencias del extremo 3' de rRNA 16S previamente identificados en nuestro análisis de conservación de secuencias de motivos. Como resultado, identificamos 96 nuevos organismos bacterianos y 39 nuevas arqueas cuyos genes de rRNA 16S carecen de la secuencia antiSD canónica. Es importante tener en cuenta que aunque los errores de secuenciación genómica pueden llevar a falsos positivos en nuestra lista de organismos, dos factores hacen que sea poco probable que esto ocurra en las secuencias encontradas. En primer lugar, el número de copias del gen rRNA 16S: casi la mitad de estos organismos tienen más de una copia pero mantienen la misma variante antiSD entre ellas. En segundo lugar, la preservación de cambios de secuencia entre especies: el hecho de que diferentes genomas de especies estrechamente relacionadas presenten la misma secuencia antiSD también ocurre en la gran mayoría de ellos. Este conjunto de organismos incluye cinco filos diferentes (Bacteroidota, Pseudomonadota, Candidatus Absconditabacteria, Candidatus Wolfbacteria y Candidatus

Campbellbacteria) y cinco clases diferentes (Bacteroidia, Cytophagia, Flavobacteriia, Alphaproteobacteria y Betaproteobacteria) en el caso de los organismos bacterianos, y tres filos diferentes (Thermoproteota, Nanoarchaeota y Crenarchaeota) y dos clases diferentes (Thermoprotei y C. Nanoarchaeia) en arquea. Este estudio proporciona la primera visión global de esta característica: la ausencia de la secuencia CCUCC, en arquea.

Debido a las anotaciones genómicas inexactas o fallas en los algoritmos computacionales, es importante tener en cuenta que la falta de identificación *in silico* de elementos antiSD en una secuencia genómica no implica necesariamente la ausencia real de la secuencia antiSD. Por ejemplo, del conjunto de organismos que carecen del CCUCC reportados en (Amin et al., 2018), nuestra metodología sí identificó en tres de ellos (*Bacillus azotoformans* LMG 9581 (GCA_000307855.1), *Bacillus bataviensis* LMG 21833 (GCA_000307875.1) y *Microcystis aeruginosa* SPC777 (GCA_000412595.1)) una secuencia antiSD canónica.

5. CONCLUSIONES

En los últimos 50 años, nuestra comprensión de los elementos que definen la iniciación de la traducción en organismos procariotas ha evolucionado significativamente desde el descubrimiento revolucionario de J. Shine y L. Dalgarno de una secuencia de consenso en el extremo 3' del rRNA 16S y su secuencia correspondiente en la región líder de los mRNAs, hasta la visión que tenemos hoy gracias a los resultados obtenidos mediante técnicas de alto rendimiento y el análisis de miles de genomas completamente secuenciados. En términos generales, dividimos a los organismos procariotas en dos grupos: aquellos que utilizan, en gran medida, las interacciones de apareamiento de bases entre el rRNA 16S y el mRNA para definir la iniciación de la traducción de sus mRNAs, y aquellos que son en su mayoría independientes de estas interacciones.

Con respecto al primer grupo, nuestros resultados amplían informes anteriores sobre variaciones en la secuencia SD observadas entre diferentes grupos filogenéticos, principalmente al comparar las secuencias de consenso de SD de bacterias y arqueas. Nuestro análisis de casi 6,500 organismos pertenecientes a 87 clases filogenéticas nos permitió definir quince consensos adicionales de antiSD que mantuvieron el CCUCC sin cambios, pero mostraron cambios particulares alrededor de este elemento central asociados con el origen filogenético de los organismos. La existencia de este conjunto de secuencias de consenso es consistente con la idea de una evolución gradual del antiSD en lugar de un cambio binario abrupto entre las secuencias antiSD de los dos grandes dominios filogenéticos, bacterias y arqueas. Esta idea se ve respaldada por nuestro hallazgo de que el Grupo 12 de nuestra clasificación de antiSD tiene miembros tanto bacterianos como arqueales. Además de las quince secuencias de consenso diferentes de antiSD, identificamos claros ejemplos de organismos en los que los rRNA 16Ss tienen cambios de nucleótidos en una de las bases del CCUCC de este elemento casi universal de los rRNA 16Ss en organismos procariotas. Curiosamente, en algunos casos (por ejemplo, *Acidilobus* sp. 7A), identificamos un cambio compensatorio de nucleótidos en la mayoría de las secuencias líder de mRNA, de manera que el grado de interacciones entre el rRNA 16S y el mRNA se mantuvo similar al observado en la mayoría de los organismos que pertenecen a cualquiera de nuestras quince categorías de antiSD, mientras que en otros casos, el cambio en el CCUCC excluyó la posición alterada de la interacción.

Nuestro análisis termodinámico de las interacciones entre el rRNA 16S y el mRNA también nos permitió identificar organismos cuya iniciación de la traducción es independiente de la interacción de nucleótidos entre estas dos moléculas de RNA. Nuestros resultados concuerdan con los informes previos sobre mRNAs sin líder en muchas arqueobacterias y en algunos otros organismos bacterianos de la clase Flavobacteriia, donde se ha informado sobre la ausencia del CCUCC en las regiones del extremo 3'. Además, hemos significativamente ampliado esta lista de organismos con la ausencia del núcleo, casi duplicándola, descubriendo organismos representativos de cinco filos previamente no explorados y cinco clases distintas. Este hallazgo respalda la

idea de que la pérdida de la vía de iniciación de la traducción, dependiente de las interacciones de apareamiento de bases entre el rRNA 16S y el mRNA, podría ocurrir a través de múltiples eventos evolutivos.

En conjunto, nuestros resultados iluminan la rica diversidad de estrategias de iniciación de la traducción en el dominio procariota, así como la increíble adaptabilidad y diversidad de la vida a nivel molecular.

6. PERSPECTIVAS

- Caracterizar y ampliar de manera extensa la anotación de genes sin región 5' UTR (sin líder). La clasificación de este tipo de genes ha recibido menor atención, en gran parte debido a la escasez de información y secuencias de RNA disponibles hasta hace algunos años. Nuestra estrategia implica entrenar un modelo de inteligencia artificial utilizando dos conjuntos de datos: uno que consta de regiones genómicas río arriba del AUG en genes validados mediante el mapeo de RNAseq al genoma de los organismos, y otro que comprende genes que utilizan la interacción SD/antiSD, identificados en este estudio.
- Desarrollar métodos computacionales nuevos y mejorados para la identificación de secuencias SD. Aunque la metodología utilizada en este estudio para caracterizar genes con posibles secuencias SD nos proporcionó una buena aproximación, creemos que considerar otras variables adicionales para validar las secuencias SD específicas de cada organismo, mediante la aplicación de estadística bayesiana en la predicción, podría resultar en una menor tasa de error.

7. APÉNDICE

7.1 Algunos de los scripts de perl utilizados durante el estudio:

1) Extraer los datos de las bases que superaron el límite del 40 por ciento.

```
use Math::Round;

# Extract the data from the bases that passed the 40 percent cutoff
my (%data);
my $mayor= 0;
my $file_to_open= shift;
(my $name= $file_to_open)=~ s/\.txt//;
open IN, $file_to_open or die "Cant read $file_to_open\n";
while(<IN>){
    chomp;
    $data{(split)[1]}= (split)[2];
    $mayor < (split)[2] && ($mayor = (split)[2]);
}
my $thold= $mayor*40/100;
my @pos;
if( $mayor >= 10 ){
    print "$name\t",sprintf("%.2f", $mayor),"\t",sprintf("%.2f", $thold),"\t";
    foreach my $pos ( sort {$a<=>$b} keys %data ){
        print "$pos," if round($data{$pos}) >= $thold;
        push @pos,$pos if round($data{$pos}) >= $thold;
    }
    print "\t",sprintf("%.1f", mean(\@pos)), "\n";
}
sub mean{
    my($data) = @_;
```

```

if (not @$data) {
    die("Empty array\n");
}
my $total = 0;
foreach (@$data) {
    $total += $_;
}
my $average = $total / @$data;
return $average;
}

```

2) Generar histogramas por especie

```

foreach $renglon (split(/\n/, `cat tail_30_16S`))
{
    ($org,$tmp) = split(/\t/, $renglon);
    $colita{$org} = $tmp;
}

foreach $renglon (split(/\n/, `cat orgs`))
{
    ($org,$filo) = split(/\t/, $renglon);
    @x = split(/\./, $filo);

    print "$x[20]\t";
    $titulo = $x[6]. " .. ".$x[20];
    $n++;
    print "$n\t$org\t";
    $colita = $colita{$org};
    $infile = "Analisis_de_RNAhybrid/".$org.".txt";
    $outfile = "Graficas_SD/".$org.".svg";
}

```

```

    $mayor = `cat $infile | cut -f3 | sort -n | tail -1`;
    chomp($mayor);
    $divisiones = 200;
if($mayor <= 90) {$divisiones = 6};
if($mayor <= 80) {$divisiones = 4};
if($mayor <= 70) {$divisiones = 4};
    if($mayor <= 60) {$divisiones = 4};
    if($mayor <= 50) {$divisiones = 4};
    if($mayor <= 40) {$divisiones = 4};
    if($mayor <= 30) {$divisiones = 3};
    if($mayor <= 20) {$divisiones = 2};
    if($mayor <= 10) {$divisiones = 0.5};
    if($mayor <= 5) {$divisiones = 0.1};
    if($mayor <= 1) {$divisiones = 0.05};
    if($mayor <= 0.1) {$divisiones = 0.005};
    print "mayor = $mayor ... $divisiones\n";
    &grafica;
}
sub grafica
{
open(S,">datos_para_grafica.txt");
print S "#\tA\tC\tG\tU\n";
foreach $linea (split(/\n/,`nl $infile | sort -nr | cut -f2-`))
{
    $i++;
    @ele = split (/^t/, $linea);
    $base[$i] = ".";
    if ($ele[7] > $ele[8] && $ele[7] > $ele[9] && $ele[7] > $ele[10] ) {$base[$i] = "A"}
    if ($ele[8] > $ele[7] && $ele[8] > $ele[9] && $ele[8] > $ele[10] ) {$base[$i] = "C"}
    if ($ele[9] > $ele[7] && $ele[9] > $ele[8] && $ele[9] > $ele[10] ) {$base[$i] = "G"}
}
}

```

```

        if ($ele[10] > $ele[7] && $ele[10] > $ele[8] && $ele[10] > $ele[9]) {$base[$i] = "U"}
        print S "$base[$i]\t$ele[3]\t$ele[4]\t$ele[5]\t$ele[6]\n";
    }
close(S);
open(I,">instru");
print I "set terminal svg\n";
print I "set output \'$outfile\'\n";
print I "set title \'$titulo\'\n";
    print I "set key right\n";
print I "set grid y\n";
print I "set style data histograms\n";
print I "set style histogram rowstacked\n";
print I "set boxwidth 0.5\n";
print I "set style fill solid 1.0 border -1\n";
print I "set ytics $divisiones nomirror\n";
print I "set yrange [:$mayor]\n";
print I "set ylabel \'Percentage (%)\'\n";
print I "set ytics $divisiones\n";
print I "plot \'datos_para_grafica.txt\' using 2 t \'A\', \'\' using 3 t \'C\', \'\' using 4 t \'G\', \'\'
using 5:xtic(1) t \'T\'\n";
close(I);
system "gnuplot < instru";
}

```

3) Generar histograma para las clases

```

foreach $renglon (split(/\n/, `ls Analisis_de_RNAhybrid_por_clase_filogenetica`))
{
    ($gpo) = split(/\./,$renglon);
    $titulo = $gpo;
    $titulo =~ tr/_/ /;
}

```

```

    $infile = "Analisis_de_RNAhybrid_por_clase_filogenetica/".$renglon;
    $outfile = "Graficas_SD_por_clase_filogenetica/".$gpo.".svg";
print "renglon:$renglon\ninfile:$infile\noutfile:$outfile\n";

    $mayor = `cat $infile | cut -f3 | sort -n | tail -1`;
    chomp($mayor);
print "mayor $mayor\n";

    if($mayor <= 60) {$divisiones = 6};
    if($mayor <= 50) {$divisiones = 5};
    if($mayor <= 40) {$divisiones = 4};
    if($mayor <= 30) {$divisiones = 3};
    if($mayor <= 20) {$divisiones = 1};
    if($mayor <= 10) {$divisiones = 0.5};
    if($mayor <= 5) {$divisiones = 0.25};
    if($mayor <= 1) {$divisiones = 0.05};
    if($mayor <= 0.1) {$divisiones = 0.005};
    print "mayor = $mayor ... $divisiones\n";
    &grafica;
}

sub grafica
{
open(S,">datos_para_grafica.txt");
print S "#\tA\tC\tG\tU\n";
foreach $linea (split(/\n/,`nl $infile | sort -nr | cut -f2-`))
{
    $i++;
    @ele = split(/\t/, $linea);
    if ($ele[7] > $ele[8] && $ele[7] > $ele[9] && $ele[7] > $ele[10] ) {$base[$i] = "A"}
    if ($ele[8] > $ele[7] && $ele[8] > $ele[9] && $ele[8] > $ele[10] ) {$base[$i] = "C"}
    if ($ele[9] > $ele[7] && $ele[9] > $ele[8] && $ele[9] > $ele[10] ) {$base[$i] = "G"}
    if ($ele[10] > $ele[7] && $ele[10] > $ele[8] && $ele[10] > $ele[9] ) {$base[$i] = "U"}
}
}

```

```

        print S "$base[$i]\t$ele[3]\t$ele[4]\t$ele[5]\t$ele[6]\n";
    }
close(S);
$scale = 1063.0/420.0;
open(I,">instru");
print I "set terminal svg\n";
print I "set output \"$outfile\"\n";
print I "set title \"$titulo\"\n";
    print I "set key right\n";
print I "set grid y\n";
print I "set style data histograms\n";
print I "set style histogram rowstacked\n";
print I "set boxwidth 0.5\n";
print I "set style fill solid 1.0 border -1\n";
print I "set ytics $divisiones nomirror\n";
print I "set yrange [:$mayor]\n";
print I "set ylabel \"Percentage (%)\"\n";
print I "set ytics $divisiones\n";
print I "plot \"datos_para_grafica.txt\" using 2 t \"A\", \"\" using 3 t \"C\", \"\" using 4 t \"G\", \"\"
using 5:xtic(1) t \"T\"\n";
close(I);
system "gnuplot < instru";
}

```

4) Analizar la salida de RNAhybrid y convertirla en una tabla de frecuencias por posición.

```

%rel;
$file_to_open= "colitas_30_16S_new5";
open IN, $file_to_open or die "Cant read $file_to_open\n";
while(<IN>){
    chomp;

```

```

    $rel{(split)[0]}= length((split)[1]);
}
foreach $linea (split(/\n/,`cat orgs_S21_verificadas`))
{
    ($org,$filo) = split(/\t/, $linea);
    ($kingdom,$phylum,$class) = (split(/./,$filo))[0,3,6];
    $llave = $kingdom."_".$phylum."_".$class;
    print "$org ... $llave\n";
    $n_orgs{$llave}++;
$filas{$llave}= 0;
    $infile = "Analisis_de_RNAhybrid_8dic/".$org.".txt";
    foreach $ren (split(/\n/,`cat $infile`))
    {
        @ele = split(/\t/, $ren);
        $p = $ele[1];
        foreach $i (2..10)
        {
            $acu{$llave}{$p}{$i} = $acu{$llave}{$p}{$i} + $ele[$i];
        }
        ($filas{$llave}= $p) && ($p > $filas{$llave});
    }
}
foreach $llave (sort %n_orgs)
{
    print "$llave ..... $n_orgs{$llave} ....\n";
    if ($n_orgs{$llave} > 0)
    {
        $outfile = "Analisis_de_RNAhybrid_por_clase_filogenetica_8dic/".$llave.".txt";
        open(S,">$outfile");
    }
}

```

```

foreach $i (1..$filas{$llave})
{
    print S "$llave\t$i\t";
    foreach $j (2..10)
    {
        $n = int($acu{$llave}[$i][$j]* 1000000/$n_orgs{$llave}+0.5) / 1000000;
        print S "$n\t";
    }
    print S "\n";
}
close(S);
} }

```

8. BIBLIOGRAFÍA

- Abolbaghaei, A., Silke, J. R., & Xia, X. (2017). How Changes in Anti-SD Sequences Would Affect SD Sequences in and. *G3*, 7(5), 1607–1615. <https://doi.org/10.1534/g3.117.039305>
- Accetto, T., & Avguštin, G. (2011). Inability of *Prevotella bryantii* to form a functional Shine-Dalgarno interaction reflects unique evolution of ribosome binding sites in Bacteroidetes. *PloS One*, 6(8), e22914. <https://doi.org/10.1371/journal.pone.0022914>
- Amin, M. R., Yurovsky, A., Chen, Y., Skiena, S., & Fitcher, B. (2018). Re-annotation of 12,495 prokaryotic 16S rRNA 3' ends and analysis of Shine-Dalgarno and anti-Shine-Dalgarno sequences. *PloS One*, 13(8), e0202767. <https://doi.org/10.1371/journal.pone.0202767>
- Baez, W. D., Roy, B., McNutt, Z. A., Shatoff, E. A., Chen, S., Bundschuh, R., & Fredrick, K. (2019). Global analysis of protein synthesis in *Flavobacterium johnsoniae* reveals the use of Kozak-like sequences in diverse bacteria. *Nucleic Acids Research*, 47(20), 10477–10488. <https://doi.org/10.1093/nar/gkz855>
- Bailey, T. L., Boden, M., Buske, F. A., Frith, M., Grant, C. E., Clementi, L., Ren, J., Li, W. W., &

- Noble, W. S. (2009). MEME SUITE: tools for motif discovery and searching. *Nucleic Acids Research*, 37(Web Server issue), W202–W208. <https://doi.org/10.1093/nar/gkp335>
- Bailey, T. L., Williams, N., Misleh, C., & Li, W. W. (2006). MEME: discovering and analyzing DNA and protein sequence motifs. *Nucleic Acids Research*, 34(Web Server issue), W369–W373. <https://doi.org/10.1093/nar/gkl198>
- Benelli, D., Maone, E., & Londei, P. (2003). Two different mechanisms for ribosome/mRNA interaction in archaeal translation initiation. *Molecular Microbiology*, 50(2), 635–643. <https://doi.org/10.1046/j.1365-2958.2003.03721.x>
- Betlach, M., Friedman, J., Boyer, H. W., & Pfeifer, F. (1984). Characterization of a halobacterial gene affecting bacterio-opsin gene expression. *Nucleic Acids Research*, 12(20), 7949–7959. <https://doi.org/10.1093/nar/12.20.7949>
- Boni, I. V., Isaeva, D. M., Musychenko, M. L., & Tzareva, N. V. (1991). Ribosome-messenger recognition: mRNA target sites for ribosomal protein S1. *Nucleic Acids Research*, 19(1), 155–162. <https://doi.org/10.1093/nar/19.1.155>
- Brock, J. E., Pourshahian, S., Giliberti, J., Limbach, P. A., & Janssen, G. R. (2008). Ribosomes bind leaderless mRNA in Escherichia coli through recognition of their 5'-terminal AUG. *RNA*, 14(10), 2159–2169. <https://doi.org/10.1261/rna.1089208>
- Condò, I., Ciammaruconi, A., Benelli, D., Ruggero, D., & Londei, P. (1999). Cis-acting signals controlling translational initiation in the thermophilic archaeon Sulfolobus solfataricus. *Molecular Microbiology*, 34(2), 377–384. <https://doi.org/10.1046/j.1365-2958.1999.01615.x>
- Duval, M., Korepanov, A., Fuchsbaauer, O., Fechter, P., Haller, A., Fabbretti, A., Choulier, L., Micura, R., Klaholz, B. P., Romby, P., Springer, M., & Marzi, S. (2013). Escherichia coli ribosomal protein S1 unfolds structured mRNAs onto the ribosome for active translation initiation. *PLoS Biology*, 11(12), e1001731.

- <https://doi.org/10.1371/journal.pbio.1001731>
- Espejo, R. T., & Plaza, N. (2018). Multiple Ribosomal RNA Operons in Bacteria; Their Concerted Evolution and Potential Consequences on the Rate of Evolution of Their 16S rRNA. *Frontiers in Microbiology*, 9, 1232. <https://doi.org/10.3389/fmicb.2018.01232>
- Estrada, K., & Merino, E. (2020). *Ribbon (Ribosomal Bona-fide)* (Version 1.2.0) [Linux]. <https://doi.org/10.5281/zenodo.7796827>
- Green, R., & Noller, H. F. (1997). Ribosomes and translation. *Annual Review of Biochemistry*, 66, 679–716. <https://doi.org/10.1146/annurev.biochem.66.1.679>
- Grill, S., Gualerzi, C. O., Londei, P., & Bläsi, U. (2000). Selective stimulation of translation of leaderless mRNA by initiation factor 2: evolutionary implications for translation. *The EMBO Journal*, 19(15), 4101–4110. <https://doi.org/10.1093/emboj/19.15.4101>
- Gruber, A. R., Lorenz, R., Bernhart, S. H., Neuböck, R., & Hofacker, I. L. (2008). The Vienna RNA websuite. *Nucleic Acids Research*, 36(Web Server issue), W70–W74. <https://doi.org/10.1093/nar/gkn188>
- Hussain, T., Llácer, J. L., Wimberly, B. T., Kieft, J. S., & Ramakrishnan, V. (2016). Large-Scale Movements of IF3 and tRNA during Bacterial Translation Initiation. *Cell*, 167(1), 133–144.e13. <https://doi.org/10.1016/j.cell.2016.08.074>
- Jha, V., Roy, B., Jahagirdar, D., McNutt, Z. A., Shatoff, E. A., Boleratz, B. L., Watkins, D. E., Bundschuh, R., Basu, K., Ortega, J., & Fredrick, K. (2021). Structural basis of sequestration of the anti-Shine-Dalgarno sequence in the Bacteroidetes ribosome. *Nucleic Acids Research*, 49(1), 547–567. <https://doi.org/10.1093/nar/gkaa1195>
- Klappenbach, J. A., Dunbar, J. M., & Schmidt, T. M. (2000). rRNA operon copy number reflects ecological strategies of bacteria. *Applied and Environmental Microbiology*, 66(4), 1328–1333. <https://doi.org/10.1128/AEM.66.4.1328-1333.2000>
- Komarova, A. V., Chufistova, L. S., Supina, E. V., & Boni, I. V. (2001). [Extensive

- complementarity of the Shine-Dalgarno region and 3'-terminal sequence of 16S ribosomal RNA is inefficient for translation in vivo]. *Bioorganicheskaya Khimiya*, 27(4), 282–290. <https://doi.org/10.1023/a:1011356520576>
- Letunic, I., & Bork, P. (2016). Interactive tree of life (iTOL) v3: an online tool for the display and annotation of phylogenetic and other trees. *Nucleic Acids Research*, 44(W1), W242–W245. <https://doi.org/10.1093/nar/gkw290>
- McNutt, Z. A., Gandhi, M. D., Shatoff, E. A., Roy, B., Devaraj, A., Bundschuh, R., & Fredrick, K. (2021). Comparative Analysis of anti-Shine-Dalgarno Function in and. *Frontiers in Molecular Biosciences*, 8, 787388. <https://doi.org/10.3389/fmolb.2021.787388>
- McNutt, Z. A., Roy, B., Gemler, B. T., Shatoff, E. A., Moon, K.-M., Foster, L. J., Bundschuh, R., & Fredrick, K. (2023). Ribosomes lacking bs21 gain function to regulate protein synthesis in *Flavobacterium johnsoniae*. *Nucleic Acids Research*, 51(4), 1927–1942. <https://doi.org/10.1093/nar/gkad047>
- Moll, I., Grill, S., Gualerzi, C. O., & Bläsi, U. (2002). Leaderless mRNAs in bacteria: surprises in ribosomal recruitment and translational control. *Molecular Microbiology*, 43(1), 239–246. <https://doi.org/10.1046/j.1365-2958.2002.02739.x>
- Nakagawa, S., Niimura, Y., & Gojobori, T. (2017). Comparative genomic analysis of translation initiation mechanisms for genes lacking the Shine-Dalgarno sequence in prokaryotes. *Nucleic Acids Research*, 45(7), 3922–3931. <https://doi.org/10.1093/nar/gkx124>
- Nakagawa, S., Niimura, Y., Miura, K.-I., & Gojobori, T. (2010). Dynamic evolution of translation initiation mechanisms in prokaryotes. *Proceedings of the National Academy of Sciences of the United States of America*, 107(14), 6382–6387. <https://doi.org/10.1073/pnas.1002036107>
- Nawrocki, E. P. (2009). *Structural RNA Homology Search and Alignment Using Covariance Models*.

- https://books.google.com/books/about/Structural_RNA_Homology_Search_and_Align.html?hl=&id=15-BAQAACAAJ
- Omotajo, D., Tate, T., Cho, H., & Choudhary, M. (2015). Distribution and diversity of ribosome binding sites in prokaryotic genomes. *BMC Genomics*, *16*(1), 604.
<https://doi.org/10.1186/s12864-015-1808-6>
- Posada, D. (2008). jModelTest: phylogenetic model averaging. *Molecular Biology and Evolution*, *25*(7), 1253–1256. <https://doi.org/10.1093/molbev/msn083>
- Rehmsmeier, M., Steffen, P., Hochsmann, M., & Giegerich, R. (2004). Fast and effective prediction of microRNA/target duplexes. *RNA*, *10*(10), 1507–1517.
<https://doi.org/10.1261/rna.5248604>
- Scharff, L. B., Childs, L., Walther, D., & Bock, R. (2011). Local absence of secondary structure permits translation of mRNAs that lack ribosome-binding sites. *PLoS Genetics*, *7*(6), e1002155. <https://doi.org/10.1371/journal.pgen.1002155>
- Shine, J., & Dalgarno, L. (1975). Determinant of cistron specificity in bacterial ribosomes. *Nature*, *254*(5495), 34–38. <https://doi.org/10.1038/254034a0>
- Srivastava, A., Gogoi, P., Deka, B., Goswami, S., & Kanaujia, S. P. (2016). In silico analysis of 5'-UTRs highlights the prevalence of Shine-Dalgarno and leaderless-dependent mechanisms of translation initiation in bacteria and archaea, respectively. *Journal of Theoretical Biology*, *402*, 54–61. <https://doi.org/10.1016/j.jtbi.2016.05.005>
- Starmer, J., Stomp, A., Vouk, M., & Bitzer, D. (2006). Predicting Shine-Dalgarno sequence locations exposes genome annotation errors. *PLoS Computational Biology*, *2*(5), e57.
<https://doi.org/10.1371/journal.pcbi.0020057>
- Tolstrup, N., Sensen, C. W., Garrett, R. A., & Clausen, I. G. (2000). Two different and highly organized mechanisms of translation initiation in the archaeon *Sulfolobus solfataricus*. *Extremophiles: Life under Extreme Conditions*, *4*(3), 175–179.

<https://doi.org/10.1007/s007920070032>

Torarinsson, E., Klenk, H.-P., & Garrett, R. A. (2005). Divergent transcriptional and translational signals in Archaea. *Environmental Microbiology*, 7(1), 47–54.

<https://doi.org/10.1111/j.1462-2920.2004.00674.x>

Tzareva, N. V., Makhno, V. I., & Boni, I. V. (1994). Ribosome-messenger recognition in the absence of the Shine-Dalgarno interactions. *FEBS Letters*, 337(2), 189–194.

[https://doi.org/10.1016/0014-5793\(94\)80271-8](https://doi.org/10.1016/0014-5793(94)80271-8)

Wei, Y., Silke, J. R., & Xia, X. (2017). Elucidating the 16S rRNA 3' boundaries and defining optimal SD/aSD pairing in *Escherichia coli* and *Bacillus subtilis* using RNA-Seq data.

Scientific Reports, 7(1), 17639. <https://doi.org/10.1038/s41598-017-17918-6>

Wen, J.-D., Kuo, S.-T., & Chou, H.-H. D. (2021). The diversity of Shine-Dalgarno sequences sheds light on the evolution of translation initiation. *RNA Biology*, 18(11), 1489–1500.

<https://doi.org/10.1080/15476286.2020.1861406>



UNIVERSIDAD AUTÓNOMA DEL
ESTADO DE MORELOS



DRA. LINA ANDREA RIVILLAS ACEVEDO
COORDINADORA DEL POSGRADO EN CIENCIAS
Presente

Atendiendo a la solicitud para emitir DICTAMEN sobre la revisión de la tesis titulada: “Análisis in silico de las secuencias de unión a ribosoma que definen el inicio de la traducción en procariontes”, que presenta el alumno **M. en C. Karel Johan Estrada Guerra** (10021247) para obtener el título de Doctor en Ciencias.

Director de tesis: Dr. Enrique Merino Pérez
Unidad Académica: Instituto de Investigación en Ciencias Básicas y Aplicadas (IICBA)

Nos permitimos informarle que nuestro voto es:

NOMBRE	DICTAMEN	FIRMA
Dr. Ramón Antonio González García Conde CIDC-UAEM	APROBADO	
Dr. Ramón Alberto Batista García CIDC-UAEM	APROBADO	
Dr. Guillermo Gosset Lagarda IBT - UNAM	APROBADO	
Dr. Miguel Ángel Carlos Cevallos Gaos CCG - UNAM	APROBADO	
Dra. Verónica Mercedes Narvárez Padilla CIDC-UAEM	APROBADO	
Dra. María Eugenia Núñez Valdez CIDC-UAEM	APROBADO	
Dr. Edgar Dantán González CEIB - UAEM	APROBADO	





UNIVERSIDAD AUTÓNOMA DEL
ESTADO DE MORELOS

Se expide el presente documento firmado electrónicamente de conformidad con el ACUERDO GENERAL PARA LA CONTINUIDAD DEL FUNCIONAMIENTO DE LA UNIVERSIDAD AUTÓNOMA DEL ESTADO DE MORELOS DURANTE LA EMERGENCIA SANITARIA PROVOCADA POR EL VIRUS SARS-COV2 (COVID-19) emitido el 27 de abril del 2020.

El presente documento cuenta con la firma electrónica UAEM del funcionario universitario competente, amparada por un certificado vigente a la fecha de su elaboración y es válido de conformidad con los LINEAMIENTOS EN MATERIA DE FIRMA ELECTRÓNICA PARA LA UNIVERSIDAD AUTÓNOMA DE ESTADO DE MORELOS emitidos el 13 de noviembre del 2019 mediante circular No. 32.

Sello electrónico

MIGUEL ÁNGEL CARLOS CEVALLOS GAOS | Fecha:2023-11-01 13:59:53 | Firmante

t7NuuAUSGEROqGwMh7RjSdiYCNZEX14f5mDBtRg5dCIma0jiz/Pr+h4LWlnL0w4x4TYLN0Vs4pFmVl4VAuZKHE+rUdJmF5DFhCohAdtUs/44aF4dR4PLU/Rp0i+VMtGUW09oXEPSweUJUOPbiwfbesOe+n0i4yT4I8gl997yo3dVzN6LC36kBOWANza7Fbf+6N/4RH5oG/ryqUtA2QvBLx4BLj0mHnmHHnAF6h66sqQxdPflX0Nfr1w/WjpuVzPbGRZDTwvquP/aGLexuFc1T9NigAlFYRNWdwQCmqHnjclYX/KzZVKK9INXN+c6CXTuPECcAn1FDDLrmZGhzgWg==

VERONICA MERCEDES NARVAEZ PADILLA | Fecha:2023-11-01 14:59:01 | Firmante

L4hTFC036867+CYEno88nvJC9Vs3/B1vbsu7K5Rek64EVwW9AmPnTWjD++a7SwvRQJBYNJ4YUuki583Nk/iLW1GuesOP6FmS7ok/06tzD6XwBF8VQ+QYz/ciDjmlnE5882LxvELoZCOPdDtk/aCyc5gOhFtbPlwvcwjlFd1vg0TXpK+PiXHzDjWtwQcQASJT7nL/FRK9VwWLFzX+nzirmGcXDxdvXph3/BYw8n0hVT+cgU4qNFKXvQQiJTIHRUpf28qe1dExfykaZi9NiyOcyZxeNrsyR54nQbZtANQhK796vKnjKNijDtGv1Wlc7GT3s149KKNczIkY5vQ==

RAMON ALBERTO BATISTA GARCIA | Fecha:2023-11-01 15:38:43 | Firmante

aCd4kAcZyXkvGA6PAXUYT1vhcGEn7wLUqnnxOxwSwbS6xdKJkpeXzaLUUFzIrwY/cRMDEgrqIGmOSH+YppB1/cl5Xoo7bITbpGUUKLzTY6uJQHwn03Jn6d3Rcu+TRuLAXwkbPTcgNjC6v7xPKtDACT/+uOXuOoXGY+K8XgZBK4WUrgQacoByWyt3PJHAdXN8ByFojT7pUQepnJEjWJRR814pNDXIML9ZL114pDSPGzuM4RwF1p4NqouVUWWMKUovAf7n0DQdxetQQRnEpONNxx0kBl3zB7npDhtPdDhphHn3DPvic4g9006a1O2mkieDRke0YgVJlwPk9FWg==

GUILHERMO GOSSET LAGARDA | Fecha:2023-11-01 18:37:34 | Firmante

P4hQC9GJpXmXfWxdS/ICqh9OUZnJmSj+TX7/Xgh7f5r8EW3nMGhyYrRoERiABZiG4NQ+omg2Uf9K1B7sSibXZ/EO2SIXrdxr6GB8m3ur+T+Xyl7123+aO9w1SDEePM6NdwA2yLUUIF7hW/fhymz6tNqWhYOhXzn/LsctIP3GM7maVhGcPR+tnS0z8M5xQliHmav0UvWVNdEMdPLrLiOrsM7R5SBKXGylmNpDFY6Hata2ASMUWc9Nfy2UMwug5rJPB83zOSr8kujDu7IL6ZLaGanAARul3wK3kwxLLs2XWNX2vK0RNxKdLs5Ru/DF2MHZSm+2AQGGpBq0y3U4/w==

MARIA EUGENIA NUÑEZ VALDEZ | Fecha:2023-11-02 19:06:55 | Firmante

RzNlx+oYpc3a0sA14p1yAF5cZ/B5AbiUaJiVkJ7FVWHBwAEuNRRiZgr+bonU540LsmMnB/vHPBaH6qGJPeMqCbAtW8dQNvzlhG5GQhm2LjEBx8i6dE0B2ggF1kzBFoCcJGoaDBW0/GkBlm3TVkjNmY7zLvr+KvYJ58fT6xdhQQWUAZiPnD4pAmxzB3pV+MbDrr8UfGwVU2zcDo1nWk5RrgcbN1i3BOfvbsw97rp3bkTS74fYLDZF+YpTfDgPlzFe7a76rFRhWCF83EMS16+YhJu+15abH2YUfxCckXE7AUeusSyavpY/MASebet4hUfSgU1wetbwg1kgzjg==

RAMON ANTONIO GONZALEZ GARCIA CONDE | Fecha:2023-11-06 09:35:40 | Firmante

Tq3tHYX3mivJrxm/cGhVuELXOwqH4fFXVmhDAVDvtDAQYn3Zgk14JbQCh4mpTXB7KT+HxI9FAsAY950sqV+TcRdhtni4IDcf9xHXG3EO4Xc06LFMJ+gxKfhhnK1jrsIID2sjkFoRnhFsi7CpCpeY/yhIFGyGyFuM8CrkhDrTijO3Gzrl0G4s6QdBLfa8c8Lb05VHGqr3NR5kUR5qYQed2P+Pujb9IRzVRpE+xcroIbUpE4I7ACan3+KCKWHufapvdy5GTnvORk3NIbwqldnSNPFxw14Pr6iJncEKOTRuAF8ono2FjEi0UImSZ2ly13Zgq01eUnALbpmK3NsQA==

EDGAR DANTAN GONZALEZ | Fecha:2023-11-06 09:56:11 | Firmante

mkVxf7fLgo/WqcBcKITB6JA/AVBdu95CrEgA5RL/zYbRjtD/rbkwEJnHF9SUMhsTCW82f2uj+Oe3M4dc1CDRchXftny4b5jO1K2bo16iPb15QeHoqRIMLgZBjaKSVyIQ9JOBqwrS35yqu4jCJJ7AQoD187Zblb2Osoi/cxWDXJV0zxiZo2jOaT66pPFVfBWabDTAcagmOZuDawnkHTpsQT8NBdGaVHEZH5O1W6/wH47xt4N8cwKNKZmyA+90Jn2iPGE/buDY9VJpcsPjYMkjcoduGleUrhNApjZEydsdWYalEG2x2kRT+Gt+C457d1TTbHTa+d0jYh625bUdn2I9fQ==

Puede verificar la autenticidad del documento en la siguiente dirección electrónica o escaneando el código QR ingresando la siguiente clave:



MJ41pVL92

<https://efirma.uaem.mx/noRepudio/6FboUJUqstRFRu0ZiH4onLSnKqM6iamb>

