



**UNIVERSIDAD NACIONAL AUTÓNOMA
DE MÉXICO**

FACULTAD DE CIENCIAS

**Filogenia molecular de Csp's: proteínas de un sistema de
respuesta al frío**

T E S I S

QUE PARA OBTENER EL TÍTULO DE:

BIÓLOGA

P R E S E N T A :

MARIANA MUÑOZ ARGOTT



**DIRECTORA DE TESIS:
M. en C. FABIOLA RAMÍREZ CORONA
2014**



Universidad Nacional
Autónoma de México



UNAM – Dirección General de Bibliotecas
Tesis Digitales
Restricciones de uso

DERECHOS RESERVADOS ©
PROHIBIDA SU REPRODUCCIÓN TOTAL O PARCIAL

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

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

1. Datos de la alumna
Muñoz
Argott
Mariana
54 43 10 86
Universidad Nacional Autónoma de México
Facultad de Ciencias
Biología
302044270

2. Datos de la Tutora
M. en C.
Fabiola
Ramírez
Corona

3. Datos de la sinodal 1
M. en C.
Sara Ernestina
Islas
Graciano

4. Datos del sinodal 2
M. en C.
Alfonso José
Vilchis
Peluyera

5. Datos del sinodal 3
Dr.
Arturo Carlos II
Becerra
Bracho

6. Datos del sinodal 4
Dr.
Bernardo
Pérez
Zamorano

7. Datos del trabajo escrito
Filogenia molecular de Csp's
Proteínas de un sistema de respuesta al frío
59 p
2014

RESÚMEN

Todos los organismos cuentan con la capacidad para adaptarse a cambios en su medio, dichos cambios van desde la disposición de nutrientes hasta parámetros fisiológicos como presión osmótica y temperatura. Particularmente las bacterias, ante una disminución drástica de la temperatura, cuentan con diversas estrategias para enfrentar este cambio, entre éstas se encuentran unas proteínas denominadas proteínas cold shock (CSPs). Las CSPs tienen una estructura general de barril β plegado compuesto de 5 cadenas β antiparalelas y funcionan como chaperonas de RNAs lo que hace que estén presentes en prácticamente todos los pasos de la expresión genética incluyendo la transcripción y la traducción. Se han encontrado secuencias similares a las de las CSPs de bacterias en proteínas de Archaeas y Eucariones a lo que se le denomina el dominio cold shock (CSD). En este trabajo se pretende resaltar la importancia adaptativa y evolutiva de las proteínas cold shock y del dominio cold shock, por lo que se realizó una búsqueda bibliográfica y bioinformática extensa empleando las bases de datos que existen en la actualidad, se construyó una base de datos con secuencias de CSPs de bacterias, archaeas y eucariontes, se construyó una filogenia molecular utilizando secuencias de los dominios Bacteria, Archaea y Eucaria con herramientas bioinformáticas actualizadas, se obtuvo el número de parálogos de las 9 proteínas CSPs de *E. coli* en los genomas completos de bacterias psicrófilas y psicrotolerantes. Se obtuvo una base de datos con 240 secuencias, de las cuales 153 secuencias pertenecen a bacterias, 57 a eucariontes y 30 a archaeas. Se construyó una filogenia molecular en la cual se pudo observar una clara separación por *phylla* entre las secuencias de los tres dominios y en menor grado un agrupamiento por tipo de proteína. Se obtuvo el número de parálogos de las 9 proteínas CSPs de *E. coli* en los genomas completos de bacterias psicrófilas y psicrotolerantes. Se pudo concluir que las proteínas CSPs y el dominio CSD se encuentran altamente conservados en los 3 dominios de la vida y derivado a este análisis bibliográfico y bioinformático proponemos que el dominio CSD es muy importante en el mecanismo de unión al RNA y que las proteínas CSPs son de suma importancia para la adaptación al frío de las bacterias aún si son psicrófilas y psicrotolerantes.

ÍNDICE

INTRODUCCIÓN	6
ANTECEDENTES.....	7
Bajas temperaturas y los organismos.....	7
Respuesta al choque frío en bacterias	8
Proteínas cold shock (CSPs).....	9
Estructura de las CSPs.....	9
Miembros de la familia de CSPs	10
Respuesta de las CSPs al choque frío	13
Regulación de la expresión del gen cspA	14
Reinicio del crecimiento después del choque frío.....	17
Dominio Cold Shock.....	18
Dominio Cold Shock en eucariontes	18
Familia Y-box	18
Proteína Lin 28.....	19
Proteína UNR.....	20
Proteína glycine- rich	21
Proteínas Cold Shock en Archaeas	23
Evolución de la proteína Cold Shock y el dominio Cold Shock	23
JUSTIFICACIÓN	24
OBJETIVOS	25
Objetivo general.....	25
Objetivos particulares.....	25
METODOLOGÍA.....	26
Diseño, síntesis y elución de los oligonucleótidos.....	26
Amplificación del gen cspA por Reacción en Cadena de la Polimerasa	26
Secuenciación.....	27
Búsqueda bibliográfica.....	27
Elaboración de la Base de datos	27
Construcción de la filogenia molecular	29
Parálogos de CSPs en los genomas completos de Bacterias Psicrófilas y Psicrotolerantes.	30
RESULTADOS.....	31

Búsqueda bibliográfica.....	31
Base de datos.....	31
Alineamiento de secuencias de bacterias, archaeas y eucariontes	32
Reconstrucción filogenética	33
Búsqueda de parálogos de CSPs en los genomas completos de Bacterias Psicrófilas y Psicrotolerantes.	39
DISCUSIÓN	44
CONCLUSIONES	47
REFERENCIAS.....	49

INTRODUCCIÓN

Los proyectos de secuenciación genómica en los diversos organismos que habitan el planeta han generado que se tenga un amplio repertorio de proteínas conocidas. El análisis de estos datos no sólo nos ha ayudado en el entendimiento de los procesos por los cuales se forman nuevas proteínas sino también a trazar la relación evolutiva de éstas (Chothia & Gough, 2009). Las proteínas pueden estar formadas por 1, 2 o más dominios (término empleado como unidad estructural y evolutiva). Los dominios cuyos miembros descienden de un ancestro en común se pueden agrupar en familias o superfamilias (Chothia & Gough, 2009).

El dominio cold shock (CSD) se ha encontrado en proteínas de bacterias, archaeas y eucariontes, el cual consiste en un barril β plegado compuesto de 5 cadenas β antiparalelas y sirve para facilitar la unión a ácidos nucleicos (Mihailovich *et al.*, 2010). El primer dominio cold shock (CSD) se encontró en las proteínas cold shock (CSPs) de bacterias (Mihailovich *et al.*, 2010). Varios miembros de esta familia de proteínas son importantes para la adaptación de las bacterias ante una disminución drástica de la temperatura, así como en condiciones normales de crecimiento (Ermolenko & Makhatazde, 2002; Graumann & Marahiel, 1998; Horn *et al.*, 2007). Las CSPs de bacterias se unen a ácidos nucleicos de cadena simple y funcionan como chaperonas de RNAs lo que hace que estén presentes en prácticamente todos los pasos de la expresión genética incluyendo la transcripción, la traducción y la entrega de RNAs (Mihailovich *et al.*, 2010).

Desde el descubrimiento de las proteínas CSPs, el dominio cold shock (CSD) se ha encontrado en un gran número de proteínas de bacterias, eucariontes y también en archaeas, aunque en menor número, por lo que en el presente trabajo se pretende enfatizar la importancia del dominio cold shock y de las proteínas CSPs así como de su importancia evolutiva.

ANTECEDENTES

Bajas temperaturas y los organismos

Todos los organismos cuentan con la capacidad para adaptarse a cambios en su medio, desde cambios en la disposición de nutrientes hasta parámetros fisiológicos como presión osmótica y temperatura. La temperatura ambiental es un factor limitante para los organismos, tanto en su distribución como en sus actividades (Horn *et al.*, 2007). Ante una rápida disminución de la temperatura los parámetros fisicoquímicos de las células procariontes y eucariontes se ven sometidos a un estrés, por lo que hay cambios en la fluidez de la membrana, la cinética enzimática, la topología, la difusión de solutos y transporte, modificación en la estructura del DNA/RNA, en la transcripción y traducción de genes y el bloqueo de la función ribosomal (Morita, 1975; Phadtare, 2004; Thieringer, Jones, & Inouye, 1998).

En el caso de los animales cuando sufren congelación, el fluido extracelular tiende a congelarse primero, provocando que la concentración de solutos aumente, ocasionando la expulsión del agua intracelular, llegando a un punto en que la concentración intracelular es tan alta, que los procesos vitales se alteran.(Fanjul & Hiriart, 2008). En cambio, los organismos unicelulares sólo pueden responder de forma directa al cambio de temperatura (Horn *et al.*, 2007); para responder a estas condiciones nuevas, las bacterias cambian su expresión genética (Graumann & Marahiel, 1996) deteniendo su crecimiento y proliferación, además hay cambios en la expresión de proteínas (Jones, Bogelen, & Neidhardt, 1987).

Respuesta al choque frío en bacterias

La fase inicial de la respuesta a un choque frío es decir, cuando la temperatura baja rápidamente, de 37 °C a 15 °C, es llamada fase de aclimatación (Jones *et al.*, 1987), ésta se caracteriza por tener altos niveles de síntesis de proteínas CIPs, por sus siglas en inglés (cold induced proteins) (Fig. 1), como son NusA, la PNPasa, la RNasa R, RecA, los factores de iniciación IF1,IF2 e IF3, la DNA girasa y la helicasa Csd, entre otras (Chen & Deutscher, 2005; Giuliodori, 2004; Graumann & Marahiel, 1996; Jones *et al.*, 1987; Xia, Ke, & Inouye, 2001), las cuales se acumulan durante el choque frío. Durante la fase de aclimatación de *E. coli* se sobreexpresan 24 proteínas de las cuales 14 se expresan de forma transitoria (Graumann & Marahiel, 1996). Del grupo de proteínas CIPs se deriva un subgrupo, la familia de proteínas cold shock (CSPs). Al final de la fase de aclimatación la síntesis de CIPs, entre ellas las CSPs, disminuye hasta niveles constantes y la producción de proteínas de estructura se retoma, entrando en la fase de crecimiento en frío (Horn *et al.*, 2007).

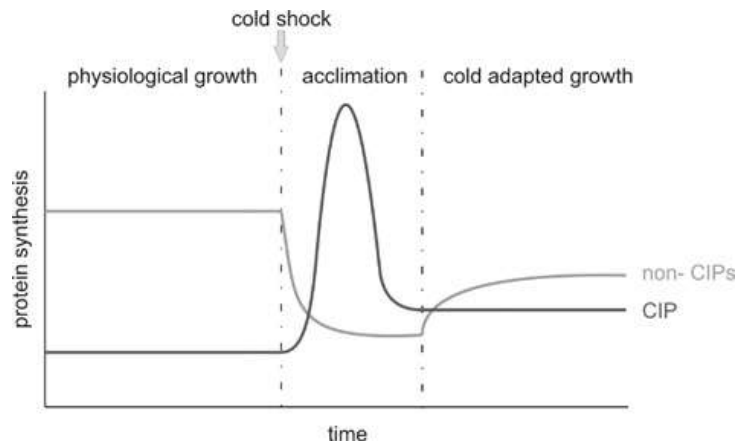


Fig. 1. Esquema de expresión de proteínas después de un choque frío. Muestra como la expresión de CIPs aumenta durante el choque frío mientras que el de las otras proteínas disminuye. Después, la expresión de CIPs declina y la de las otras proteínas se reactiva, empezando el periodo de adaptación al frío (tomado de Horn *et al.*, 2007).

Proteínas cold shock (CSPs)

La primer CSP fué identificada en 1987 (Jones *et al.*, 1987), codificada por uno de los genes del “cold-shock stimulon” en *Escherichia coli*. Desde entonces, se han identificado CSP en casi todas las bacterias psicrófilas, mesófilas, termófilas e hipertermófilas estudiadas, incluyendo *Thermotoga* (Perl *et al.*, 1998) y *Aquifex* (Deckert *et al.*, 1998), siendo *Helicobacter pylori* y *Mycoplasma genitalium* las únicas 2 en las que no se ha encontrado alguna CSP (Graumann & Marahiel, 1996). A pesar de esto, la mayoría de los trabajos se han realizado con las proteínas CSP de *Escherichia coli* y únicamente los trabajos de Yamanaka, (1999) y Graumann & Marahiel, (1998) contienen información de la evolución de la familias de las proteínas CSP de esta bacteria.

Estructura de las CSPs

Es una familia altamente conservada de proteínas pequeñas de bajo peso molecular, aproximadamente de 7.4 KDa (Lim, Thomas, & Cavicchioli, 2000; Perl *et al.*, 1998), y constan de un barril constituido por 5 cadenas β plegadas anti-paralelas ($\beta 1$ - $\beta 5$) (Fig.2), 2 de estas cadenas ($\beta 2$ y $\beta 3$ a partir del extremo N-terminal) incluyen 2 motivos altamente conservados denominados con las siglas RNP1 (cadena $\beta 2$, secuencia consenso KGYGFIEV) y RNP2 ($\beta 3$, secuencia consenso VFVHF)(Fig.2), las cuales forman el dominio típico conocido como “dominio de choque frío” (CSD, por sus siglas en ingles cold-shock domain) (Max *et al.*, 2007; Murata & Los, 1997; Weber & Marahiel, 2002) (Fig. 2). Los residuos aromáticos presentes en ambos motivos son esenciales para mediar la interacción con moléculas de RNA de cadena simple, que en condiciones de baja temperatura, estabilizan sus estructuras secundarias para que interactúen con el ribosoma de forma habitual, como los mRNA codificantes para las CIPs (entre ellas las propias CSPs) (Max *et al.*, 2007). Se ha demostrado que las CSPs también se unen a moléculas de DNA de cadena simple con mayor afinidad que con las moléculas de RNA (Weber *et al.*, 2001). Esta propiedad de unirse con alta afinidad a ácidos nucleicos permite un amplio espectro de

interacciones a las proteínas de esta familia, cuyos miembros poseen diversos niveles de afinidad de unión y especificidad de secuencia (Vigh, Maresca, & Harwood, 1998).

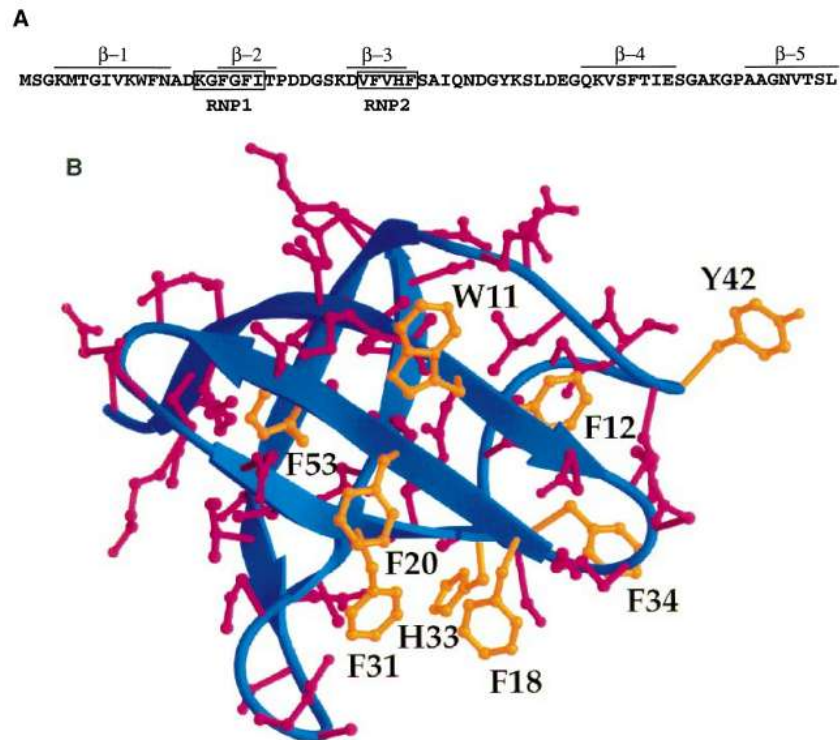


Fig. 2. Secuencia primaria y estructura terciaria de la CSPA.

A. Secuencia primaria de aminoácidos de la CSP. Las cadenas $\beta 1$, $\beta 2$, $\beta 3$, $\beta 4$ y $\beta 5$ se encuentran indicadas con una línea en la parte superior de la secuencia y los motivos RNP1 Y RNP2 están dentro de las cajas.

B. Estructura terciaria de CSPA. Los residuos aromáticos se encuentran coloreados de amarillo (tomado de K Yamanaka, Fang, & Inouye, 1998)

Miembros de la familia de CSPs

Las proteínas cold shock (CSPs) se encuentran ampliamente distribuidas en las bacterias Gram-positivas y Gram-negativas las cuales presentan una identidad mayor al 45% (Graumann *et al.*, 1997); a pesar de esto, los principales estudios se han realizado en las proteínas CSPs de *E. coli* debido a que se identificaron por primera vez en este organismo.

Escherichia coli posee una familia de proteínas CSPs, constituida por nueve miembros, a los cuales se les ha nombrado de forma alfabética desde la CSPA a la CSPI. Dichas proteínas comparten de 29 a 83% de similitud (Fig. 3 B) (Goldstein, Pollitt, & Inouye, 1990; Lee *et al.*, 1994; Nakashima *et al.*, 1996; Phadtare & Severinov, 2009; Phadtare *et al.*, 2009; Yamanaka *et al.*, 1998) y los genes *csp* correspondientes a cada proteína son duplicaciones del gen *cspA* y poseen funciones muy similares (Fig 3. A) (Bae, Jones, & Inouye, 1997).

Estas proteínas se agrupan en dos tipos: las de clase I y las de clase II. Las de la clase I son proteínas que se expresan poco a 37 °C pero su tasa de expresión aumenta drásticamente cuando baja la temperatura (CSPA, CSPB, CSPG y CSPI) (Phadtare & Severinov, 2009). Las de la clase II son proteínas que están presentes a 37 °C y que no son funcionales a 15 °C e incluyen a CSPC, CSPD, CSPE, CSPF y CSPH

Las CSPE y CSPC se han visto involucradas en la regulación de la transcripción y la condensación de los cromosomas (Hu *et al.*, 1996; Yamanaka *et al.*, 1994). También se encontró que CSPE presenta una actividad similar a CSPA en cuanto a la desnaturalización de ácidos nucleicos (Phadtare & Severinov, 2005). La CSPD se induce en condiciones de estrés nutricional (Yamanaka & Inouye, 1997). CSPA, CSPB, CSPG y CSPI son mayormente producidas en un choque frío (Lee *et al.*, 1994; Nakashima *et al.*, 1996; Wang, Yamanaka, & Inouye, 1999), mientras que la función de CSPF y CSPH es desconocida hasta el momento (Yamanaka *et al.*, 1998) y nadie ha trabajado con éstas hasta la fecha.

Ya que presentan funciones muy similares se puede pensar que las CSPs tienen una actividad redundante, y estudios de mutagénesis muestran que el fenotipo de sensibilidad al frío únicamente se obtiene por la inactivación genética simultánea de *cspA*, *cspB*, *cspG* y *cspI*. De modo interesante, este fenotipo se revierte por la expresión en *trans* de un homólogo estructural de CSPA, el dominio de unión al RNA S1, que está ampliamente distribuido en distintas proteínas de diversos organismos (Xia *et al.*, 2001).

De forma similar *Bacillus subtilis* posee una familia de proteínas CSPA-like formada por 3 proteínas homólogas a las CSPs de *E. coli* (CSPB, CSPC y CSPD) (Graumann *et al.*, 1996), las cuales son esenciales para el crecimiento a temperatura óptima así como para la adaptación a bajas temperaturas y sobrevivencia durante la fase estacionaria de esta bacteria (Graumann *et al.*, 1997). Las proteínas CSPB y CSPD incrementan su expresión ante el descenso de la temperatura mientras que la CSPC muestra un patrón de expresión asociado exclusivamente a bajas temperaturas (Schindler *et al.*, 1999). Las CSPs de *B. subtilis* muestran entre un 72% y 80% de similitud entre sí, y son de hecho capaces de complementarse unas a otras (Schindler *et al.*, 1999); estudios muestran que la presencia de al menos una de las CSP es esencial para la supervivencia de esta bacteria, incluso bajo condiciones óptimas de crecimiento (Weber & Marahiel, 2002).

A nivel de aminoácidos, CSPA tiene una identidad de 43% con otro dominio de choque frío (CSD) encontrado en la familia de proteínas eucariontes Y-box (Jiang, Hou, & Inouye, 1997). Esta familia de proteínas participa en el silenciamiento traduccional de los mRNAs maternos y se ha demostrado que el dominio CSD es el responsable de la actividad de unión a ácidos nucleicos (Bouvet, Matsumoto, & Wolffe, 1995). Lo anterior sugiere que estos dominios estructurales juegan un papel importante en el metabolismo del RNA y su traducción, y que probablemente estén relacionados evolutivamente (Xia *et al.*, 2001).

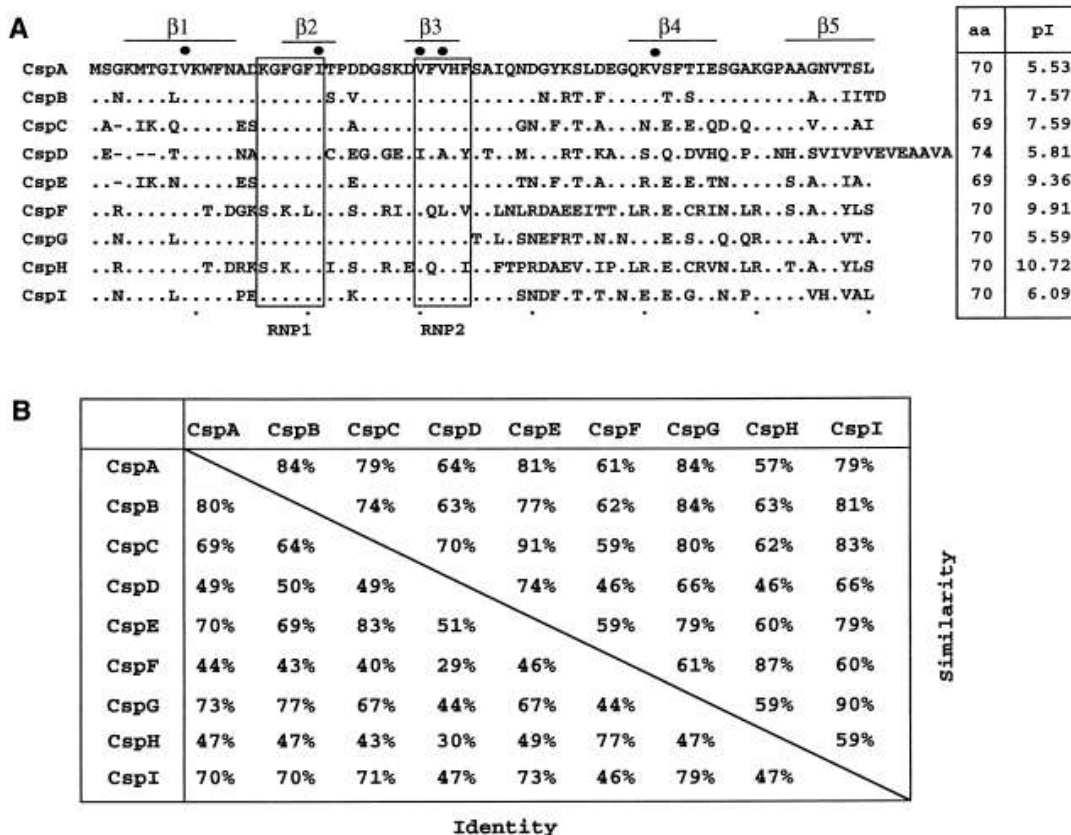


Fig 3. Familia de proteínas CSPs en *E.coli*. A) Alineamiento de las secuencias de CSPA, CSPB, CSPC, CSPD, CSPE, CSPF, CSPG, CSPH, CSPI. Dentro de los cuadros están los motivos RNP1 y RNP2 y las cadenas β1-β5 están indicadas con líneas en la parte superior. B) Similitud e identidad entre los miembros de la familia de proteínas CSPs comparadas contra la secuencia de 70 residuos de CSPA (tomado de Yamanaka *et al.*, 1998).

Respuesta de las CSPs al choque frío

En *E. coli* son cuatro las CSPs principales que se expresan en un choque frío, CSPA, CSPB, CSPG y CSPI (Phadtare & Severinov, 2005), siendo la más abundante la CSPA, la cual alcanza niveles superiores al 13% del total de proteínas sintetizadas. CSPA es una proteína citoplasmática con un peso de 7.4 KDa (Goldstein *et al.*, 1990; Jones *et al.*, 1987), esta proteína es más abundante en el choque frío pero también se encuentra presente durante el crecimiento exponencial a 37 °C con una concentración aproximada de 50 μM.

También los genes de CSPA se encuentran altamente transcritos pero el mRNA no se traduce efectivamente (Bae *et al.*, 1997; Mitta, Fang, & Inouye, 1997), además su tiempo de vida es mucho menor (5 segundos), a diferencia de cuando se encuentra en un choque frío que el mRNA es altamente estable y su tiempo de vida es de 12 segundos, esto provoca que la concentración de CSPA aumente a aproximadamente a 100 μ M (Brandi, Spurio, Gualerzi, & Pon, 1999). Adicionalmente consta de un 5'-UTR, que es inusualmente larga, de 159 pb, y que tiene una secuencia altamente conservada, llamada cold-box, lo que le facilita su transcripción en estas condiciones en comparación a los otros mRNAs celulares (Bae *et al.*, 1997; Jiang *et al.*, 1997; Mitta *et al.*, 1997). CSPA actúa como chaperona de RNAs, cuando la proteína llega a una concentración óptima se une cooperativamente a la estructura del RNA, desnaturalizándolo y facilitando su traducción o procesamiento post-transcripcional (Jiang *et al.*, 1997).

Regulación de la expresión del gen *cspA*

El gen *cspA* tiene un promotor que se activa tanto a 37 °C como a 15 °C (Fang *et al.*, 1997; Goldenberg *et al.*, 1997; Mitta *et al.*, 1997), y está equipado con dos secuencias características del gen. La primera es una secuencia rica en AT ubicada inmediatamente a la región -35 río arriba, esta secuencia se llama elemento "UP" (Fig. 4), la cual es reconocida directamente por la subunidad α de la RNA polimerasa (Goldenberg *et al.*, 1997; Mitta *et al.*, 1997), lo que le confiere una actividad fuerte de transcripción (Ross *et al.*, 1993). La segunda es una secuencia TGn situada río arriba inmediato a la región -10, esta secuencia, junto con la región -10 constituye lo que se llama la región -10 extendida (Kumar *et al.*, 1993) (Fig. 4). Esta dos secuencias son las responsables de que el gen *cspA* cuente con una actividad tanto a 37 °C como a 15 °C (Yamanaka, 1999), y se encuentran presentes en los genes *cspB*, *cspG* y *cspI* los cuales también se activan en un choque frío (Wang *et al.*, 1999).

Sin embargo, es importante tener en cuenta que el mRNA de los genes *cspA*, *cspB*, *cspG* y *cspI* tiene una región 5'-UTR (Untranslated, highly conserved region) inusualmente larga, de 159, 161, 156 y 145 bases respectivamente (Tanabe *et al.*, 1992; Wang *et al.*, 1999) (Fig. 4) y que es extremadamente inestable a 37 °C, teniendo una vida media de aproximadamente 12 segundos (Brandi *et al.*, 1996; Fang *et al.*, 1997; Goldenberg, Azar, & Oppenheim, 1996). Una vez que se presenta el choque frío el mRNA se torna estable, teniendo ahora una vida media de aproximadamente 20 minutos. Esta inestabilidad se ha visto que es provocada por la presencia de un sitio de reconocimiento a enzimas de restricción RNasas E **putativas**, situado río arriba junto a la secuencia SD (Fang *et al.*, 1997). Estas enzimas presentan la misma actividad catalítica a 37 °C como a 15 °C pero el mRNA a 15 °C forma estructuras secundarias que evitan la unión de estas enzimas (Hankins *et al.*, 2007). Sin embargo el mecanismo exacto por el cual el mRNA se estabiliza no se conoce aún (Yamanaka, 1999).

La 5' UTR cuenta también con una secuencia única altamente conservada en los genes de choque frío (*cspA*, *cspB*, *cspG* y *cspI*) llamada cold box, de 11 pares de bases (Bae *et al.*, 1997) (Fig. 4), la cual forma una estructura estable "stem-loop" (Jiang, Fang, & Inouye, 1996) y está involucrada en la represión del gen *cspA* (Bae *et al.*, 1997; Fang *et al.*, 1997).

La expresión del gen *cspA* también es regulada a nivel de su traducción (Phadtare, Alsina, & Inouye, 1999). El mRNA contiene un sitio de unión al ribosoma llamado Downstream Box (DB) (Lee *et al.*, 1994), de 14 pares de bases localizada río abajo del codón de inicio (Etchegaray & Inouye, 1999; Mitta *et al.*, 1997) (Fig. 4). Esta secuencia DB es complementaria a la región llamada secuencia anti-DB de la subunidad 16S del rRNA, este complejo es responsable de la iniciación de la traducción de proteínas CSPA (Sprengart, Fuchs, & Porter, 1996). Se ha comprobado que la DB es esencial para la traducción de proteínas del choque frío durante la fase de aclimatación mientras que la traducción de las otras proteínas están bloqueadas (Etchegaray & Inouye, 1999; Mitta *et al.*, 1997), además de que esta secuencia también se encuentra presente en otros genes inducidos en choque frío de clase I (*cspB*, *cspG* y *cspI*) (Mitta *et al.*, 1997).

Se ha propuesto que los ribosomas actúan como sensores fisiológicos del estrés ante un choque frío (VanBogelen & Neidhardt, 1990), cuando se presenta un choque frío se bloquea la traducción de proteínas. Después de la fase de aclimatación las proteínas del choque frío transforman a los ribosomas no adaptados al frío en ribosomas adaptados al frío, los cuales ahora pueden traducir mRNAs para proteínas que no sean de choque frío (Bae *et al.*, 1997; Jones & Inouye, 1996), por lo anterior, se puede decir que la traducción es la clave de la adaptación a las bajas temperaturas (Yamanaka, 1999).

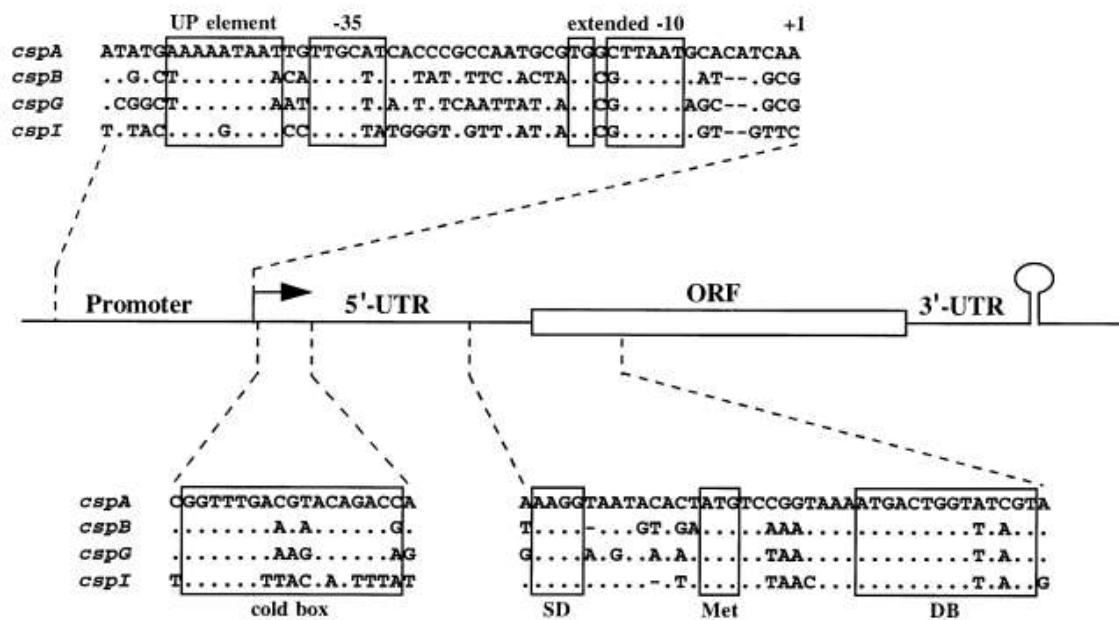


Fig. 4. Comparación de las secuencias de los genes de *cspA*, *cspB*, *cspG* y *cspI*. Las estructuras características (UP element, región -35, región -10 y cold box) del gen *cspA* y para la transcripción (Shine-Dalgarno, codón de inicio y Downstream Box), se encuentran dentro de las cajas (tomado de Yamanaka, 1999).

Reinicio del crecimiento después del choque frío

A la fecha se han descrito dos eventos de regulación post-transcripcional fundamentales para el reinicio del crecimiento celular después del choque frío, que son la degradación de los mRNAs de las CSPs mediante la helicasa CsdA (Khemici *et al.*, 2004) y la regulación de los niveles de la RNasa R. Estudios en *Y. enterocolitica*, *B. subtilis* y *E. coli* muestran que la actividad fosforolítica, principalmente de la PNPasa (RNasa de la familia de la RNasa PH, que es dependiente de fosfato y que forma parte del degradosoma de RNA), se induce y es esencial para el crecimiento óptimo a bajas temperaturas (Goverde *et al.*, 1998; Yamanaka & Inouye, 2001; Zhou & Deutscher, 1997). El crecimiento celular solamente se restablece cuando el mRNA de CSPA es degradado, independientemente de la acumulación de las chaperonas de RNA en la célula, y se ha descrito que la PNPasa interviene en la degradación específica del transcrito *cspA* (Neuhaus, Rapposch, Francis, & Scherer, 2000; Yamanaka & Inouye, 2001). A pesar de haberse identificado un degradosoma especializado para choque frío (Prud'homme-Généreux *et al.*, 2004), se ha encontrado que esta actividad específica se realiza por vía independiente del degradosoma (Yamanaka & Inouye, 2001). En las mutantes nulas de PNPasa se acumula el transcrito *cspA*, y dada su traducción preferente en frío, los ribosomas son secuestrados y la traducción del mRNA celular se inhibe, abatiendo el crecimiento bacteriano (Jiang *et al.*, 1996). Por otro lado, se ha observado que durante el choque frío en cepas de *E. coli* mutantes nulas de PNPasa, hay una degradación permanente del rRNA 23S, lo que genera ribosomas defectuosos que no presentan un ensamblaje adecuado (Zhou & Deutscher, 1997). Por todo esto, resulta evidente que la degradación fosforolítica es fundamental para el crecimiento celular a bajas temperaturas. Considerando la amplia distribución de la PNPasa (Bermúdez *et al.*, 2005), así como la ventaja energética de la fosforólisis del RNA respecto a la hidrólisis (Deutscher & Reuven, 1991), es posible que este mecanismo se encuentre altamente conservado en el árbol de la vida como un elemento de adaptación a bajas temperaturas.

Dominio Cold Shock

El dominio Cold Shock (CSD) es un dominio ancestral que se ha encontrado en varias proteínas de los tres dominios del árbol de la vida, y probablemente estuvo presente al comienzo de la evolución de la célula, hace 3.5 billones de años (Graumann & Marahiel, 1998). El dominio conserva la estructura de barril β plegable, descrita en las proteínas CSPs de las bacterias, a través de los distintos organismos y las funciones de las proteínas en las que se encuentra contenido (Mihailovich *et al.*, 2010) (fig. 5). Su función particular es, en la mayoría de los casos, de regulación post-transcripcional de genes a través de la unión a ácidos nucleicos (Mihailovich *et al.*, 2010).

Dominio Cold Shock en eucariontes

La lista de proteínas que contienen el dominio CSD o CSD *like* en eucariontes ha aumentado con el tiempo; dicha lista contiene proteínas que están implicadas en varios procesos relacionados con el metabolismo del RNA, esta diversidad funcional está proporcionada en parte por la combinación del dominio CSD con muchos otros dominios funcionales, así como por la alta gama de secuencias que son reconocidas por este dominio y su afinidad de reconocimiento, lo que le proporciona a éste un amplio repertorio de RNAs diana (Mihailovich *et al.*, 2010). A continuación se describirán algunas de las proteínas que contienen el dominio CSD:

Familia Y-box

La familia Y-box se encuentra en vertebrados e invertebrados y está compuesta por las proteínas denominadas DBPA (DNA binding protein A, también conocida como CSDA), DbpB (comúnmente conocida como factor de unión Y-box-1, YB-1) y DBPC

(frog germ-cell-specific o Y-box protein 2, FRGY2). Se ha demostrado que DBPA reprime la transcripción de varios genes, en ocasiones en conjunto con DbpB o YB-1 (Coles *et al.*, 1996). YB-1 es un componente principal del mensajero citoplasmático de ribonucleoproteínas (mRNPs), tiene funciones en el splicing (empalme) del pre-mRNA, en la estabilidad del mRNA y en la traducción de estas proteínas (Matsumoto & Wolffe, 1998). YB-1 contiene un único dominio CSD (Fig. 5) con preferencia por la unión a secuencias ricas en pirimidinas de una sola hebra (Kloks *et al.*, 2002). El dominio CSD es necesario para la importación nuclear de la proteína, así como su reconocimiento, pero no parece tener otras funciones para la proteína (Jurchott *et al.*, 2003).

La proteína YB-1 puede estimular o reprimir la traducción, dependiendo de su cantidad con respecto al RNA diana. Cuando existe una relación baja entre YB-1 y el mRNA, YB-1 promueve la iniciación de la traducción, mientras que un aumento de esta relación inhibe fuertemente la traducción (Evdokimova *et al.*, 1998; Pisarev *et al.*, 2002), relación parecida a la de las CSP en bacterias.

La proteína FRGY-2 es también un componente importante de RNPs almacenadas en oocitos del genero *Xenopus*. FRGY-2 empaqueta el mRNA en RNPs compactos que provocan la represión de la traducción y la estabilización del mRNA (Bouvet & Wolffe, 1994; Matsumoto *et al.*, 2003). Sin embargo el dominio CSD no se requiere para la represión de traducción (Matsumoto, Meric, & Wolffe, 1996), pero es necesario para el reconocimiento del RNA (Bouvet *et al.*, 1995).

Estas proteínas desempeñan papeles similares a las proteínas CSPs de las bacterias.

Proteína Lin 28

La proteína Lin 28 (abnormal cell lineage 28) es una proteína citoplasmática pequeña (~ 28 kDa) la cual posee un dominio CSD (Fig. 5) y dos dedos de Zinc (CCHC)

los cuales cooperan con el dominio CSD para el reconocimiento de los ácidos nucleicos (RNA) diana (Moss & Tang, 2003). Esta proteína se encuentra conservada en géneros como *Drosophila*, *Xenopus*, ratón y humano (Mihailovich *et al.*, 2010). Lin 28 regula el metabolismo de micro-RNAs (miRNA). Los miRNA son RNAs pequeños de aproximadamente 22 nucleótidos, que regulan la expresión genética mediante la hibridación a sitios que por lo general están ubicados en el extremo 3' del UTR de los RNAs diana. La interacción entre Lin 28 y los miRNAs es crítica para el desarrollo, la diferenciación y la progresión del cáncer (Nimmo & Slack, 2009).

Proteína UNR

El gen *unr* es esencial, estudios demuestran que los ratones que carecen de éste mueren a los 10 días del desarrollo embrionario (Doniger *et al.*, 1992), y se encuentra ubicado aproximadamente 150 pb río arriba del locus N-ras en múltiples especies (Jeffers, Paciucci, & Pellicer, 1990). La proteína del gen *unr* (proteína UNR) se expresa de forma ubicua y se encuentra principalmente en el citoplasma, ya sea en forma soluble o asociada a las membranas, en particular al retículo endoplásmico (Jacquemin-Sablon *et al.*, 1994). También se ha encontrado en el núcleo aunque en pequeñas proporciones, sin embargo el papel de la UNR en éste sigue siendo poco clara (Ferrer *et al.*, 1999; Patalano *et al.*, 2009).

La proteína UNR es única entre las proteínas eucariontes que contienen el dominio CSD ya que contiene cinco copias de este dominio (Doniger *et al.*, 1992) (Fig. 5). La proteína UNR tiene un alto grado de conservación entre los eucariontes, sólo está ausente en *C. elegans*. Lo anterior muestra que la tasa de conservación del dominio CSD, especialmente de CSD1, es más alta con respecto a la proteína CSP bacteriana (Mihailovich *et al.*, 2010).

Los CSDs de la proteína UNR presentan una función tanto en la proteína como en el ácido nucleico al que se une la proteína. Además UNR muestra preferencia por DNA de

una sola cadena (ssDNA) y RNA (Ferrer *et al.*, 1999; Jacquemin-Sablon *et al.*, 1994). Esta proteína ha sido encontrada en los complejos que intervienen en la regulación de la traducción y la estabilidad del mRNA. En la traducción, la UNR se comporta como un factor IRES trans-actin (ITAF), que regula la actividad de varios IRES virales y celulares (Brown & Jackson, 2004). También se ha propuesto que funciona como una chaperona de RNA, pero no está claro si la UNR permanece unida al RNA o puede disociarse de él como lo hacen normalmente las otras proteínas chaperonas (Evans *et al.*, 2003; Hunt *et al.*, 1999; Mitchell *et al.*, 2003).

Proteína glycine- rich

El dominio CSD se encuentra ampliamente extendido entre las plantas, incluyendo las monocotiledóneas y dicotiledóneas (Karlson & Imai, 2003). Este dominio se encuentra asociado a una gran región rica en glicina intercalada con dedos de zinc (CCHC) en el extremo c-terminal (Karlson *et al.*, 2002), esta estructura es muy parecida a la de la proteína Lin-28 (Sasaki & Imai, 2011) (Fig. 5). La proteína rica en glicina puede tener de 2 a 7 dedos de zinc (CCHC) dependiendo de la especie, las monocotiledóneas contienen por lo general 2 o 4, mientras que las dicotiledóneas contienen 1, 2, 4, 5 o 7 de estos dedos de zinc (CCHC). Las regiones ricas en glicina y los dedos de zinc (CCHC) probablemente están implicadas en la unión de ácidos nucleicos y a la unión de otras proteínas (Karlson *et al.*, 2002; Nakaminami, Karlson, & Imai, 2006; Sasaki, Kim, & Imai, 2007) ya que las diversas combinaciones son necesarias para dichos procesos (Sasaki & Imai, 2011).

Mientras que se han realizado considerables investigaciones de la función del dominio CSD en proteínas de bacterias y animales, se sabe poco de sus funciones en plantas. La primera proteína caracterizada en plantas que poseía el dominio fue WCSP1 del trigo (Karlson *et al.*, 2002). Esta proteína está compuesta por el dominio CSD y por una región rica en glicina intercalada con tres dedos de zinc (CCHC). Su mRNA se encuentra regulado por una respuesta al frío y la proteína se acumula considerablemente en el tejido

de la corona durante la aclimatación al frío prolongado, además de que sus niveles de transcripción no son modulados por otras tensiones ambientales como la concentración de sal, la sequía y el calor, o tratamiento con ácido abscísico (Karlson *et al.*, 2002; Nakaminami *et al.*, 2005, 2006). Se cree que la función de la proteína WCSP1 es específica a la adaptación al frío en plantas (Sasaki & Imai, 2011). También se ha demostrado que mutantes de CSP de *E. coli* con WCSP1 complementan el fenotipo sensible al frío de esta bacteria (Nakaminami *et al.*, 2006); esto último y el hecho de que la proteína WCSP1 se una al DNA y al RNA (Karlson *et al.*, 2002) sugiere que la WCSP1 comparte una función conservada con la CSP de *E. coli* y está implicada en la regulación de la aclimatación al frío (Sasaki & Imai, 2011).

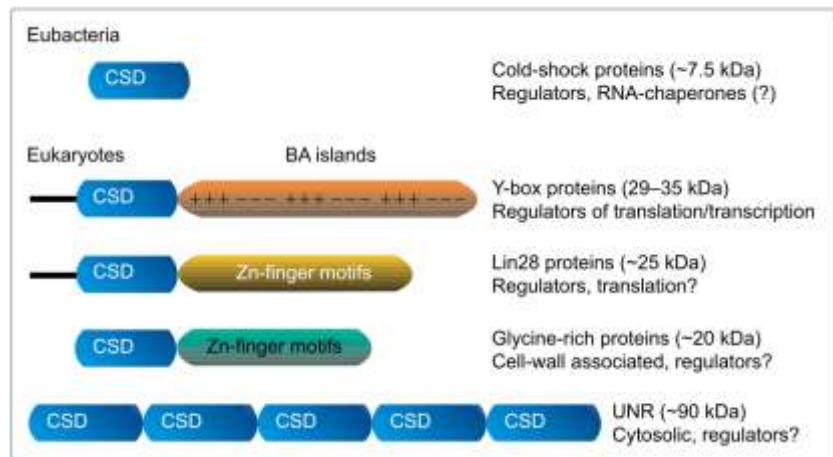


Fig. 5. Proteínas que contienen el dominio CSD. Las proteínas CSP consisten en un solo dominio CSD, el cual se une a RNA y ssDNA. La proteína Y-box está constituida por un dominio CSD unido a una isla BA (basic/aromatic), las proteínas Lin 28 y Glycine-rich por un dominio CSD y diversos números de dedos de zinc, mientras que la proteína UNR consta de 5 dominios CSD (tomado de P. L. Graumann & Marahiel, 1998)

Proteínas Cold Shock en Archaeas

En Archaeas se han identificado, hasta el año 2010, 16 especies que poseen proteínas con secuencias homólogas a las proteínas CSPs de las bacterias, y son principalmente del clado Halobacteriaceae (Mihailovich *et al.*, 2010). La similitud entre secuencias de CSP de Archeas y la CSPA de *E. coli* varía entre el 36 y el 59%, siendo la CSP de *M. frigidum* la que tiene mayor similitud (59%), sin embargo las CSPs de Archaeas son más similares entre sí (68 a 91%) (Giaquinto *et al.*, 2007).

Los estudios muestran que la CSP de Archaeas tiene una función en células que crecen a 4 °C (Goodchild *et al.*, n.d.); ésto, junto con el hecho de que las especies de Archaeas que tienen esta proteína son en su mayoría mesófilas y psicrófilas, refuerza la idea de que son importantes para el crecimiento a bajas temperaturas (Giaquinto *et al.*, 2007); ilustrando así, que en algunos casos, el dominio CSD ha conservado, en alto grado, una similitud funcional con sus homólogas bacterianas a lo largo de la evolución de las Archeas (Giaquinto *et al.*, 2007).

Evolución de la proteína Cold Shock y el dominio Cold Shock

La función de la proteína CSP en el metabolismo del RNA es similar al de la proteína ribosomal S1 en bacterias Gram-negativas. S1 consta de cinco dominios, se asocia con el ribosoma, se une a RNAs y es esencial para la iniciación de la traducción de varios mRNAs, principalmente de los que contienen 5'UTRs largas (Tedin, Resch, & Bläsi, 1997). Así mismo se ha demostrado que el dominio S1 tiene un pliegue similar a la proteína CSP (Bycroft *et al.*, 1997) y otras proteínas de unión a oligosacáridos (Murzin, 1993).

Aunque la homología de los dominios CSD y S1 no se puede demostrar mediante la secuencia, su alta similitud estructural sugiere que los dos dominios se originaron a partir de una proteína ancestral común (Bycroft *et al.*, 1997), a partir de la cual, los dominios

CSD y S1 actuales se separaron (Mihailovich *et al.*, 2010). Esto es apoyado por el hecho de que ambos dominios están altamente distribuidos a través de los 3 dominios de la vida, aunque en el caso del dominio CSD se piensa que las Archaeas lo adquirieron por transferencia horizontal ya que en solo muy pocas especies se ha detectado (Graumann & Marahiel, 1998; Mihailovich *et al.*, 2010).

De esta forma, la proteína CSP y el dominio CSD constituyen un ejemplo interesante de cómo durante la evolución, el dominio de una proteína antigua puede divergir a lo largo del árbol de la vida, formando arquitecturas funcionales diversas, que van desde que una proteína tenga un solo dominio multifuncional (como en el caso de las bacterias), hasta proteínas con múltiples copias de CSD (como UNR) y además que se pueden combinar con otros dominios reguladores. Esta capacidad de combinación junto con la flexibilidad del dominio CSD para el reconocimiento de ácidos nucleicos, da como resultado un repertorio creciente de proteínas implicadas en múltiples aspectos de la fisiología celular (Mihailovich *et al.*, 2010).

JUSTIFICACIÓN

Actualmente la información estructural sobre el dominio CSD se ha acumulado, sin embargo no existe un estudio que integre toda la información existente hasta el momento en cuanto a los organismos en los que se ha detectado la presencia de este dominio. Además la mayoría de los trabajos se han hecho con proteínas Heat Shock y son escasos los trabajos con las proteínas Cold Shock y el dominio CSD por lo que es importante realizar más estudios sobre éste con el fin de esclarecer aun más la historia evolutiva de este dominio y su importancia para los organismos que lo poseen.

OBJETIVOS

Objetivo general

Construir la filogenia molecular de las proteínas Cold Shock y el dominio Cold Shock utilizando secuencias proteicas de Bacterias, Archaeas y Eucariontes.

Objetivos particulares

- 1) Obtener la secuencia de la proteína CSPA del genoma completo de la cepa de *Escherichia coli* DH5 α mediante técnicas moleculares.
- 2) Llevar a cabo una búsqueda bibliográfica exhaustiva de estudios sobre la filogenia de la proteína Cold Shock y el dominio Cold Shock así como del papel que desempeña tanto en bacterias, archeas y eucariontes.
- 3) Construir una base de datos de las secuencias de los tres dominios utilizando la herramienta BLAST de las bases de datos KEGG, EMBL-EBI y HMMER y las herramientas BLAST y PSI-BLAST de la base de datos NCBI, usando como referencias la secuencia de la proteína CSPA obtenida en el laboratorio.
- 4) Realizar la filogenia individual de cada uno de los tres dominios y una que contenga todas las secuencias de los tres dominios. Utilizar métodos actualizados que tomen en cuenta la estructura terciaria de la proteína para realizar la filogenia molecular de los tres dominios.

METODOLOGÍA

Diseño, síntesis y elución de los oligonucleótidos

El diseño de los oligonucleótidos fue realizado en colaboración con el Dr. Alexander de Luna Fors en el laboratorio de Sistemas Biológicos Genéticos del Instituto LANGEBIO del CINVESTAV unidad Irapuato, México. Las secuencias fueron nombradas Cryo_pr1Fw (GTTTCTTCGAATTCGCGGCCGCTTCTAGAGCGGCATTAAGTAAGCAGTTGATG) que corresponde al oligo que va río arriba y Cryo2_pr3 Rv (TACTAGTAGCGGCCGCTGCAGGAAGAA-ACGTCAGGA-GTGATGAAGCCGAAGCC) para el oligo que va río abajo, las cuales amplifican la secuencia del *gen cspA*.

Los oligonucleótidos originales estaban liofilizados y se prepararon a una concentración de 200 mM. Posteriormente se hicieron alícuotas a una concentración de 10 µM que se utilizaron en el proceso de PCR.

Se extrajo DNA de *E.coli* cepa DH5α con el kit DNA easy Blood & Tissue de Qiagen® siguiendo las especificaciones del fabricante a partir de un cultivo overnight en LB.

Amplificación del gen *cspA* por Reacción en Cadena de la Polimerasa

La PCR se realizó usando el kit REDTaq ReadyMix PCR Reaction Mix, con MgCl₂, número de catalogo R2648, 20 mM Tris-HCl, pH 8.3, con 100 mM KCl, 3 mM MgCl₂, 0.002 % de gelatina, 0.4 mM dNTP mix (dATP, dCTP, dGTP, TTP), estabilizadores, y 0.06 unit/mL de Taq DNA Polimerasa. Proporcionado en 20 reacciones o 5X20 reacciones. Agua, reactivo PCR, número de catalogo W1754. Provisto en un tubo de 1.5 ml. Las condiciones fueron: 95 °C por 5 min, 35 ciclos de 95 °C por 1 min, 58 °C por 1 min, 72 °C por 1 min, y una extensión final a 72°C por 74 min y el producto se almacenó a

4 °C. La PCR se efectuó usando un termociclador Techne, Flexigene®, modelo FFG02HSD y una vez realizada se corrió en un gel de agarosa al 1% 1X TBE a 100 volts por 40 min para corroborar la eficacia del proceso.

Secuenciación

El producto de la amplificación de la PCR se mandó a secuenciar al Centro High Throughput Genomics en Seattle, Washington y al Instituto de Biología de la UNAM, D.F., México por duplicado.

Búsqueda bibliográfica

Se llevó a cabo una búsqueda bibliográfica en las siguientes bases de datos: Web of Science, Web of Knowledge, JSTOR, NCBI y ScienceDirect. Utilizando palabras claves como proteínas CSP, familia de proteínas CSP, proteínas CSP en Bacterias, filogenia molecular de CSP, evolución de CSP, dominio CSD en eucariontes, en animales, en plantas, proteínas CSP en Archeas, evolución del dominio CSD, entre otros. Con el fin de buscar los trabajos más recientes sobre la evolución y función de las proteínas CSPs y el dominio CSD

Elaboración de la Base de datos

Se realizó una búsqueda de secuencias homólogas mediante la herramienta BLAST (Altschul *et al.*, 1997) contra todas las secuencias existentes en cada base de datos: NCBI, KEGG (<http://www.genome.jp/kegg/>), EMBL-EBI y HMMER utilizando como referencia la secuencia, en aminoácidos, de la proteína CSPA obtenida en el laboratorio y las opciones

por default del BLAST de cada base de datos. Posteriormente se hizo la búsqueda con el BLAST contra las secuencias de Bacterias, Archaeas y Eucariontes por separado, en las bases de datos NCBI y KEGG, con las opciones por default de cada base, esto con el fin de esclarecer cuántas secuencias correspondían a cada dominio.

Una vez obtenidos los resultados de la búsqueda con BLAST de cada uno de los dominios, en las dos bases de datos, se realizaron búsquedas con la herramienta PSI-BLAST (Altschul *et al.*, 1997) de la base de datos NSBI contra las secuencias de Bacterias, Archaeas y Eucariontes por separado, utilizando las opciones estándar del mismo, con la finalidad de depurar las secuencias obtenidas.

La base de datos se construyó en Excel® tomando en cuenta: la secuencia, el número de acceso, los datos arrojados por el PSI-BLAST (la secuencia, el número de acceso, el porcentaje de identidad, el porcentaje de posiciones que son idénticos, alineamiento, longitud de alineación, número de desajustes, número de huecos, inicio de la alineación de la secuencia de consulta, fin de la alineación de la secuencia de consulta, inicio de la alineación de las secuencias de datos, fin de la alineación de la secuencia de las bases de datos, valor *e*, score (bit), el organismo al que pertenecía la secuencia, el rango de temperatura de crecimiento, el hábitat y la clasificación taxonómica.), el organismo al que pertenecía la secuencia, el rango de temperatura de crecimiento, el hábitat y la clasificación taxonómica. Para la inclusión de la secuencias en la base, se revisó que tuvieran un *e value* menor a 1×10^{-12} y un porcentaje de traslape mayor al 30%, a continuación se excluyeron las secuencias redundantes para cada organismo, tomando como criterio la cantidad y calidad de información existente, también se excluyeron las secuencias que eran hipotéticas, predictivas o putativas. En el caso particular de *E. coli* se quitaron todas las secuencias arrojadas por el PSI-BLAST y se trabajó con las secuencias de las nueve proteínas CSPs (CSPA-CSPI) utilizadas en trabajos anteriores (Etchegaray, Jones, & Inouye, 1996; Goldstein *et al.*, 1990; Lee *et al.*, 1994; Nakashima *et al.*, 1996; Wang *et al.*, 1999; Yamanaka & Inouye, 1997), con la finalidad de eliminar redundancias y quedarse con las secuencias que contaban con mayores estudios.

Construcción de la filogenia molecular

Una vez construida la base de datos se hicieron los alineamientos de las secuencias de cada dominio por separado y un alineamiento con las secuencias de los tres dominios, tomando como referencia la secuencia de la proteína CSPA para cada alineamiento, para esto se utilizó la herramienta bioinformática MAFFT (Multiple alignment program for amino acid or nucleotide sequences) versión 7 (Kato & Standley, 2013), ya que esta herramienta toma en cuenta las propiedades fisicoquímicas de los aminoácidos y supone que las secuencias de entrada descienden de un ancestro común por lo que el orden de las letras en cada secuencia de inicio siempre se conserva (Kato & Standley, 2013).

Una vez obtenidos los alineamientos se hicieron los árboles de máxima verosimilitud usando el programa PhyML 3.0 (Guindon *et al.*, 2010) utilizando las opciones por default, con un árbol inicial construido con BIONJ (Gascuel, 1997) y un modelo de sustitución LG (el cual incorpora la variabilidad de las tasas de evolución a través de los sitios para la estimación de matriz y usa una base de datos mucho más grande y diversa) (Le & Gascuel, 2008), posteriormente el programa construyó los árboles con máxima verosimilitud con un modelo de sustitución EX_EHO, este modelo se basa en la observación de que los aminoácidos externos de las proteínas presentan una tasa de sustitución mayor que los aminoácidos que se encuentran en las partes más internas de las proteínas (Le, Lartillot, & Gascuel, 2008) por lo que toma en cuenta la estructura secundaria de las proteínas, y un soporte de ramas aLRT SH-like. Los árboles fueron visualizados con el programa Archaeopteryx: Visualization, Analysis, and Editing of Phylogenetic Trees (Han & Zmasek, 2009) e iTol (interactive Tree Of Life) (Letunic & Bork, 2011). Por último el árbol con las secuencias de los tres dominios se colapsó con ayuda del programa iTOL.

Parálogos de CSPs en los genomas completos de Bacterias Psicrófilas y Psicrotolerantes.

Con el fin de comparar el número de parálogos según el hábitat de bacterias psicrófilas y psicrotolerantes, se extrajeron las secuencias de los genomas completos de *Aeromonas salmonicida*, *Aequorivita sublithicola*, *Bacillus weihenstephanensis*, *Colwellia psychrerythraea*, *Desulfotalea psychrophila*, *Exiguobacterium sibiricum*, *Flavobacterium psychrophilum*, *Glaciecola sp.*, *Glaciecola psychrophila*, *Idiomarina loihiensis*, *Leuconostoc citreum*, *Leuconostoc gasicomitatum*, *Leuconostoc kimchii*, *Listeria monocytogenes*, *Marinobacter sp*, *Methylocella silvestris*, *Octadecabacter arcticus*, *Octadecabacter antarcticus*, *Psychrobacter arcticus*, *Psychrobacter cryohalensis*, *Pseudoalteromonas haloplanktis*, *Psychromonas ingrahamii*, *Polomonas naphthalenivorans*, *Polaribacter irgensii*, *Photobacterium profundum*, *Psychroflexus torquis*, *Rhodiferax ferrireducens*, *Shewanella báltica*, *Saccharophagus degradans*, *Shewanella halifaxensis*, *Shewanella oneidensis*, *Shewanella violácea*, *Shewanella woodyi*, *Shewanella piezotolerans*, *Aliivibrio salmonicida*, de la base KEGG (<http://www.genome.jp/kegg/>) y se realizó un BLAST contra las secuencias de las 9 proteínas CSPs (CSPA-CSPI) de *E. coli*, con un script desarrollado en el Laboratorio de Origen de la Vida en la Facultad de Ciencias, UNAM, en lenguaje Perl.

RESULTADOS

Búsqueda bibliográfica

La búsqueda bibliográfica no arrojó muchos artículos recientes que estudiaran la evolución de la familia de las proteínas CSPs y del dominio CSD, siendo sólo los trabajos de Graumann & Marahiel (1998); Yamanaka *et. al.*, (1989); Yamanaka (1999); Giaquinto *et. al.*(2007), Mihailovich *et. al.* (2010) y Mojib *et. al* (2011), los que tratan de alguna forma este tema, sin embargo no existe hasta el momento, un estudio donde se construya una filogenia molecular que englobe secuencias de Bacterias, Archaeas y de Eucariontes.

Base de datos

Los datos arrojados por el BLAST de cada base de datos y con cada estrategia de búsqueda se muestran en la Tabla 1

	NCBI	KEGG	EMBL- EBI	JACKHMMER
BLAST general	13,130 hits	7,582hits	7,403 hits	35,343 hits
BLAST bacterias	11,366 hits	6,931 hits	-	-
BLAST archaeas	309 hits	125 hits	-	-
BLAST eucariontes	1651 hits	617 hits	-	-
PSI-BAST bacterias	1001 hits	-	-	-
PSI-BLAST archaeas	232 hits	-	-	-
PSI-BLAST eucariontes	500 hits	-	-	-

Tabla 1. Resultados del BLAST obtenidos con cada base de datos y con cada estrategia de búsqueda.

Del PSI-BLAST en NCBI se tomaron 153 secuencias de Bacterias, 57 secuencias de Eucariontes y las 30 secuencias de Archeas, que resultaron de la depuración con los criterios de exclusión descritos en la metodología, para la construcción de la base de datos y de la filogenia, esta base de datos se puede consultar en el servidor del Taller de Geo geografía y Sistemática en la Facultad de Ciencias, UNAM.

Alineamiento de secuencias de bacterias, archaeas y eucariontes

El alineamiento fue construido con el programa MAFFT (Katoh & Standley, 2013), este programa parte de la idea de que la frecuencia de sustitución entre 2 aminoácidos depende en gran medida de la diferencia de las propiedades fisicoquímicas, particularmente del volumen y polaridad de dichos aminoácidos (Miyata, Miyazawa, & Yasunaga, 1979), por lo que la sustitución entre aminoácidos con propiedades físico-químicas similares tienden a preservar la estructura de la proteína, asumiéndolas así como sustituciones neutrales que se han ido acumulando en las moléculas a lo largo de la evolución (Kimura, 1983). Este programa fue de gran ayuda en este trabajo debido a la diversidad de organismos con los que se trabajó ya que se obtuvo un alineamiento sin muchos *gaps*.

La Fig. 6 muestra el alineamiento de las secuencias de los 3 dominios en donde se puede observar que las secuencias de los “motivos” RNP1 y RNP2 y la de las cadenas $\beta 2$ y $\beta 3$ se encuentra bien conservada dentro de las secuencias de Bacterias, Archaeas y Eucariontes, y en menor grado la de las cadenas $\beta 1$ y $\beta 4$. Esto también se puede observar en la gráfica de barras en la parte de abajo del alineamiento, donde las barras de amarillo más claro indican un mayor grado de conservación.

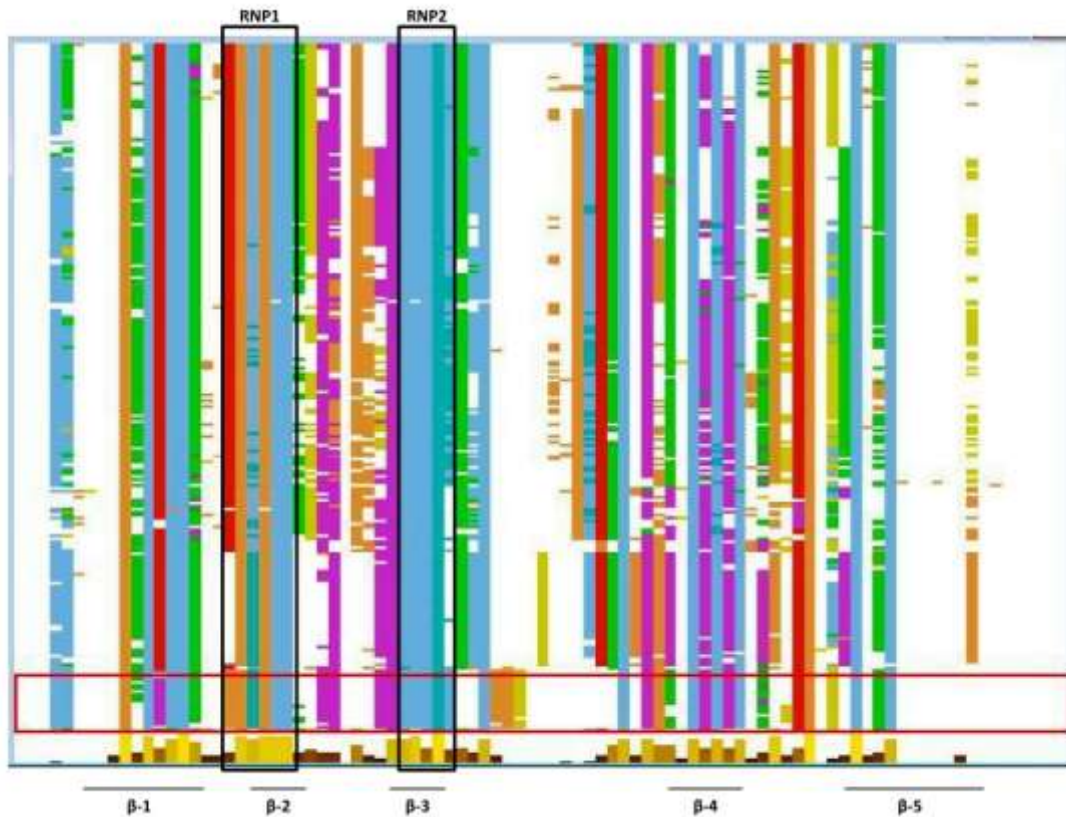


Fig. 6. Alineamiento de secuencias de Bacterias, Archeas y Eucariontes. Las cadenas β_1 , β_2 , β_3 , β_4 y β_5 se encuentran indicadas con una línea en la parte inferior de las secuencias y los motivos RNP1 Y RNP2 están dentro de las cajas.

Reconstrucción filogenética

Utilizando el programa PhyML 3.0 (Guindon *et al.*, 2010) se obtuvieron las reconstrucciones filogenéticas con un árbol inicial construido con BIONJ (Gascuel, 1997) y un modelo de sustitución de aminoácidos LG y posteriormente el programa reconstruyó el árbol con máxima verosimilitud y un modelo de sustitución EX_EHO (Fig. 7 y 8).

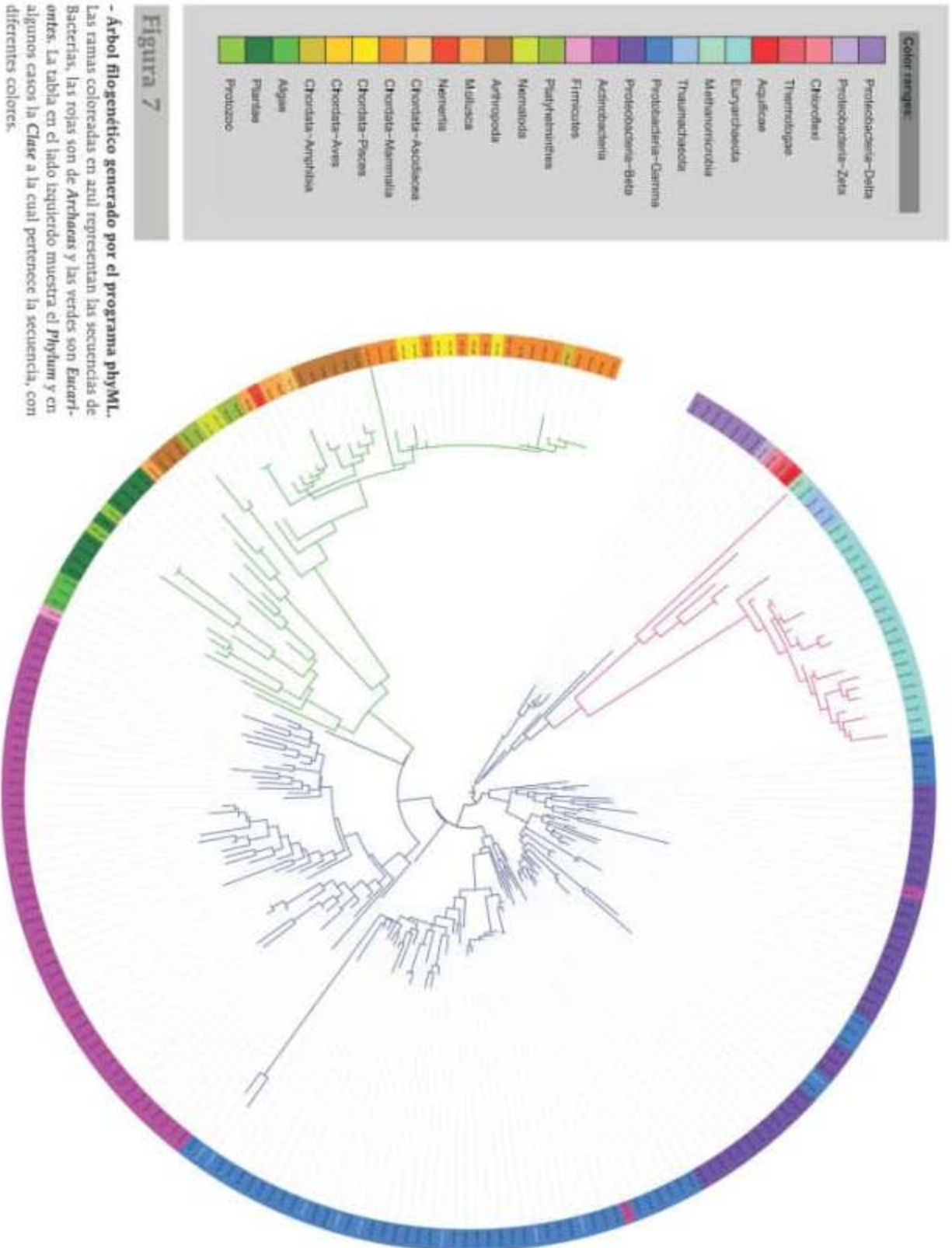
Lo primero que se puede observar en el árbol (Fig. 7 y 8) es que los 3 dominios se encuentran claramente separados (las ramas en color azul pertenecen a secuencias de bacterias, las de color rojo son de archaeas y las de color verde son de eucariontes). Más a

detalle se pudo observar que las secuencias de bacterias se separaron y agruparon principalmente según el *PHYLUM*, la *CLASE* y el *ORDEN*, y se puede apreciar mediante los diferentes colores asignados a cada *PHYLUM* en la figura 7. También se vio que algunos clados estaban conformados por el mismo tipo de proteína (CSPA, CSPB, CSPC, CSPD, CSPE, CSPF, CSPG, CSPH y CSPI, respectivamente) y que las que pertenecen a CSPA, CSPB, CSPG y CSPI se encuentran en ramas más cercanas en el árbol (Fig. 7 y 8), aunque no es posible apreciarlo del todo ya que la mayoría de las secuencias no están completamente tipificadas, en la figura 8 es donde se encuentran etiquetadas, mediante diferentes colores, dependiendo del tipo de proteína al que pertenecen. Además, las 9 proteínas de *E. coli* no se encuentran agrupadas en un mismo clado con la excepción de CSPF y CSPH. Es importante mencionar que ningún otro organismo de la base de datos creada para este trabajo presenta una secuencia identificada como homóloga CSPF o CSPH.

Para el caso de las secuencias de Eucariontes se puede observar que los principales clados están formados en principio por el tipo de proteína (Y-box, CSDA, Lin28 y Glycin-rich) (Fig. 8), las secuencias de Y-box, se separan en Y-B1 y Y-B2, y en segundo término por el organismo al que pertenecen (Fig. 7).

Las secuencias de las Archaeas se separaron principalmente en 2 clados, el que contiene más secuencias está formado por Archaeas del *Phylum* Euryarchaeota, *Clase* Halobacteria y el clado con menos secuencias está formado por 3 secuencias del *Phylum* Thaumarchaeota, 1 secuencia del *Phylum* Methanomicrobia y 1 secuencia sin clasificar (Fig. 7).

También se pudieron observar 46 nodos con un valor de aLRT SH-like menor a 0.5, los cuales se colapsaron (Fig. 8), a pesar de eso quedaron ramas no resueltas.



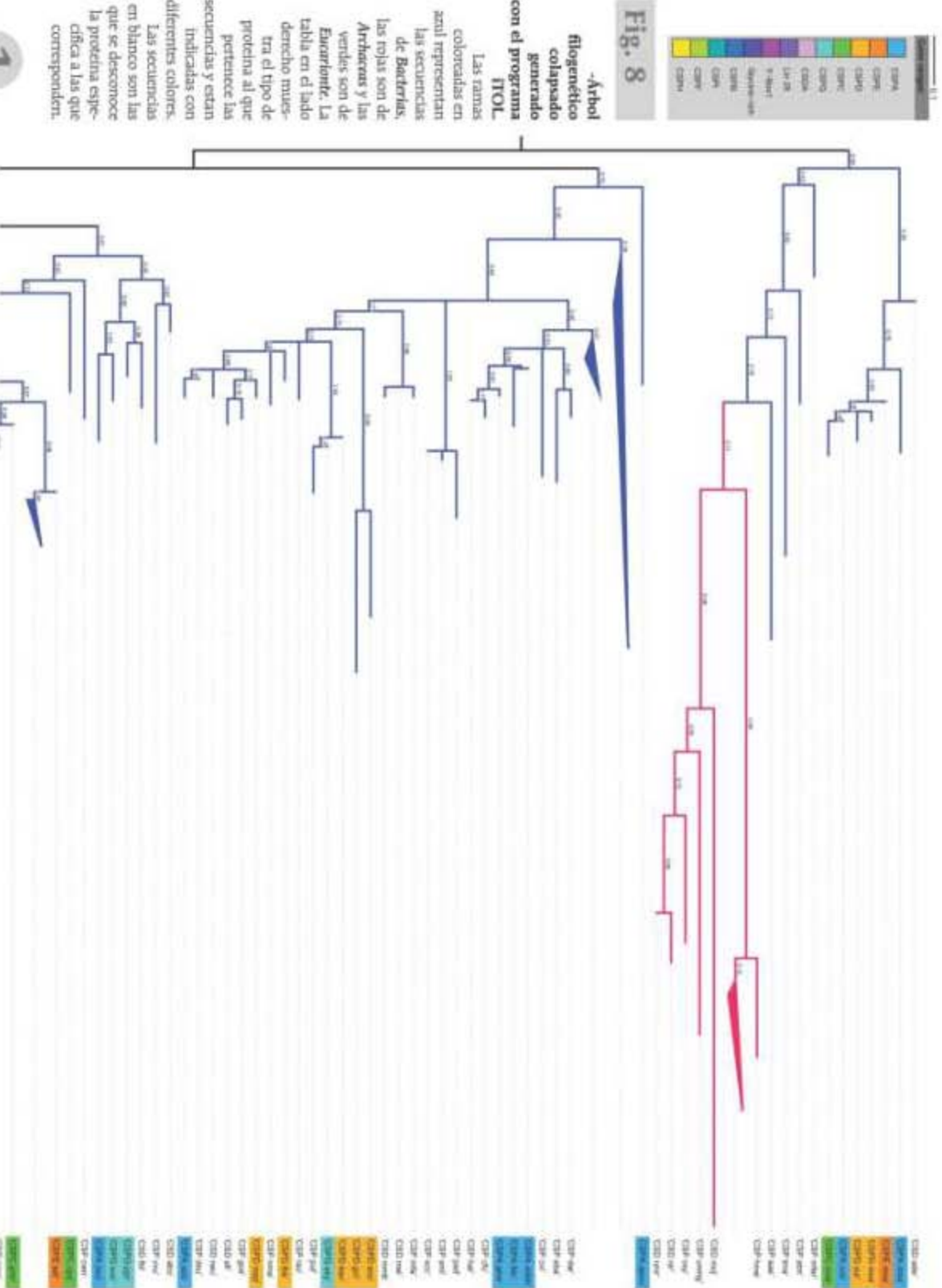


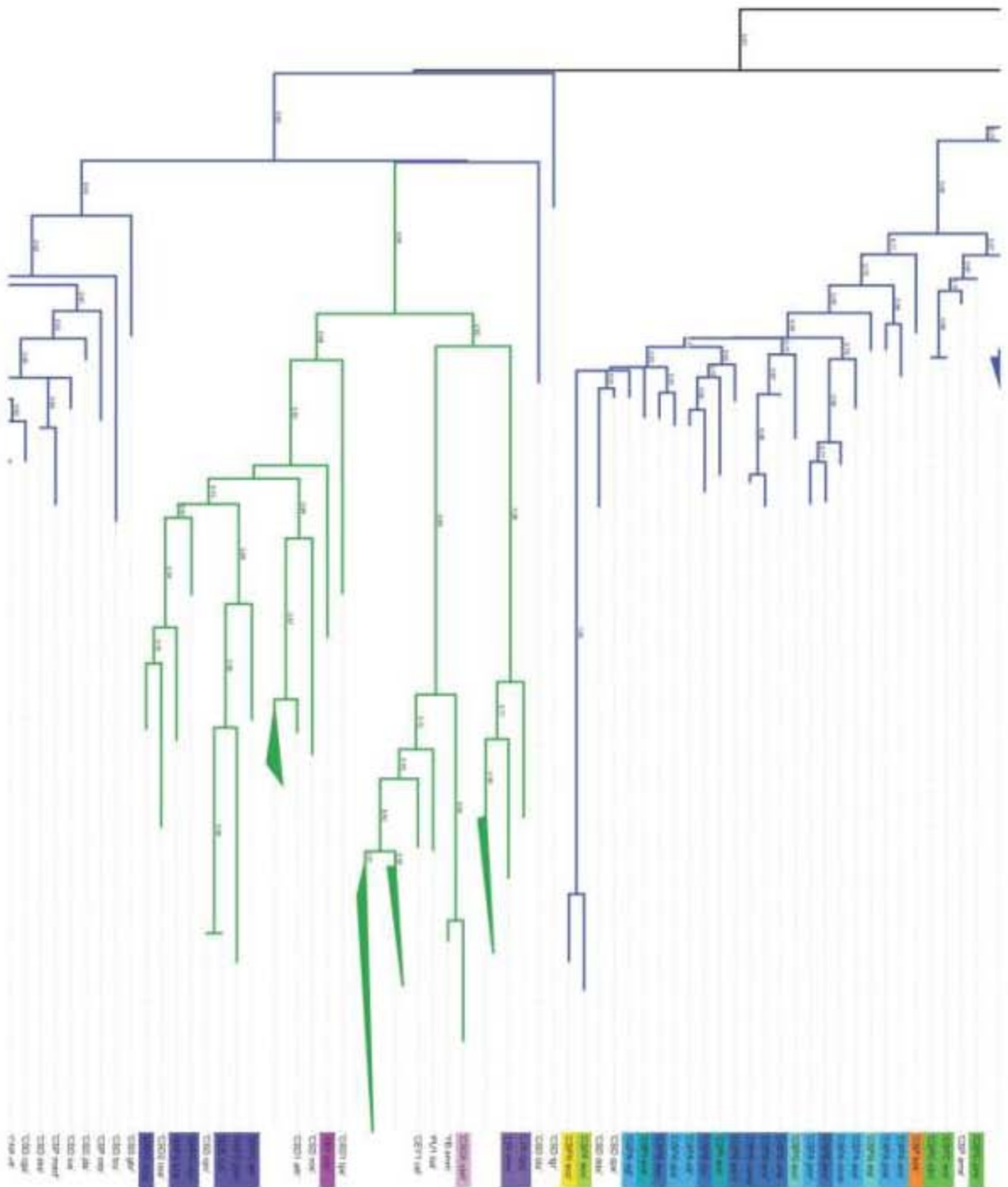
Fig. 8

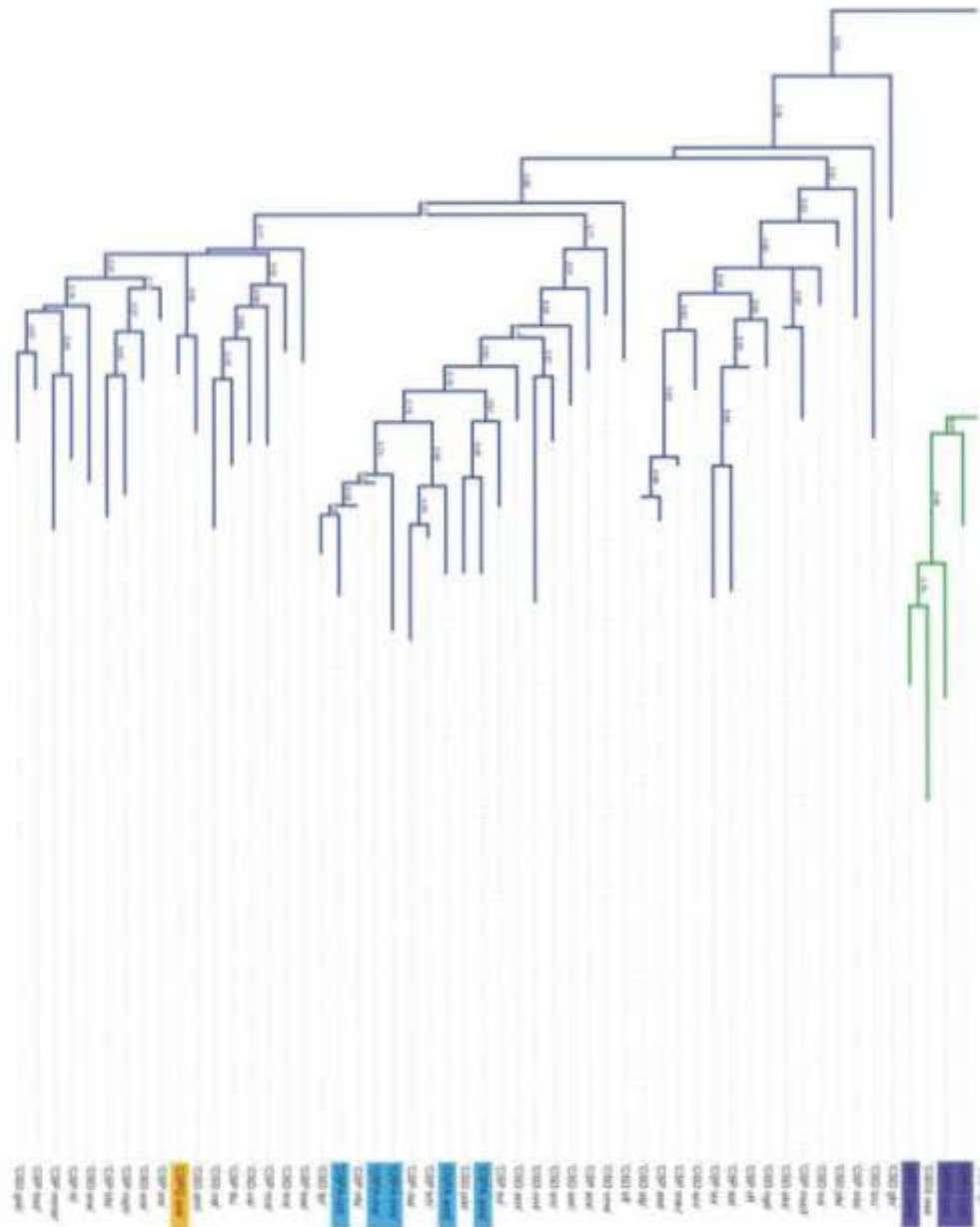
-Árbol filogenético colapsado generado con el programa IQ-TREE.

Las ramas coloreadas en azul representan las secuencias de *Bacterias*, las rojas son de *Archaeos* y las verdes son de *Eucariotes*. La tabla en el lado derecho muestra el tipo de proteína al que pertenece las secuencias y están indicadas con diferentes colores. Las secuencias en blanco son las que se desconoce la proteína específica a las que corresponden.

1

2





Búsqueda de parálogos de CSPs en los genomas completos de Bacterias Psicrófilas y Psicrotolerantes.

En la mayoría de los genomas que se utilizaron no se encontró que hubiera un gran número de copias homólogas a la secuencia de la proteína CSPA, CSPB, CSPC, CSPD, CSPE, CSPF, CSPG, CSPH y CSPI de *E. coli*, aunque se puede observar que la mayoría (94.3 %) presenta mínimamente una copia y la moda y mediana son de 3 copias.

Los organismos que presentaron mayor número de copias son en primer lugar *Psychromonas ingrahamii* con 11 homólogas (Tabla 2) y en segundo lugar *Pseudoalteromonas haloplanktis* con 9 homólogas (Tabla 2) estas bacterias son psicrófilas y habitan en aguas marinas de alguno de los 2 polos, también *Saccharophagus degradans*, habitante de aguas marina, presenta un número significativo de copias (6 copias) aunque esta es una bacteria psicrotolerante (Tabla 2). Las bacterias que habitan en ambientes marinos acuáticos presentan un promedio de 3.5 copias y una moda y mediana de 3 copias.

Nombre	Aequorivita sublithicola	Glaciecola psychrophila	Idiomarina loihiensis	Pseudoalteromonas haloplanktis	Psychromonas ingrahamii	Polaribacter irgensii	Psychroflexus torquis	Saccharophagus degradans	Shewanella baltica
Rango de temperatura	Psicrotolerante	Psicrofila	Psicrotolerante	Psicrofila	Psicrofila	Psicrofila	Psicrofila	Psicrotolerante	Psicrotolerante
Habitat	Marina, acuatica, Antartica	Marina, acuatica	Fuente hidrotermal, marina	Acuatica marina, costa antartica	Marina, acuatica, polar Artico	Aguas superficiales marinas Antarticas	Marina, acuatica, Antartico	Marina, Acuatica, marisma	Marina, acuatica, mar baltico
Name	asl.faa	gps.faa	ilo.faa	pha.faa	pin.faa	pom.faa	ptq.faa	sde.faa	sbp.faa
CspA	2	3	3	9	11	1	2	6	3
CspB	1	3	3	9	11	0	2	6	3
CspC	3	3	3	9	11	3	2	6	3
CspD	2	3	3	9	11	2	2	6	3
CspE	2	3	3	9	11	3	2	6	3
CspF	0	0	1	0	0	0	0	0	0
CspG	2	4	3	9	11	1	2	6	3
CspH	0	0	1	0	0	0	0	0	1
CspI	2	3	3	9	11	1	2	6	3

Tabla 2. Homólogos de CSPs en bacterias psicrófilas y psicrotolerantes acuáticas marinas. La tabla muestra el número de secuencias homólogas a cada una de las proteínas CSPs dentro del genoma completo del organismo.

Las bacterias que habitan en sedimentos marinos (Tabla3) presentan un promedio de copias de 3.9 con una moda de 3 y una mediana de 4, también es el ambiente más homogéneo en cuanto al número de copias y con el promedio mayor, en comparación a los demás hábitats estudiados en este trabajo. Además *Photobacterium profundum* (habitante de sedimentos en mar profundo), *Desulfotalea psychrophila* (habitante en sedimentos del mar ártico) y *Shewanella violacea* (habitante de sedimentos de mar profundo) tienen un número significativo de copias, de 8,7 y 6 respectivamente (Tabla3).

Nombre	Colwellia psychrerythraea	Desulfotalea psychrophila	Marinobacter sp	Photobacterium profundum	Shewanella halifaxensis	Shewanella violacea	Shewanella woodyi	Shewanella piezotolerans
Rango de temperatura	Psicrofila	Psicrofila	Psicrofila	Psicrotoletante	Psicrofila	Psicrofila	Psicrofila	Psicrotolerante
Habitat	Sedimento marino Artico	Sedimento marino Artico de Svalbard	Sedimentos de mar profundo del Artico	Sedimentos de mar profundo	Sedimentos, Oceano Atlantico	Sedimento del mar profundo	Sedimento de mar profundo	Sedimento de mar profundo
Name	cps.faa	dps.faa	mbs.faa	ppr.faa	shl.faa	svo.faa	swd.faa	swp.faa
CspA	4	7	5	8	3	6	4	3
CspB	4	6	5	8	3	6	4	3
CspC	4	6	5	8	3	6	4	3
CspD	4	6	5	8	3	6	4	3
CspE	4	6	5	8	3	6	4	3
CspF	0	0	0	1	0	0	0	1
CspG	5	6	5	8	3	6	4	3
CspH	0	0	0	3	0	0	0	1
CspI	4	6	5	8	3	6	4	3

Tabla 3. Homólogos de CSPA en bacterias psicrofilas y psicrotolerantes de sedimentos marinos. La tabla muestra el número de secuencias homólogas a cada una de las proteínas CSPs dentro del genoma completo del organismo.

Las bacterias habitantes de ambientes dulceacuícolas presentan un promedio de 1.4 copias, una moda de 0 y una mediana de 2, teniendo así de los valores más bajos de número de copias de los hábitats en este trabajo y una de las 2 bacterias que en este estudio no presentaron ningún homólogo a alguna proteína CSPs de *E. coli* (*Polomonas naphthalenivorans*) (Tabla 4), aunque de esta bacteria se reporta al menos una copia (Siddiqui *et al.*, 2013), además de que en el NCBI (<http://www.ncbi.nlm.nih.gov/taxonomy>) está clasificada como bacteria mesófila y en esa revisión como psicrofílica.

Nombre	Glaciecola sp.	Octadecabacter antarcticus	Octadecabacter arcticus	Polomonas naphthalenivorans
Rango de temperatura	Psicrofila	Psicrofila	Psicrofila	Mesofila 20 °C
Habitat	Agua fresca	Acuatica,hielo, Antartica	Agua fresca, Marina, Artico	Acuatica, sedimentos
Name	gag.faa	oat.faa	oar.faa	pna.faa
CspA	3	3	2	0
CspB	3	2	1	0
CspC	3	3	2	0
CspD	3	3	2	0
CspE	3	3	2	0
CspF	0	0	0	0
CspG	3	3	2	0
CspH	0	0	0	0
CspI	3	2	1	0

Tabla 4. Homólogos de CSPs en bacterias psicrofilas y psicrotolerantes de agua dulce. La tabla muestra el número de secuencias homólogas a cada una de las proteínas CSPs dentro del genoma completo del organismo.

Las bacterias que tienen como hábitat sedimentos dulceacuícolas presentan un promedio de 1.2, una moda y mediana de 0, siendo así, junto con las bacterias que viven en agua dulce, las bacterias que presentan el menor número de copias de los habitas estudiados en este trabajo. (Tabla 4 y 5). *Rhodoferrax ferrireducens* habitante de sedimentos dulceacuícolas tampoco presento algún homólogo a CSPs de *E. coli* (Tabla5).

Nombre	Rhodoferax ferrireducens	Shewanella oneidensis
Rango de temperatura	Psicrotolerante	Mesofila
Habitat	Sedimento, aguas residuales y estancadas, sedimento marino, aguas saladas	Sedimento, agua fresca, acuatico
Name	rfr.faa	son.faa
CspA	0	3
CspB	0	3
CspC	0	3
CspD	0	3
CspE	0	3
CspF	0	0
CspG	0	3
CspH	0	1
CspI	0	3

Tabla 5. Homólogos de CSPs en bacterias psicrófilas y psicrotolerantes de sedimentos de agua dulce. La tabla muestra el número de secuencias homólogas a cada una de las proteínas CSPs dentro del genoma completo del organismo.

Las bacterias habitantes de suelos tienen un promedio de copias de 3.1, una moda y mediana de 3, en donde resalta *Exiguobacterium sibiricum* con 6 homólogos de CSPs de *E. coli* aunque esta bacteria también se ha encontrado en mar abierto (Tabla 6).

Nombre	Exiguobacterium sibiricum	Methylocella silvestris	Psychrobacter arcticus	Psychrobacter cryohalensis
Rango de temperatura	Psicrofila	Psicrotolerante	Psicrofila	Psicrofila
Habitat	Suelo congelado de Siberia, mar abierto	Suelo acido, cambisol	Suelos congelados de Siberia	Suelos congelados de Siberia
Name	esi.faa	rsl.faa	par.faa	pcr.faa
CspA	6	4	4	3
CspB	6	4	3	3
CspC	6	4	3	3
CspD	6	4	3	3
CspE	6	4	3	3
CspF	0	0	0	0
CspG	6	4	3	3
CspH	0	0	0	0
CspI	6	4	3	3

Tabla 6. Homólogos de CSPs en bacterias psicrófilas y psicrotolerantes de suelo. La tabla muestra el número de secuencias homólogas a cada una de las proteínas CSPs dentro del genoma completo del organismo.

Las bacterias patógenas de peces presentan un promedio de copias de 2.5, una moda de 4 y una mediana de 3. *Aeromonas salmonicida* tiene 4 copias homólogas a CSPs de *E. coli* y es patógena de un pez de agua dulce mientras que *Aliivibrio salmonicida* también tiene 4 copias y es patógena de un pez marino (Tabla 7).

Nombre	Aeromonas salmonicida	Aliivibrio salmonicida	Flavobacterium psychrophilum	Listeria monocytogenes
Rango de temperatura	Psicrofila	Psicrofila	Psicrofila	Psicrotolerante
Habitat	Patogena de la Trucha, agua fresca	Marina, Patogena de Bacalao Atlantico	Patogeno de peces, agua fresca	Patogeno transmitido por alimentos, agua fresca, suelo
Name	asa.faa	vsa.faa	fps.faa	lml.faa
CspA	4	4	1	3
CspB	4	4	1	3
CspC	4	4	1	3
CspD	4	4	1	3
CspE	4	4	1	3
CspF	1	1	0	0
CspG	4	4	1	3
CspH	2	2	0	0
CspI	4	4	1	3

Tabla 7. Homólogos de CSPs en bacterias psicrofilas y psicrotolerantes patógenas de peces. La tabla muestra el número de secuencias homólogas a cada una de las proteínas CSPs dentro del genoma completo del organismo.

Las bacterias que tienen como hábitat alimentos presentaron un promedio de copias homólogas de CSPs de *E. coli* de 2.52, una moda y mediana de 2 copias. *Bacillus weihenstephanensis* presente en leche pasteurizada resalta en este grupo ya que tiene 7 copias aunque los demás integrantes de este grupo son todos del género *Leuconostoc* y presentan valores iguales de copias (Tabla 8).

Nombre	Bacillus weihenstephanensis	Leuconostoc citreum	Leuconostoc gasicomitatum	Leuconostoc kimchi
Rango de temperatura	Psicrotolerante	Psicrofila	Psicrofila	Psicrofila
Habitat	Comida: leche pasteurizada, Sue- lo	Comida, Producto fermentado	Comida	Comida, producto fermentado
Name	bwe.faa	kci.faa	lgs.faa	lki.faa
CspA	7	2	2	2
CspB	7	2	2	2
CspC	7	2	2	2
CspD	7	2	2	2
CspE	7	2	2	2
CspF	0	0	0	0
CspG	7	2	2	2
CspH	0	0	0	0
Cspi	7	2	2	2

Tabla 8. Homólogos de CSPs en bacterias psicrófilas y psicrotolerantes presentes en alimentos. La tabla muestra el número de secuencias homólogas a cada una de las proteínas CSPs dentro del genoma completo del organismo.

DISCUSIÓN

En las bases de datos se encuentran una gran cantidad de secuencias homólogas a la proteína CSPA de *Escherichia coli*, sin embargo la mayoría de ellas no cuentan con información que permita verificar su autenticidad y muchas de estas no han sido utilizadas en algún trabajo donde se analice más a fondo su función. También en muchos casos son secuencias putativas o hipotéticas, y para el caso específico de la base de datos NCBI muchas de las secuencias que poseen no han sido completamente revisadas.

Todo esto provocó que no se pudiera incluir una mayor diversidad de secuencias de organismos en la construcción de la filogenia, las secuencias de Bacterias son principalmente de proteobacterias (63.4 %) de las cuales son mayormente gamma-proteobacterias, 34 % son de actinobacterias y el 2.6% pertenecen a otros *phylla*. De las secuencias de las Archaeas sólo se utilizaron 30 aún cuando al quitar las redundancias se obtuvieron 81 secuencias, ya que no presentaban suficiente información de los estudios realizados con ellas además de que no estaban completamente avaladas por el NCBI.

Las CSPs de las Bacterias no están bien clasificadas según el tipo de proteína CSP al que pertenecen, por lo que no se puede observar con mucha claridad si se agrupan por tipo de CSP (CSPA-CSPI) debido a que muchas de estas secuencias provienen de secuenciaciones completas de los genomas de los organismos y las bases de datos las incluyen sin que tengan algún otro tipo de estudio. De las que sí se sabe qué tipo de CSP son, hay clados donde se agrupan y se puede pensar que probablemente las demás secuencias que lo conforman pertenecen al mismo tipo de CSP y que presenten una función similar (Sleator, 2012).

A pesar de que el alineamiento de las secuencias de los 3 dominios no presenta muchos *gaps* y de que los motivos RNP1 y RNP2 están altamente conservados, en el árbol filogenético algunas ramas quedaron como no resueltas. La capacidad de detectar las relaciones evolutivas de los dominios por similitud de secuencia es limitada porque con frecuencia divergen más allá del punto en el que las relaciones de confianza pueden ser reconocidas por las herramientas de comparación de secuencias (Chothia & Gough, 2009). Si se conoce la estructura tridimensional, la función y la secuencia de una proteína, la estructura del dominio y las relaciones evolutivas por lo general pueden ser reconocidas (Chothia & Gough, 2009; Murzin, 1998). Por lo que fue de suma importancia que para este trabajo se utilizaran programas como el MAFFT que toma en cuenta las propiedades fisicoquímicas de los aminoácidos y con esto considera las sustituciones de aminoácidos con propiedades fisicoquímicas similares como sustituciones neutrales (Katoh, 2002; Katoh & Standley, 2013) y el PhyML 3.0 que considera la estructura secundaria de las proteínas (Le & Gascuel, 2010) para la reconstrucción de la filogenia del dominio cold shock (CSD) y las proteínas CSPs en los 3 dominios del árbol de la vida. Además, en este trabajo se presenta la primera construcción filogenética del dominio cold shock y las proteínas CSPs con secuencias de bacterias, archaeas y eucariontes basada en la construcción de una base de datos exhaustiva y modelos de sustitución de aminoácidos novedosos.

En la búsqueda de parálogos dentro del genoma completo de bacterias psicrotolerantes y psicrófilas no se observa un gran número de homólogos con respecto a las 9 proteínas homólogas de *E. coli*, sin embargo el hecho de que la mayoría posea al

menos un homólogo a estas proteínas a pesar de tener adaptaciones para vivir a bajas temperaturas, como tener membranas con mayor contenido de ácidos grasos metilramificados insaturados y polisaturados y /o una longitud más corta en las cadenas *acyl* (Chintalapati, Kiran, & Shivaji, 2004; Russell, 1997) lo que se cree que tiene un papel clave en el aumento de la fluidez de membrana mediante la introducción de restricciones estéricas que cambian el orden de empaquetamiento o reducen el número de interacciones en la membrana; la trehalosa y los exopolisacáridos (EPSs) también pueden tener un papel importante de crioprotección en psicrófilos, la trehalosa se cree que tiene un efecto coligativo, pero probablemente también ayuda en la desnaturalización y la agregación de proteínas (Phadtare, 2004). Los EPSs modifican el entorno fisicoquímico de las bacterias, participan en la adhesión celular y retención de agua, favorecen la captación y la concentración de nutrientes, conservan y protegen a las enzimas extracelulares contra la desnaturalización por frío (Nichols *et al.*, 2005); y también se ha visto que las CSPs de las bacteria psicrófilas actúan de forma similar que las CSPs de bacterias mesófilas, implicadas en diversos procesos celulares, como la transcripción, la traducción, plegamiento de proteínas y la regulación de la fluidez de la membrana (Phadtare, 2004). Éstas se encuentran de forma constitutiva, siendo más bien proteínas de aclimatación al frío (CAPs) (Berger *et al.*, 1996; Ray, Kumar, & Shivaji, 1994), lo que sugiere que las proteínas CSPs juegan un papel importante en la adaptación al frío, así como en los procesos básicos de funcionamiento celular.

Ya que se ha visto que las proteínas cold shock (CSPs) son de suma importancia para muchos organismos a lo largo del árbol de la vida y que no existen muchos estudios evolutivos como es el caso de las proteínas heat shock (HSP); es imprescindible que se realicen más trabajos evolutivos a mayor profundidad y no sólo de relevancia biotecnológica. En el futuro se espera poder mejorar este trabajo depurando más las secuencias, construyendo un árbol filogenético con raíz y generando un reloj molecular para poder compararlo con eventos climatológicos presentes a lo largo de la evolución de la vida.

CONCLUSIONES

1. Se realizó una búsqueda bibliográfica y bioinformática exhaustiva empleando diversas bases de datos de las proteínas cold shock y del dominio cold shock así como de la evolución de estas proteínas y de este dominio.
2. De la búsqueda bibliográfica se pudo observar que no hay trabajos muy recientes que estudien la evolución de las proteínas CSPs y el dominio CSD, siendo solo los trabajos de Yamanaka (1999); Giaquinto *et. al.*(2007), Mihailovich *et. al.* (2010) y Mojib *et. al* (2011) los únicos que tratan algún punto de éste tema
3. Se obtuvo satisfactoriamente la secuencia de la proteína CSPA de *E.coli* cepa DH5 α mediante técnicas moleculares.
4. Mediante la búsqueda bioinformática se observó que a pesar de la gran cantidad de secuencias obtenidas, muchas de éstas carecían de la información que acredite su veracidad por lo que no se pudieron utilizar para la construcción de la base de datos.
5. Se construyó una base de datos con 240 secuencias de las cuales 153 secuencias son de bacterias, 30 de archaeas y 57 de eucariontes, la base de datos contiene: la secuencia, el número de acceso, el porcentaje de identidad, el porcentaje de posiciones que son idénticos, alineamiento, longitud de alineación, número de desajustes, número de huecos, inicio de la alineación de la secuencia de consulta, fin de la alineación de la secuencia de consulta, inicio de la alineación de las secuencias de datos, fin de la alineación de la secuencia de las bases de datos, valor *e*, score (bit), el organismo al que pertenecía la secuencia, el rango de temperatura de crecimiento, el hábitat y la clasificación taxonómica.
6. Se construyó una filogenia molecular utilizando secuencias de los dominios Bacteria, Archaea y Eucaria con modelos novedosos de alineamiento y

reconstrucción de árboles filogenéticos, en la cual se pudo observar una clara separación por *phyla* entre las secuencias de estos tres dominios y en menor grado un agrupamiento por tipo de proteína sobre todo en las secuencias de eucariontes. Proponemos que las CSPs y dominio CSD se encuentran altamente conservadas entre las bacterias, eucariontes y archaeas debido a que juegan un papel muy importante en el mecanismo de unión al RNA.

7. Se obtuvo el número de parálogos de las 9 proteínas CSPs de *E. coli* en los genomas completos de bacterias psicrófilas y psicrotolerantes, donde la mayoría presenta al menos un homólogo a estas proteínas a pesar de que cuentan con otras estrategias de adaptación al frío, por lo que se puede decir que estas proteínas son muy importantes para la adaptación al frío y en condiciones normales de crecimiento. También se puede decir que, en este tipo de bacterias, las que presentan más copias de genes homólogos de proteínas CSPs de *E. coli* son las que viven en ambientes marinos.

REFERENCIAS

- Altschul, S. F., Madden, T. L., Schäffer, A. A., Zhang, J., Zhang, Z., Miller, W., & Lipman, D. J. (1997). Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Research*, 25(17), 3389–402.
- Bae, W., Jones, P. G., & Inouye, M. (1997). CspA, the major cold shock protein of *Escherichia coli*, negatively regulates its own gene expression. *Journal of Bacteriology*, 179(22), 7081–8.
- Berger, F., Morellet, N., Menu, F., & Potier, P. (1996). Cold shock and cold acclimation proteins in the psychrotrophic bacterium *Arthrobacter globiformis* SI55. *Journal of Bacteriology*, 178(11), 2999–3007.
- Bermúdez-Cruz, R. M., Fernández-Ramírez, F., Ramírez, F., Kameyama-Kawabe, L., & Montañez, C. (2005). Conserved domains in polynucleotide phosphorylase among eubacteria. *Biochimie*, 87(8), 737–45. doi:10.1016/j.biochi.2005.03.005
- Bouvet, P., Matsumoto, K., & Wolffe, A. P. (1995). Sequence-specific RNA recognition by the *Xenopus* Y-box proteins. An essential role for the cold shock domain. *The Journal of Biological Chemistry*, 270(47), 28297–303.
- Bouvet, P., & Wolffe, A. P. (1994). A role for transcription and FRGY2 in masking maternal mRNA within *Xenopus* oocytes. *Cell*, 77(6), 931–41.
- Brandi, A., Pietroni, P., Gualerzi, C. O., & Pon, C. L. (1996). Post-transcriptional regulation of CspA expression in *Escherichia coli*. *Molecular Microbiology*, 19(2), 231–40.
- Brandi, A., Spurio, R., Gualerzi, C. O., & Pon, C. L. (1999). Massive presence of the *Escherichia coli* “major cold-shock protein” CspA under non-stress conditions. *The EMBO Journal*, 18(6), 1653–9. doi:10.1093/emboj/18.6.1653
- Brown, E. C., & Jackson, R. J. (2004). All five cold-shock domains of unr (upstream of N-ras) are required for stimulation of human rhinovirus RNA translation. *The Journal of General Virology*, 85(Pt 8), 2279–87. doi:10.1099/vir.0.80045-0
- Bycroft, M., Hubbard, T. J., Proctor, M., Freund, S. M., & Murzin, A. G. (1997). The solution structure of the S1 RNA binding domain: a member of an ancient nucleic acid-binding fold. *Cell*, 88(2), 235–42.
- Cairrão, F., Cruz, A., Mori, H., & Arraiano, C. M. (2003). Cold shock induction of RNase R and its role in the maturation of the quality control mediator SsrA/tmRNA. *Molecular Microbiology*, 50(4), 1349–60.

- Chen, C., & Deutscher, M. P. (2005). Elevation of RNase R in response to multiple stress conditions. *The Journal of Biological Chemistry*, 280(41), 34393–6. doi:10.1074/jbc.C500333200
- Chintalapati, S., Kiran, M. D., & Shivaji, S. (2004). Role of membrane lipid fatty acids in cold adaptation. *Cellular and Molecular Biology (Noisy-Le-Grand, France)*, 50(5), 631–42.
- Chothia, C., & Gough, J. (2009). Genomic and structural aspects of protein evolution. *The Biochemical Journal*, 419(1), 15–28. doi:10.1042/BJ20090122
- Coles, L. S., Diamond, P., Occhiodoro, F., Vadas, M. A., & Shannon, M. F. (1996). Cold shock domain proteins repress transcription from the GM-CSF promoter. *Nucleic Acids Research*, 24(12), 2311–7.
- Deckert, G., Warren, P. V., Gaasterland, T., Young, W. G., Lenox, A. L., Graham, D. E., ... Swanson, R. V. (1998). The complete genome of the hyperthermophilic bacterium *Aquifex aeolicus*. *Nature*, 392(6674), 353–8. doi:10.1038/32831
- Deutscher, M. P., & Reuven, N. B. (1991). Enzymatic basis for hydrolytic versus phosphorolytic mRNA degradation in *Escherichia coli* and *Bacillus subtilis*. *Proceedings of the National Academy of Sciences of the United States of America*, 88(8), 3277–80.
- Doniger, J., Landsman, D., Gonda, M. A., & Wistow, G. (1992). The product of unr, the highly conserved gene upstream of N-ras, contains multiple repeats similar to the cold-shock domain (CSD), a putative DNA-binding motif. *The New Biologist*, 4(4), 389–95.
- Ermolenko, D. N., & Makhatadze, G. I. (2002). Bacterial cold-shock proteins. *Cellular and Molecular Life Sciences : CMLS*, 59(11), 1902–13.
- Etchegaray, J. P., Jones, P. G., & Inouye, M. (1996). Differential thermoregulation of two highly homologous cold-shock genes, cspA and cspB, of *Escherichia coli*. *Genes to Cells : Devoted to Molecular & Cellular Mechanisms*, 1(2), 171–8.
- Etchegaray, J.-P., & Inouye, M. (1999). DB or not DB in translation? *Molecular Microbiology*, 33(2), 438–439. doi:10.1046/j.1365-2958.1999.01487.x
- Evans, J. R., Mitchell, S. A., Spriggs, K. A., Ostrowski, J., Bomsztyk, K., Ostarek, D., & Willis, A. E. (2003). Members of the poly (rC) binding protein family stimulate the activity of the c-myc internal ribosome entry segment in vitro and in vivo. *Oncogene*, 22(39), 8012–20. doi:10.1038/sj.onc.1206645

- Evdokimova, V. M., Kovrigina, E. A., Nashchekin, D. V., Davydova, E. K., Hershey, J. W., & Ovchinnikov, L. P. (1998). The major core protein of messenger ribonucleoprotein particles (p50) promotes initiation of protein biosynthesis in vitro. *The Journal of Biological Chemistry*, 273(6), 3574–81.
- Fang, L., Jiang, W., Bae, W., & Inouye, M. (1997). Promoter-independent cold-shock induction of *cspA* and its derepression at 37 degrees C by mRNA stabilization. *Molecular Microbiology*, 23(2), 355–64.
- Fanjul, María Luisa, Hiriart, M. (2008). *Biología funcional de los animales I / Functional Biology of Animals I* (pp. 55–86). México, D. F.: Siglo XXI.
- Ferrer, N., Garcia-Espana, A., Jeffers, M., & Pellicer, A. (1999). The *unr* gene: evolutionary considerations and nucleic acid-binding properties of its long isoform product. *DNA and Cell Biology*, 18(3), 209–18. doi:10.1089/104454999315420
- Gascuel, O. (1997). BIONJ: an improved version of the NJ algorithm based on a simple model of sequence data. *Molecular Biology and Evolution*, 14(7), 685–95.
- Giaquinto, L., Curmi, P. M. G., Siddiqui, K. S., Poljak, A., DeLong, E., DasSarma, S., & Cavicchioli, R. (2007). Structure and function of cold shock proteins in archaea. *Journal of Bacteriology*, 189(15), 5738–48. doi:10.1128/JB.00395-07
- Giuliodori, A. M. (2004). Preferential translation of cold-shock mRNAs during cold adaptation. *RNA*, 10(2), 265–276. doi:10.1261/rna.5164904
- Goldenberg, D., Azar, I., & Oppenheim, A. B. (1996). Differential mRNA stability of the *cspA* gene in the cold-shock response of *Escherichia coli*. *Molecular Microbiology*, 19(2), 241–8.
- Goldenberg, D., Azar, I., Oppenheim, A. B., Brandi, A., Pon, C. L., & Gualerzi, C. O. (1997). Role of *Escherichia coli cspA* promoter sequences and adaptation of translational apparatus in the cold shock response. *Molecular & General Genetics : MGG*, 256(3), 282–90.
- Goldstein, J., Pollitt, N. S., & Inouye, M. (1990). Major cold shock protein of *Escherichia coli*. *Proceedings of the National Academy of Sciences of the United States of America*, 87(1), 283–7.
- Goodchild, A., Raftery, M., Saunders, N. F. W., Guilhaus, M., & Cavicchioli, R. (n.d.). Biology of the cold adapted archaeon, *Methanococcoides burtonii* determined by proteomics using liquid chromatography-tandem mass spectrometry. *Journal of Proteome Research*, 3(6), 1164–76. doi:10.1021/pr0498988

- Goverde, R. L., Huis in't Veld, J. H., Kusters, J. G., & Mooi, F. R. (1998). The psychrotrophic bacterium *Yersinia enterocolitica* requires expression of pnp, the gene for polynucleotide phosphorylase, for growth at low temperature (5 degrees C). *Molecular Microbiology*, 28(3), 555–69.
- Graumann, P., & Marahiel, M. A. (1996). Some like it cold: response of microorganisms to cold shock. *Archives of Microbiology*, 166(5), 293–300.
- Graumann, P., Schröder, K., Schmid, R., & Marahiel, M. A. (1996). Cold shock stress-induced proteins in *Bacillus subtilis*. *Journal of Bacteriology*, 178(15), 4611–9.
- Graumann, P., Wendrich, T. M., Weber, M. H., Schröder, K., & Marahiel, M. A. (1997). A family of cold shock proteins in *Bacillus subtilis* is essential for cellular growth and for efficient protein synthesis at optimal and low temperatures. *Molecular Microbiology*, 25(4), 741–56.
- Graumann, P. L., & Marahiel, M. A. (1998). A superfamily of proteins that contain the cold-shock domain. *Trends in Biochemical Sciences*, 23(8), 286–90.
- Guindon, S., Dufayard, J.-F., Lefort, V., Anisimova, M., Hordijk, W., & Gascuel, O. (2010). New algorithms and methods to estimate maximum-likelihood phylogenies: assessing the performance of PhyML 3.0. *Systematic Biology*, 59(3), 307–21. doi:10.1093/sysbio/syq010
- Han, M. V., & Zmasek, C. M. (2009). phyloXML: XML for evolutionary biology and comparative genomics. *BMC Bioinformatics*, 10, 356. doi:10.1186/1471-2105-10-356
- Hankins, J. S., Zappavigna, C., Prud'homme-Généreux, A., & Mackie, G. A. (2007). Role of RNA structure and susceptibility to RNase E in regulation of a cold shock mRNA, cspA mRNA. *Journal of Bacteriology*, 189(12), 4353–8. doi:10.1128/JB.00193-07
- Horn, G., Hofweber, R., Kremer, W., & Kalbitzer, H. R. (2007). Structure and function of bacterial cold shock proteins. *Cellular and Molecular Life Sciences : CMLS*, 64(12), 1457–70. doi:10.1007/s00018-007-6388-4
- Hu, K. H., Liu, E., Dean, K., Gingras, M., DeGraff, W., & Trun, N. J. (1996). Overproduction of Three Genes Leads to Camphor Resistance and Chromosome Condensation in *Escherichia coli*. *Genetics*, 143(4), 1521–1532.
- Hunt, S. L., Hsuan, J. J., Totty, N., & Jackson, R. J. (1999). unr, a cellular cytoplasmic RNA-binding protein with five cold-shock domains, is required for internal initiation of translation of human rhinovirus RNA. *Genes & Development*, 13(4), 437–48.

- Jacquemin-Sablon, H., Triqueneaux, G., Deschamps, S., le Maire, M., Doniger, J., & Dautry, F. (1994). Nucleic acid binding and intracellular localization of unr, a protein with five cold shock domains. *Nucleic Acids Research*, *22*(13), 2643–50.
- Jeffers, M., Paciucci, R., & Pellicer, A. (1990). Characterization of unr; a gene closely linked to N-ras. *Nucleic Acids Research*, *18*(16), 4891–9.
- Jiang, W., Fang, L., & Inouye, M. (1996). The role of the 5'-end untranslated region of the mRNA for CspA, the major cold-shock protein of *Escherichia coli*, in cold-shock adaptation. *Journal of Bacteriology*, *178*(16), 4919–25.
- Jiang, W., Hou, Y., & Inouye, M. (1997). CspA, the major cold-shock protein of *Escherichia coli*, is an RNA chaperone. *The Journal of Biological Chemistry*, *272*(1), 196–202.
- Jones, P., Bogelen, R. Van, & Neidhardt, F. (1987). Induction of proteins in response to low temperature in *Escherichia Cmh.* *Bacteriol 1987*; *169*[^]
- Jones, P. G., & Inouye, M. (1996). RbfA, a 30S ribosomal binding factor, is a cold-shock protein whose absence triggers the cold-shock response. *Molecular Microbiology*, *21*(6), 1207–18.
- Jurchott, K., Bergmann, S., Stein, U., Walther, W., Janz, M., Manni, I., ... Royer, H.-D. (2003). YB-1 as a cell cycle-regulated transcription factor facilitating cyclin A and cyclin B1 gene expression. *The Journal of Biological Chemistry*, *278*(30), 27988–96. doi:10.1074/jbc.M212966200
- Karlson, D., Nakaminami, K., Toyomasu, T., & Imai, R. (2002). A cold-regulated nucleic acid-binding protein of winter wheat shares a domain with bacterial cold shock proteins. *The Journal of Biological Chemistry*, *277*(38), 35248–56. doi:10.1074/jbc.M205774200
- Karlson, D., & Imai, R. (2003). Conservation of the cold shock domain protein family in plants. *Plant Physiology*, *131*(1), 12–5. doi:10.1104/pp.014472
- Katoh, K. (2002). MAFFT: a novel method for rapid multiple sequence alignment based on fast Fourier transform. *Nucleic Acids Research*, *30*(14), 3059–3066. doi:10.1093/nar/gkf436
- Katoh, K., & Standley, D. M. (2013). MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Molecular Biology and Evolution*, *30*(4), 772–80. doi:10.1093/molbev/mst010
- Khemic, V., Toesca, I., Poljak, L., Vanzo, N. F., & Carpousis, A. J. (2004). The RNase E of *Escherichia coli* has at least two binding sites for DEAD-box RNA helicases:

functional replacement of RhlB by RhlE. *Molecular Microbiology*, 54(5), 1422–30. doi:10.1111/j.1365-2958.2004.04361.x

- Kimura, M. (1983). *The Neutral Theory of Molecular Evolution*. Cambridge, UK: Cambridge University Press.
- Kloks, C. P. A. M., Spronk, C. A. E. M., Lasonder, E., Hoffmann, A., Vuister, G. W., Grzesiek, S., & Hilbers, C. W. (2002). The solution structure and DNA-binding properties of the cold-shock domain of the human Y-box protein YB-1. *Journal of Molecular Biology*, 316(2), 317–26. doi:10.1006/jmbi.2001.5334
- Kumar, A., Malloch, R. A., Fujita, N., Smillie, D. A., Ishihama, A., & Hayward, R. S. (1993). The minus 35-recognition region of *Escherichia coli* sigma 70 is inessential for initiation of transcription at an “extended minus 10” promoter. *Journal of Molecular Biology*, 232(2), 406–18. doi:10.1006/jmbi.1993.1400
- Le, S. Q., & Gascuel, O. (2008). An improved general amino acid replacement matrix. *Molecular Biology and Evolution*, 25(7), 1307–20. doi:10.1093/molbev/msn067
- Le, S. Q., Lartillot, N., & Gascuel, O. (2008). Phylogenetic mixture models for proteins. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*, 363(1512), 3965–76. doi:10.1098/rstb.2008.0180
- Le, S. Q., & Gascuel, O. (2010). Accounting for solvent accessibility and secondary structure in protein phylogenetics is clearly beneficial. *Systematic Biology*, 59(3), 277–87. doi:10.1093/sysbio/syq002
- Lee, S. J., Xie, A., Jiang, W., Etchegaray, J. P., Jones, P. G., & Inouye, M. (1994). Family of the major cold-shock protein, CspA (CS7.4), of *Escherichia coli*, whose members show a high sequence similarity with the eukaryotic Y-box binding proteins. *Molecular Microbiology*, 11(5), 833–9.
- Letunic, I., & Bork, P. (2011). Interactive Tree Of Life v2: online annotation and display of phylogenetic trees made easy. *Nucleic Acids Research*, 39(Web Server issue), W475–8. doi:10.1093/nar/gkr201
- Lim, J., Thomas, T., & Cavicchioli, R. (2000). Low temperature regulated DEAD-box RNA helicase from the Antarctic archaeon, *Methanococcoides burtonii*. *Journal of Molecular Biology*, 297(3), 553–67. doi:10.1006/jmbi.2000.3585
- Matsumoto, K., Meric, F., & Wolffe, A. P. (1996). Translational repression dependent on the interaction of the *Xenopus* Y-box protein FRGY2 with mRNA. Role of the cold shock domain, tail domain, and selective RNA sequence recognition. *The Journal of Biological Chemistry*, 271(37), 22706–12.

- Matsumoto, K., & Wolffe, A. P. (1998). Gene regulation by Y-box proteins: coupling control of transcription and translation. *Trends in Cell Biology*, 8(8), 318–23.
- Matsumoto, K., Tanaka, K. J., Aoki, K., Sameshima, M., & Tsujimoto, M. (2003). Visualization of the reconstituted FRGY2-mRNA complexes by electron microscopy. *Biochemical and Biophysical Research Communications*, 306(1), 53–8.
- Max, K. E. A., Zeeb, M., Bienert, R., Balbach, J., & Heinemann, U. (2007). Common mode of DNA binding to cold shock domains. Crystal structure of hexathymidine bound to the domain-swapped form of a major cold shock protein from *Bacillus caldolyticus*. *The FEBS Journal*, 274(5), 1265–79. doi:10.1111/j.1742-4658.2007.05672.x
- Mihailovich, M., Militti, C., Gabaldón, T., & Gebauer, F. (2010). Eukaryotic cold shock domain proteins: highly versatile regulators of gene expression. *BioEssays : News and Reviews in Molecular, Cellular and Developmental Biology*, 32(2), 109–18. doi:10.1002/bies.200900122
- Mitchell, S. A., Spriggs, K. A., Coldwell, M. J., Jackson, R. J., & Willis, A. E. (2003). The Apaf-1 internal ribosome entry segment attains the correct structural conformation for function via interactions with PTB and unr. *Molecular Cell*, 11(3), 757–71.
- Mitta, M., Fang, L., & Inouye, M. (1997). Deletion analysis of cspA of *Escherichia coli* : requirement of the AT-rich UP element for cspA transcription and the downstream box in the coding region for its cold shock induction. *Molecular Microbiology*, 26(02), 321–335. doi:10.1046/j.1365-2958.1997.5771943.x
- Miyata, T., Miyazawa, S., & Yasunaga, T. (1979). Two types of amino acid substitutions in protein evolution. *Journal of Molecular Evolution*, 12(3), 219–36.
- Morita, R. Y. (1975). Psychrophilic bacteria. *Bacteriological Reviews*, 39(2), 144–67.
- Moss, E. G., & Tang, L. (2003). Conservation of the heterochronic regulator Lin-28, its developmental expression and microRNA complementary sites. *Developmental Biology*, 258(2), 432–42.
- Murata, N., & Los, D. A. (1997). Membrane Fluidity and Temperature Perception. *Plant Physiology*, 115(3), 875–879.
- Murzin, A. G. (1993). OB(oligonucleotide/oligosaccharide binding)-fold: common structural and functional solution for non-homologous sequences. *The EMBO Journal*, 12(3), 861–7.
- Murzin, A. G. (1998). How far divergent evolution goes in proteins?. *Current Opinion in Structural Biology*, 8(3), 380–7.

- Nakaminami, K., Sasaki, K., Kajita, S., Takeda, H., Karlson, D., Ohgi, K., & Imai, R. (2005). Heat stable ssDNA/RNA-binding activity of a wheat cold shock domain protein. *FEBS Letters*, *579*(21), 4887–91. doi:10.1016/j.febslet.2005.07.074
- Nakaminami, K., Karlson, D. T., & Imai, R. (2006). Functional conservation of cold shock domains in bacteria and higher plants. *Proceedings of the National Academy of Sciences of the United States of America*, *103*(26), 10122–7. doi:10.1073/pnas.0603168103
- Nakashima, K., Kanamaru, K., Mizuno, T., & Horikoshi, K. (1996). A novel member of the cspA family of genes that is induced by cold shock in *Escherichia coli*. *Journal of Bacteriology*, *178*(10), 2994–7.
- Neuhaus, K., Rapposch, S., Francis, K. P., & Scherer, S. (2000). Restart of Exponential Growth of Cold-Shocked *Yersinia enterocolitica* Occurs after Down-Regulation of cspA1/A2 mRNA. *Journal of Bacteriology*, *182*(11), 3285–3288. doi:10.1128/JB.182.11.3285-3288.2000
- Nichols, C. M., Lardièrre, S. G., Bowman, J. P., Nichols, P. D., A E Gibson, J., & Guézennec, J. (2005). Chemical characterization of exopolysaccharides from Antarctic marine bacteria. *Microbial Ecology*, *49*(4), 578–89. doi:10.1007/s00248-004-0093-8
- Nimmo, R. A., & Slack, F. J. (2009). An elegant miRror: microRNAs in stem cells, developmental timing and cancer. *Chromosoma*, *118*(4), 405–18. doi:10.1007/s00412-009-0210-z
- Patalano, S., Mihailovich, M., Belacortu, Y., Paricio, N., & Gebauer, F. (2009). Dual sex-specific functions of *Drosophila* Upstream of N-ras in the control of X chromosome dosage compensation. *Development (Cambridge, England)*, *136*(4), 689–98. doi:10.1242/dev.027656
- Perl, D., Welker, C., Schindler, T., Schröder, K., Marahiel, M. A., Jaenicke, R., & Schmid, F. X. (1998). Conservation of rapid two-state folding in mesophilic, thermophilic and hyperthermophilic cold shock proteins. *Nature Structural Biology*, *5*(3), 229–35.
- Phadtare, S., Alsina, J., & Inouye, M. (1999). Cold-shock response and cold-shock proteins. *Current Opinion in Microbiology*, *2*(2), 175–80. doi:10.1016/S1369-5274(99)80031-9
- Phadtare, S. (2004). Recent developments in bacterial cold-shock response. *Current Issues in Molecular Biology*, *6*(2), 125–36.
- Phadtare, S., & Severinov, K. (2005). Nucleic acid melting by *Escherichia coli* CspE. *Nucleic Acids Research*, *33*(17), 5583–90. doi:10.1093/nar/gki859

- Phadtare, S., & Severinov, K. (2009). Comparative analysis of changes in gene expression due to RNA melting activities of translation initiation factor IF1 and a cold shock protein of the CspA family. *Genes to Cells : Devoted to Molecular & Cellular Mechanisms*, 14(11), 1227–39. doi:10.1111/j.1365-2443.2009.01346.x
- Phadtare, S., Zhu, L., Uemori, T., Mukai, H., Kato, I., & Inouye, M. (2009). Applications of nucleic acid chaperone activity of CspA and its homologues. *Journal of Molecular Microbiology and Biotechnology*, 17(3), 110–7. doi:10.1159/000226587
- Pisarev, A. V., Skabkin, M. A., Thomas, A. A., Merrick, W. C., Ovchinnikov, L. P., & Shatsky, I. N. (2002). Positive and negative effects of the major mammalian messenger ribonucleoprotein p50 on binding of 40 S ribosomal subunits to the initiation codon of beta-globin mRNA. *The Journal of Biological Chemistry*, 277(18), 15445–51. doi:10.1074/jbc.M111954200
- Prud'homme-Généreux, A., Beran, R. K., Iost, I., Ramey, C. S., Mackie, G. A., & Simons, R. W. (2004). Physical and functional interactions among RNase E, polynucleotide phosphorylase and the cold-shock protein, CsdA: evidence for a “cold shock degradosome”. *Molecular Microbiology*, 54(5), 1409–21. doi:10.1111/j.1365-2958.2004.04360.x
- Ray, M. K., Kumar, G. S., & Shivaji, S. (1994). Phosphorylation of membrane proteins in response to temperature in an Antarctic *Pseudomonas syringae*. *Microbiology (Reading, England)*, 140 (Pt 1), 3217–23.
- Ross, W., Gosink, K. K., Salomon, J., Igarashi, K., Zou, C., Ishihama, A., Gourse, R. L. (1993). A third recognition element in bacterial promoters: DNA binding by the alpha subunit of RNA polymerase. *Science (New York, N.Y.)*, 262(5138), 1407–13.
- Russell, N. J. (1997). Psychrophilic bacteria—Molecular adaptations of membrane lipids. *Comparative Biochemistry and Physiology Part A: Physiology*, 118(3), 489–493. doi:10.1016/S0300-9629(97)87354-9
- Sasaki, K., Kim, M.-H., & Imai, R. (2007). Arabidopsis COLD SHOCK DOMAIN PROTEIN2 is a RNA chaperone that is regulated by cold and developmental signals. *Biochemical and Biophysical Research Communications*, 364(3), 633–8. doi:10.1016/j.bbrc.2007.10.059
- Sasaki, K., & Imai, R. (2011). Pleiotropic roles of cold shock domain proteins in plants. *Frontiers in Plant Science*, 2, 116. doi:10.3389/fpls.2011.00116
- Schindler, T., Graumann, P. L., Perl, D., Ma, S., Schmid, F. X., & Marahiel, M. A. (1999). The family of cold shock proteins of *Bacillus subtilis*. Stability and dynamics in vitro and in vivo. *The Journal of Biological Chemistry*, 274(6), 3407–13.

- Siddiqui, K. S., Williams, T. J., Wilkins, D., Yau, S., Allen, M. A., Brown, M. V., ... Cavicchioli, R. (2013). Psychrophiles. *Annual Review of Earth and Planetary Sciences*, 41(1), 87–115. doi:10.1146/annurev-earth-040610-133514
- Sleator, R. D. (2012). Proteins: form and function. *Bioengineered Bugs*, 3(2), 80–5. doi:10.4161/bbug.18303
- Sprenghart, M. L., Fuchs, E., & Porter, A. G. (1996). The downstream box: an efficient and independent translation initiation signal in *Escherichia coli*. *The EMBO Journal*, 15(3), 665–74.
- Tanabe, H., Goldstein, J., Yang, M., & Inouye, M. (1992). Identification of the promoter region of the *Escherichia coli* major cold shock gene, *cspA*. *Journal of Bacteriology*, 174(12), 3867–73.
- Tedin, K., Resch, A., & Bläsi, U. (1997). Requirements for ribosomal protein S1 for translation initiation of mRNAs with and without a 5' leader sequence. *Molecular Microbiology*, 25(01), 189–199. doi:10.1046/j.1365-2958.1997.4421810.x
- Thieringer, H. A., Jones, P. G., & Inouye, M. (1998). Cold shock and adaptation. *BioEssays : News and Reviews in Molecular, Cellular and Developmental Biology*, 20(1), 49–57. doi:10.1002/(SICI)1521-1878(199801)20:1<49::AID-BIES8>3.0.CO;2-N
- VanBogelen, R. A., & Neidhardt, F. C. (1990). Ribosomes as sensors of heat and cold shock in *Escherichia coli*. *Proceedings of the National Academy of Sciences of the United States of America*, 87(15), 5589–93.
- Vigh, L., Maresca, B., & Harwood, J. L. (1998). Does the membrane's physical state control the expression of heat shock and other genes? *Trends in Biochemical Sciences*, 23(10), 369–74.
- Wang, N., Yamanaka, K., & Inouye, M. (1999). CspI, the ninth member of the CspA family of *Escherichia coli*, is induced upon cold shock. *Journal of Bacteriology*, 181(5), 1603–9.
- Weber, M. H., Klein, W., Müller, L., Niess, U. M., & Marahiel, M. A. (2001). Role of the *Bacillus subtilis* fatty acid desaturase in membrane adaptation during cold shock. *Molecular Microbiology*, 39(5), 1321–9.
- Weber, M. H. W., & Marahiel, M. A. (2002). Coping with the cold: the cold shock response in the Gram-positive soil bacterium *Bacillus subtilis*. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*, 357(1423), 895–907. doi:10.1098/rstb.2002.1078

- Xia, B., Ke, H., & Inouye, M. (2001). Acquisition of cold sensitivity by quadruple deletion of the cspA family and its suppression by PNPase S1 domain in *Escherichia coli*. *Molecular Microbiology*, *40*(1), 179–88.
- Yamanaka, K., Mitani, T., Ogura, T., Niki, H., & Hiraga, S. (1994). Cloning, sequencing, and characterization of multicopy suppressors of a mukB mutation in *Escherichia coli*. *Molecular Microbiology*, *13*(2), 301–312. doi:10.1111/j.1365-2958.1994.tb00424.x
- Yamanaka, K., & Inouye, M. (1997). Growth-phase-dependent expression of cspD, encoding a member of the CspA family in *Escherichia coli*. *Journal of Bacteriology*, *179*(16), 5126–30.
- Yamanaka, K., Fang, L., & Inouye, M. (1998). The CspA family in *Escherichia coli*: multiple gene duplication for stress adaptation. *Molecular Microbiology*, *27*(2), 247–55.
- Yamanaka, K. (1999). Cold shock response in *Escherichia coli*. *Journal of Molecular Microbiology and Biotechnology*, *1*(2), 193–202.
- Yamanaka, K., & Inouye, M. (2001). Selective mRNA degradation by polynucleotide phosphorylase in cold shock adaptation in *Escherichia coli*. *Journal of Bacteriology*, *183*(9), 2808–16. doi:10.1128/JB.183.9.2808-2816.2001
- Zhou, Z., & Deutscher, M. P. (1997). An essential function for the phosphate-dependent exoribonucleases RNase PH and polynucleotide phosphorylase. *Journal of Bacteriology*, *179*(13), 4391–5.